STATVERSITY

User Guide

A plain-language guide to analyzing your data with confidence.

Version 1.6.0 · November 16, 2026

Contents

Section 1: Getting started **4–25**

Chapter 1 - Welcome to Statversity 5–9

Chapter 2 - Installing and opening Statversity 10–17

Chapter 3 - Your first five minutes 18–25

Section 2: Screen by screen **26–123**

Chapter 4 - The top bar and project basics 27–36

Chapter 5 - Getting data in 37–53

Chapter 6 - Variables 54–68

Chapter 7 - Analyze 69–85

Chapter 8 - Results 86–99

Chapter 9 - Charts, Responses, Study Planner, Export and Print 100–113

Chapter 10 - Working without the Assistant 114–123

Section 3: Walkthroughs **124–333**

Chapter 11 - Did the training help? 125–158

Chapter 12 - Which intervention worked best? 159–196

Chapter 13 - Cleaning up a real export 197–224

Chapter 14 - Matching the same student across two files 225–257

Chapter 15 - Does my survey measure what I think it does? 258–292

Chapter 16 - Yes/no data, four different ways 293–323

Chapter 17 - Seeing how your survey branches 324–333

Section 4: Recipes **334–380**

Chapter 18 - What predicts the posttest score? 335–343

Chapter 19 - Can we trust the graders? 344–350

Chapter 20 - A full study at realistic size 351–362

Chapter 21 - Does tutoring work through confidence? 363–371

Chapter 22 - Durations and dates 372–380

Section 5: Understanding what you see 381–414

Chapter 23 - Reading a result when you are new to statistics 382–392

Chapter 24 - Glossary 393–404

Chapter 25 - Troubleshooting and FAQ 405–414

SECTION 1

Getting started

CHAPTER 1

Welcome to Statversity

Welcome. This guide teaches you to use Statversity, a desktop program that helps you analyze survey data. It is written for people who may never have run a statistical test before. If you can open a spreadsheet, you can follow it.

Who this guide is for

The guide assumes you are a graduate student, or someone in a similar position, who has collected survey data and now has to say something sensible about it. Maybe you ran a training and surveyed participants before and after. Maybe you compared two classes, or you want to know whether a questionnaire really measures one idea. You do not need to remember anything from a statistics course. Every technical word is explained the first time it appears, and the gray **novice** boxes give plain-language explanations you can skip once you know the idea.

You also do not need to know how to program. Everything in Statversity is done by clicking, typing a name, or choosing from a list.

What Statversity is

Statversity is a desktop app. You install it on your own computer, and your data stays there. It does four jobs, in this order:

1. **It imports survey exports.** You give it a file from Qualtrics, Excel, a CSV file, or one of the common statistics formats (SPSS, Stata, SAS). Statversity reads the file, notices the quirks of survey exports, and helps you tidy them.
2. **It helps you choose the right test through a conversation.** Instead of asking you to know the name of the test, Statversity asks plain questions: Did the same people answer twice? Are you comparing groups? Statversity then recommends a test and tells you why.
3. **It runs the test with a real statistics engine.** The calculations are done by established statistical code, not by a quick approximation, and the numbers are checked against the R programming language during development.
4. **It explains the result twice.** You get a plain-English paragraph that says what the result means, and you get the same result written the way a journal expects it, in APA form, with a sentence and a table you can copy into your paper.

> **NOTE**
> Statversity never sends your data or your results anywhere. The only thing it ever asks the internet for is which version of Statversity is newest, so it can tell you when an update is available, and you can turn even that off (see Chapter 2).

What Statversity does not do

Knowing the limits saves you time.

- **It does not decide your research question for you.** It can recommend a test, but you decide what you want to find out.
- **It does not replace a statistician for unusual designs.** The Test Advisor covers the common designs that graduate students meet. If your design is truly unusual, talk to an advisor or a methods consultant.
- **It does not write your paper.** It gives you the sentences and tables for the results section. Introduction, method, and discussion are yours.
- **It does not fix bad data.** It helps you spot duplicates, missing answers, and awkward coding, but if a survey question was badly written, no software can rescue it.
- **It does not produce syntax or scripts.** There is no code to copy. The record of what you did is the Test Log, described below.

How this guide is organised

The guide has five parts.

- **Getting started** (chapters 1 to 3) gets you installed and through your first analysis.
- **Screen by screen** (chapters 4 to 10) is the reference. Each chapter takes one area of the program and describes every control in it. Read these when you want to know what a button does.
- **Walkthroughs** (chapters 11 to 17) are seven complete tasks, from opening the data to the finished result. Chapter 17 shows how to read a survey that sends students down different branches.
- **Recipes** (chapters 18 to 22) are shorter, with the question, the steps, and what to look for.
- **Understanding what you see** (chapters 23 to 25) explains p-values, effect sizes and assumptions for beginners, and includes a glossary and a troubleshooting section.

If you are brand new, read chapters 1 to 3, then pick the walkthrough closest to your own study. If you get stuck on a word, look it up in the glossary.

How to read the instructions

Instructions use the exact words on the screen. A button or menu item appears in bold, such as **Choose files…**. Steps are numbered, one action per step, and each tells you what you should see afterward. When a chapter shows a screenshot with numbered badges drawn on it, the table directly under the picture lists what each number points at. The screenshots were captured in the dark theme. If you use the light theme, the layout is the same and only the colors differ.

The 60-second tour

Once you have a dataset open, a row of tabs appears under the top bar. The tabs are the main rooms of the program. Here is each one in a paragraph.

Data. This is your data as a table, one row per respondent and one column per question. You can click into a cell to fix a typo, add or delete rows, look for duplicates, keep only some rows with a filter, or combine this dataset with another file. Every change can be undone.

Variables. A variable is just a column, a question or a measurement. This tab lists them all and lets you say what each one is: an ID, a group label, a test question, a survey rating. It is also where you set up scales (a score made from several related questions) and calculated variables such as a gain score.

Analyze. This is where you choose and run a test. With the Statversity Assistant on, you open the Test Advisor, answer a few questions, and get a recommendation. You can also ask Statversity to plan a whole set of analyses at once.

Test Log. Every test you run is kept here, with its inputs and results, and it is saved with the project. You can reopen a result, re-run it if your data changed, or group several tests together when you want to correct for running many of them.

Charts. Build bar charts, box plots, Likert charts and more from your data, adjust their look, and save them as images for your paper.

Responses. If your survey has written answers, this is where you read them, tag them with themes, and count the themes.

Learn. Learn is not a tab under the top bar. It is a button in the top bar itself. It opens plain-language guides to every test, assumption and effect size in Statversity, and most guides have a **Run this test** button.

The screenshot below shows the Home screen, which is where Statversity opens.

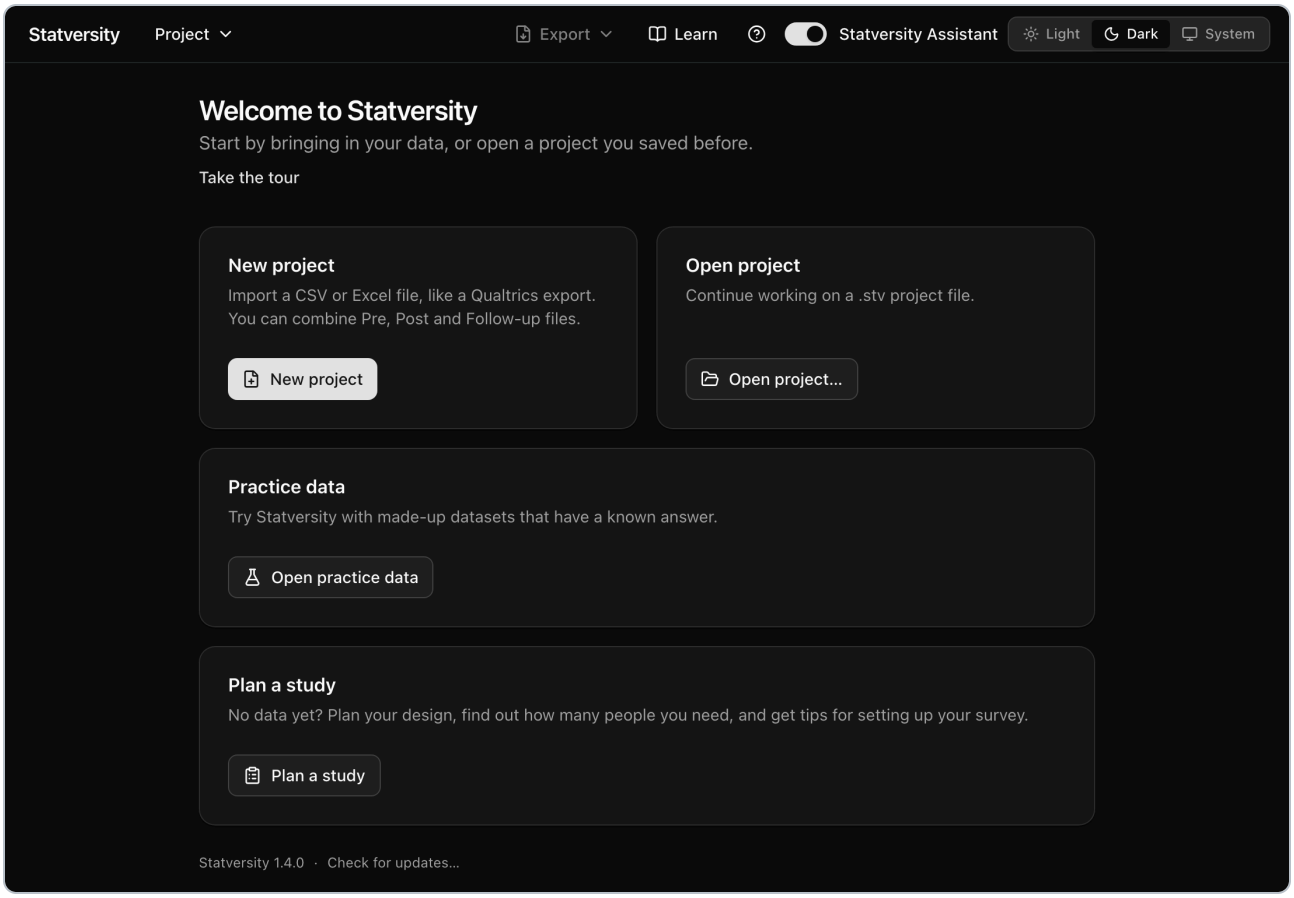


Figure 1. The Home screen when Statversity opens.

Home has four cards: **New project**, **Open project**, **Open practice data**, and **Plan a study**. If you do not have data yet, **Open practice data** gives you datasets to learn on.

Two ways to work

The top bar has a switch called **Statversity Assistant**. It is on by default, and this guide assumes it is on.

With the Assistant on, Statversity guides you. After you import data it asks about your study, then walks you through a short interview about your variables. On the Analyze tab it opens the Test Advisor, which asks questions and recommends a test. Before it runs the test it walks you through the assumption checks, which are the conditions a test needs in order to give trustworthy results, and lets you make the final choice.

With the Assistant off, Statversity gets out of your way. After import it takes you straight to the Variables screen. The Analyze tab shows a searchable list of every test, called **All tests**, and when you pick one it goes straight to choosing variables and running. This mode suits people who already know which test they want. It is covered in chapter 10.

> **IN PLAIN LANGUAGE**
> If you are not sure which to use, leave the Assistant on. It asks more questions, but those questions are what protect you from running a test that does not fit your data.

You can flip the switch at any time. It does not change your data or your saved results.

What to do next

Install the program if you have not yet (chapter 2), then follow chapter 3 for your first analysis. Come back to the reference chapters whenever you want to know exactly what a control does.

CHAPTER 2

Installing and opening Statversity

This chapter gets Statversity onto your computer and open for the first time. The Mac version is the one that is shipped today, so it comes first and gets the most detail. It takes about five minutes. Most of that time goes to one confirmation that every new Statversity user sees once, the first time the app opens.

> NOTE
>
> **Statversity never sends your data anywhere.** Your data, your results, and your project files stay on your own computer, and Statversity works without a Wi-Fi connection. The one thing it ever asks the internet for is which version of Statversity is newest, so it can tell you when an update is out. That check sends nothing about you or your work, and you can turn it off in the **Software updates** dialog (see "Updating Statversity" below). Apart from that, you only need the internet to download the installer (or you can receive the installer on a flash drive or a shared drive).

What you need

- A Mac. Statversity is built for both Apple Silicon Macs (M1, M2, M3, M4) and Intel Macs, and each has its own installer file. To find out which you have, click the Apple menu in the top-left corner of your screen and choose **About This Mac**. A line that starts with **Chip** followed by "Apple M…" means Apple Silicon. A line that starts with **Processor** followed by "Intel" means Intel.
- The Statversity installer file. It ends in `.dmg` (a "disk image," which is the standard way Mac apps are handed out). The file name tells you which Mac it is for: names containing `aarch64-apple-darwin` are for Apple Silicon, and names containing `x86_64-apple-darwin` are for Intel.
- Permission to put an app in your **Applications** folder. On a university-managed computer you may need an administrator password.

Install on a Mac

Step 1. Get the installer

Statversity is distributed by the person who runs the project. You will normally receive the `.dmg` file in one of these ways:

- A download link from your instructor, lab, or the project owner. Releases are posted on the project's GitHub **Releases** page, and the files can be downloaded without a GitHub account.
- A shared drive link (for example Google Drive) or a flash drive. Many email systems block or strip large installer attachments, so Statversity's installer is usually shared this way rather than by

email.

Save the `.dmg` somewhere you will find it again, such as your **Downloads** folder.

Step 2. Open the disk image

1. In the Finder, open **Downloads** (or wherever you saved the file).
2. Double-click the file that ends in `.dmg`.
3. After a moment, a small window opens showing the Statversity icon, usually next to a shortcut to your **Applications** folder.

Step 3. Drag Statversity into Applications

1. Drag the **Statversity** icon onto the **Applications** folder in that window (or open a second Finder window on **Applications** and drop it there).
2. Wait for the copy to finish. A progress bar appears and then disappears.
3. Close the disk image window. In the Finder sidebar, click the small eject button next to the Statversity disk image, or drag the disk image to the Trash. This does not delete Statversity; it only "unplugs" the installer.

Step 4. Open Statversity for the first time

1. Open your **Applications** folder and double-click **Statversity**.
2. The first time only, macOS asks: "Statversity" is an app downloaded from the internet. Are you sure you want to open it? Underneath it says **Apple checked it for malicious software and none was detected**. That line is there because Statversity is signed by Anatomical Consulting LLC and has been checked by Apple's notarization service.
3. Click **Open**. Statversity opens, and macOS will not ask again. There are no settings to change.

> **TIP**
>
> **If macOS does show a warning** that Statversity "cannot be verified" or is "damaged," you are almost certainly looking at an old or altered copy, for example one that was forwarded by email or synced through a cloud folder. Do not try to force it open. Delete that copy, download the installer again from **statversity.com**, and install it from the fresh disk image.

Step 5. Confirm that Statversity is ready

When Statversity opens, it starts its statistics engine, the part of the program that does the calculations. You will see an "engine ready" card that shows two version numbers: the version of the app and the version of the statistics engine. Wait for that card before you import anything. On the very first launch it can take a few seconds.

If the card does not appear, or Statversity reports an engine problem, see “Installing and opening” in the chapter on troubleshooting. Statversity keeps a log of what the engine did at `~/Library/Logs/edu.statversity.desktop/engine.log`, which is the first thing to send if you ask someone for help.

> **TRY IT YOURSELF**
> Quit Statversity with `Cmd` + `Q` and open it again from **Applications**. It should open straight away with no warning. If it does, the install is finished.

Where your work is saved

Statversity keeps your work in a **project file**, which ends in `.stv`. A project holds your data, your variable settings, your results, your charts, and your Test Log in a single file. Older `.statly` files still open. Statversity does not hide your project in some hidden location; you choose where it goes. The first time you save, Statversity shows the usual Mac save window and you choose the folder and the file name. Pick a folder you back up, such as **Documents**, a folder in your cloud drive, or your lab’s shared drive.

- Double-clicking a `.stv` file in the Finder opens Statversity with that project loaded. If Statversity is already running, it opens the file after checking whether you have unsaved changes in the project you are working on. (On a Mac, this connection between `.stv` files and Statversity starts working once you have launched Statversity once.)
- Your original data files are never changed. Statversity reads them and keeps its own copy inside the project.
- While you work, Statversity also makes automatic backups. If Statversity closes unexpectedly, the next launch offers **Recover unsaved work**. This is a safety net, not a substitute for saving. Chapter 4 covers saving and opening in detail.

Updating Statversity

Statversity does not install updates by itself. It tells you when a newer version is out, and you install it yourself. Updates are free for as long as you use Statversity.

How you find out. Each time Statversity opens, it asks statversity.com which version is newest. If yours is older, the Home screen shows a banner: **Statversity 1.5.0 is available.** (the number will match the new version) with **See what’s new** and **Not now**. **Not now** hides the banner for that version.

Checking on your own. You can check at any time. On a Mac, open the **Statversity** menu (next to the Apple menu) and choose **Check for Updates…**. On a Mac or Windows, you can also click **Check for updates…** in the footer at the bottom of the Home screen, next to “Statversity 1.4.0”. Either one opens a

Software updates dialog that says **Checking statversity.com for the latest version…** and then tells you one of two things:

- **You have the latest version of Statversity (1.4.0).** Nothing more to do.
- **Statversity 1.5.0 is available. You have 1.4.0.** The dialog lists what is new, with **Download from statversity.com** and **Not now** buttons.

Installing the new version (Mac).

1. Click **Download from statversity.com**. This opens the download page in your web browser. Download the new `.dmg`.
2. Save your work and quit Statversity with `Cmd` + `Q`.
3. Open the new `.dmg` and drag the new **Statversity** onto **Applications**, over the old one. macOS asks whether to replace it. Choose **Replace**.
4. Open Statversity. Your project files are separate from the app, so they are not affected, and you open them the same way as before.

Installing the new version (Windows). Click **Download from statversity.com**, download the new installer, quit Statversity, and run the new installer. It replaces the old copy.

What the check sends. Statversity only asks statversity.com which version is newest. It never sends your data or your project. If you would rather it did not check on its own, untick **Check for updates automatically when Statversity opens** in the Software updates dialog (it is on by default). You can still check by hand whenever you like.

Uninstalling Statversity

1. Quit Statversity.
2. Open **Applications** and drag **Statversity** to the Trash. Then empty the Trash.

This removes the app. It does **not** delete your `.stv` project files, your original data files, or exports, because those live in the folders you chose. Delete those yourself if you want them gone. Statversity also leaves its small log folder at `~/Library/Logs/edu.statversity.desktop`, which you can delete in the Finder (choose **Go**, then **Go to Folder…**, and paste that path).

Windows

A Windows 10/11 installer (a file ending in `-setup.exe`) is produced by the same build process, but the version shipped and tested for this guide is the Mac one, so the rest of this guide shows Mac screens. If you are given the Windows installer:

1. Run the downloaded `-setup.exe` and accept the defaults.

2. If Windows shows a blue **Windows protected your PC** screen, click **More info**, then **Run anyway**. This happens because the Windows installer is not yet signed with a Windows code-signing certificate, and typically only the first time.
3. To uninstall, open **Settings**, then **Apps**, then **Installed apps**, choose **Statversity**, and click **Uninstall**. The engine log is at `%LOCALAPPDATA%\edu.statversity.desktop\logs\engine.log`.

Linux

There is no Linux version of Statversity. Only the Mac version is shipped today, with a Windows installer available as described above.

If something goes wrong

The troubleshooting chapter's first section lists the problems people hit while installing and opening Statversity, with the exact fix for each.

Your free trial and license key

Statversity is free to try for **7 days from the first time you open it**. You do not need a key, an account, or an internet connection to start. During the trial, a slim bar at the top of the Home screen tells you how many days are left, for example **Free trial: 5 days left**. You can close it with the **x** on its right edge. It comes back the next time you open Statversity.

When you buy a license, you receive a **license key**: a line of text that starts with `STV-`. To enter it:

1. On the Home screen, click **License…** in the footer at the bottom (next to **Check for updates…**). You can also click **Enter a license key** in the trial bar. The **License** dialog opens.
2. Paste the key into the **License key** box, then click **Unlock**.

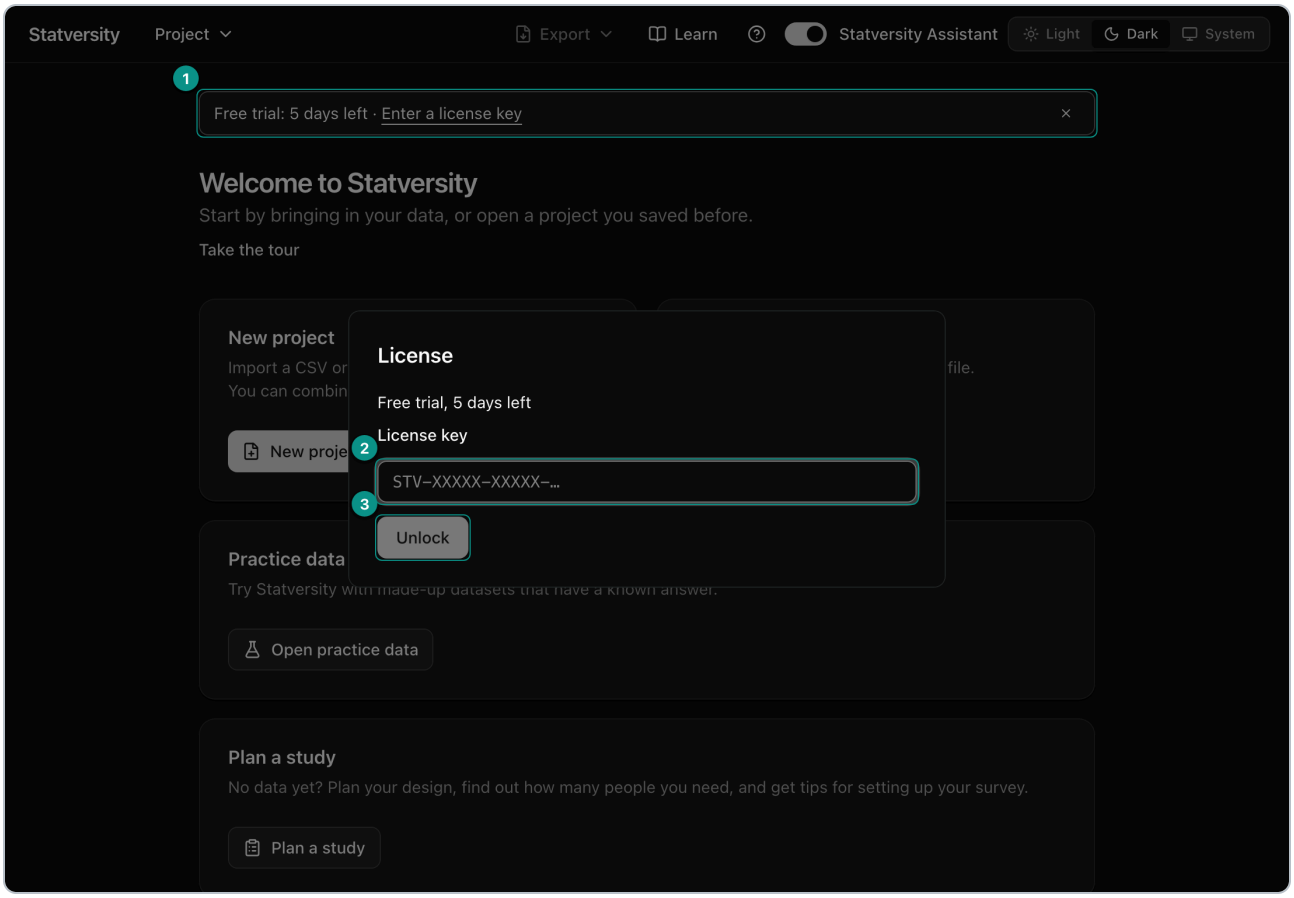


Figure 2. The Home screen with the trial bar at the top and the License dialog open, showing Free trial, 5 days left, the License key box, and the Unlock button.

#	Control	What it does
1	Trial bar	Shows the days left in your free trial. Click **Enter a license key** in it to open the dialog.
2	**License key** box	Paste your key here. Spaces and line breaks around it are ignored.
3	**Unlock** button	Checks the key. If it is accepted, the dialog shows who the license is for, and the trial bar disappears.

If the key is not accepted, a short message under the box says why. The usual cause is a missing or extra character from copying, so paste it again from the original email.

When the 7 days are over and no key has been entered, Statversity opens to a single screen instead of Home.

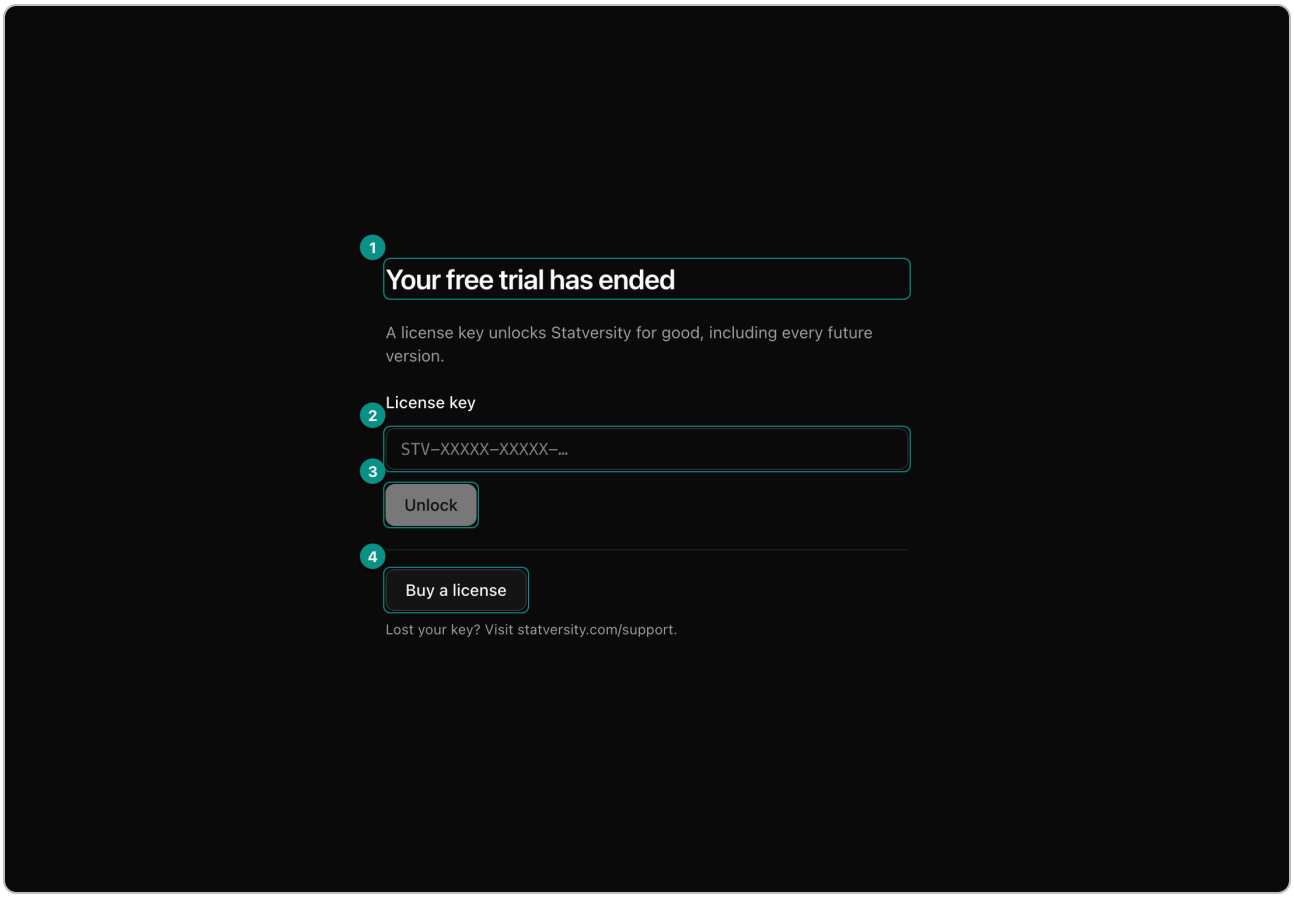


Figure 3. The screen shown after the trial ends, titled Your free trial has ended, with a License key box, Unlock, Buy a license, and a note about lost keys.

#	Control	What it does
1	**Your free trial has ended**	Tells you the trial is over. Your projects and files are untouched; they are just locked behind this screen until a key is entered.
2	**License key** box	Paste your key here.
3	**Unlock** button	Checks the key and opens Statversity.
4	**Buy a license**	Opens statversity.com/pricing in your web browser.

Good to know:

- **Nothing is sent.** Statversity checks the key inside the app, using the key itself. It does not contact any server, and it never sends your data or your key anywhere.
- **One key covers every future version.** You do not need a new key when you update.
- **Departments and labs** get one shared key. It carries a seat count (how many people it covers) and the department's name, which the License dialog shows.
- **Lost your key?** Go to statversity.com/support. To buy one, go to statversity.com/pricing.

- To see your license later, click **License…** in the Home footer. It shows who it is licensed to and the key id.

CHAPTER 3

Your first five minutes

In about five minutes you will open a practice dataset that comes with Statversity, answer three plain-language questions, and read Statversity's verdict on one real question: did scores go up after a training program? You will not need to know any statistics, and you will not break anything, because the data are made up and your original files are never changed. This chapter is a fast preview. Chapter 11 walks through the same analysis slowly and explains every screen, so whenever you want to know why something happened, that is the place to look.

The question

A group of 60 people answered a ten-question survey before a training program, and 60 answers were collected again afterward. The question is: **did the average score go up?**

Steps

1. Start Statversity. The Home screen says **Welcome to Statversity**. It offers four cards: **New project**, **Open project**, **Practice data**, and **Plan a study**.

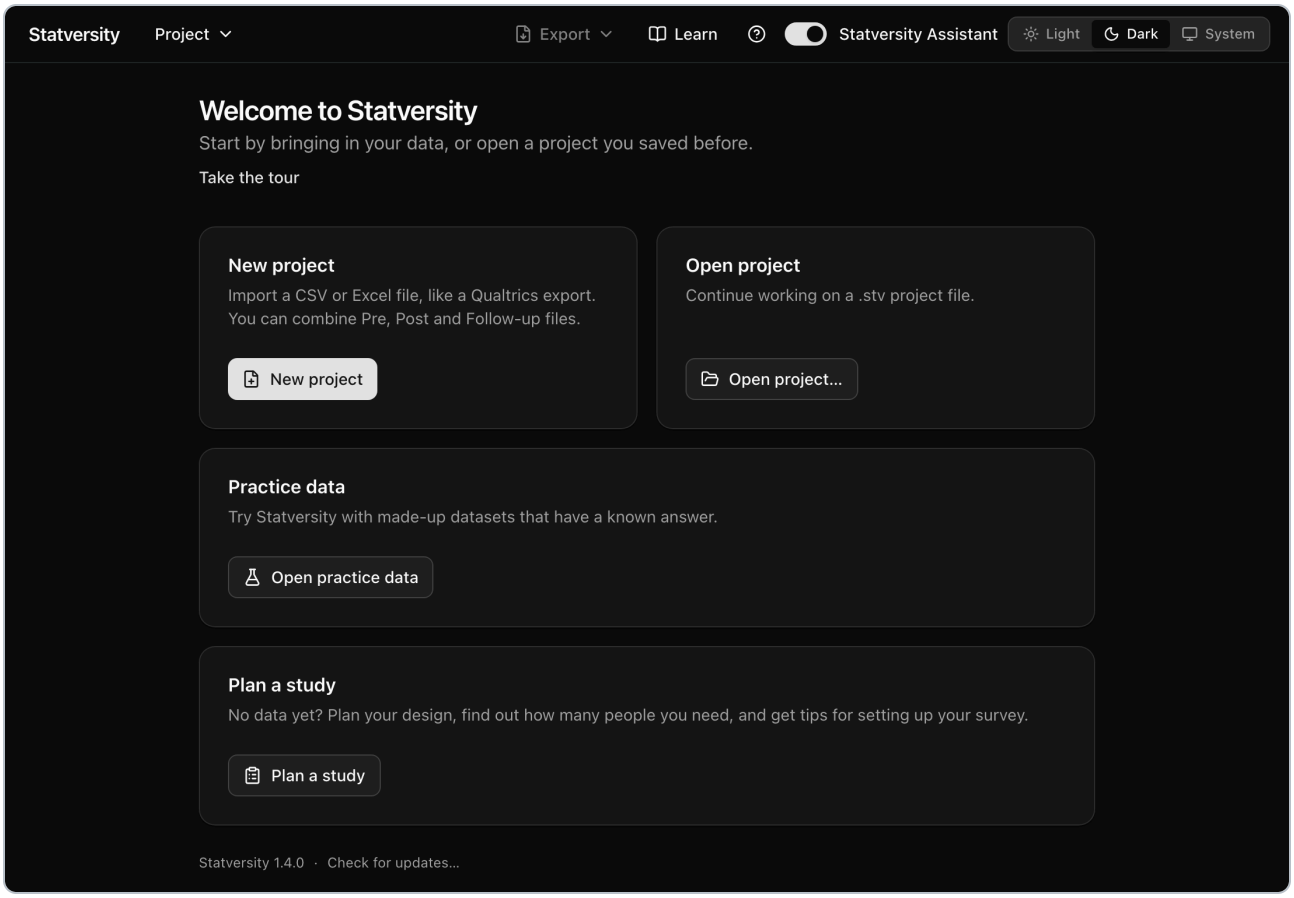


Figure 4. The Home screen, Welcome to Statversity, with the Open practice data button.

2. On the **Practice data** card, click **Open practice data**. A list of practice datasets appears.
3. Find the card titled **Did the training help?** and click **Import this dataset** on it. The screen shows **Choose files**, with `pre.csv` and `post.csv` already listed. Statversity loaded both for you.

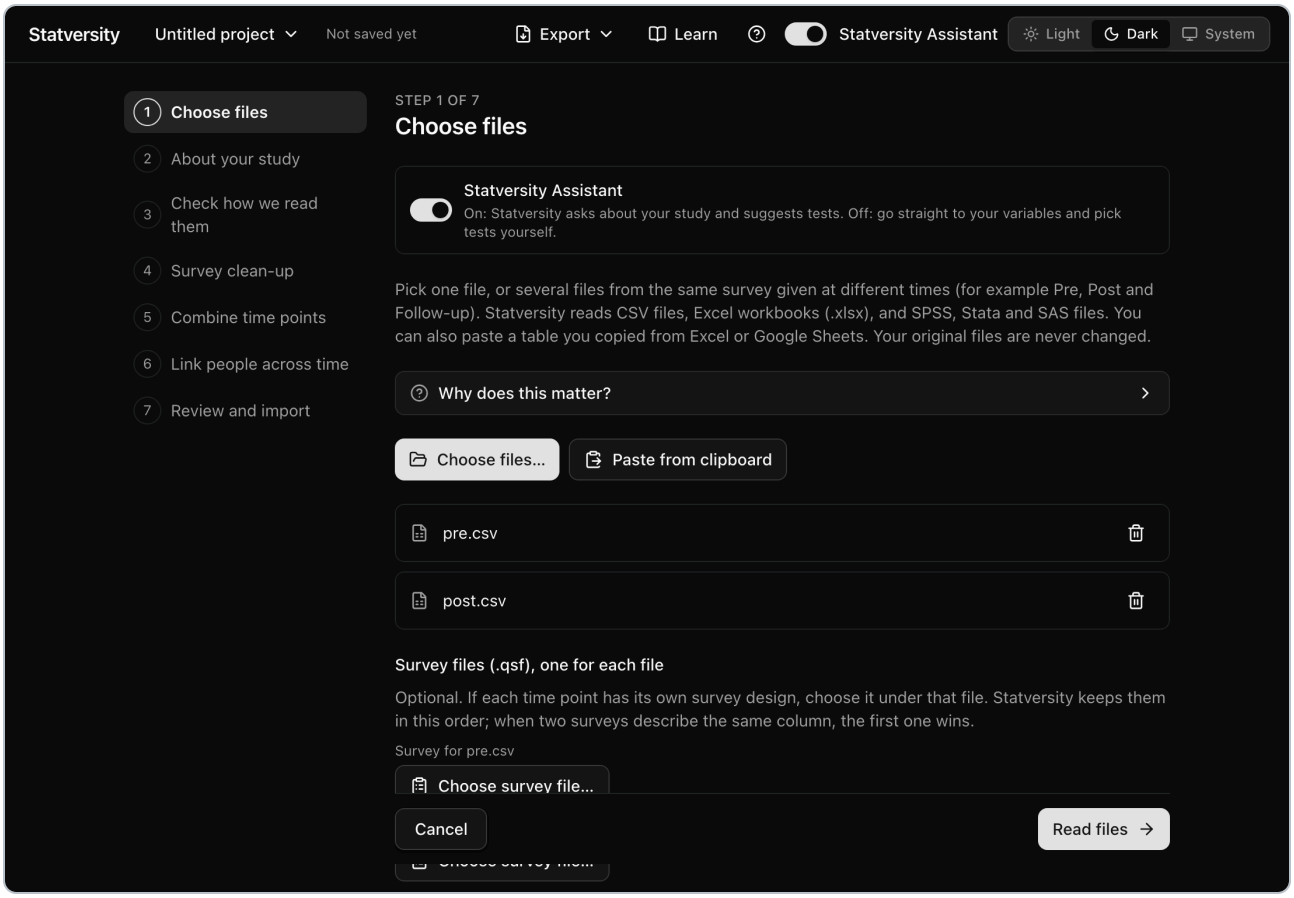


Figure 5. Step 1 of 7, Choose files, showing pre.csv and post.csv.

4. Click **Read files**. The screen shows **About your study**. Click **Skip these questions**. (You will answer the questions that matter later.)

5. Press **Continue** on each remaining import step: **Check how we read them**, **Survey clean-up**, **Combine time points**, and **Link people across time**. Statversity has already picked sensible settings, and chapter 11 explains what each one does. Leave **Don't link (default)** selected on the last of these, because these two files have no shared ID that could match one person's "before" to their "after."

6. The screen shows **Review and import**. This is Statversity's summary of what it is about to do: 120 rows read, 27 columns kept, and the time points labeled Pre and Post. Click **Import**.

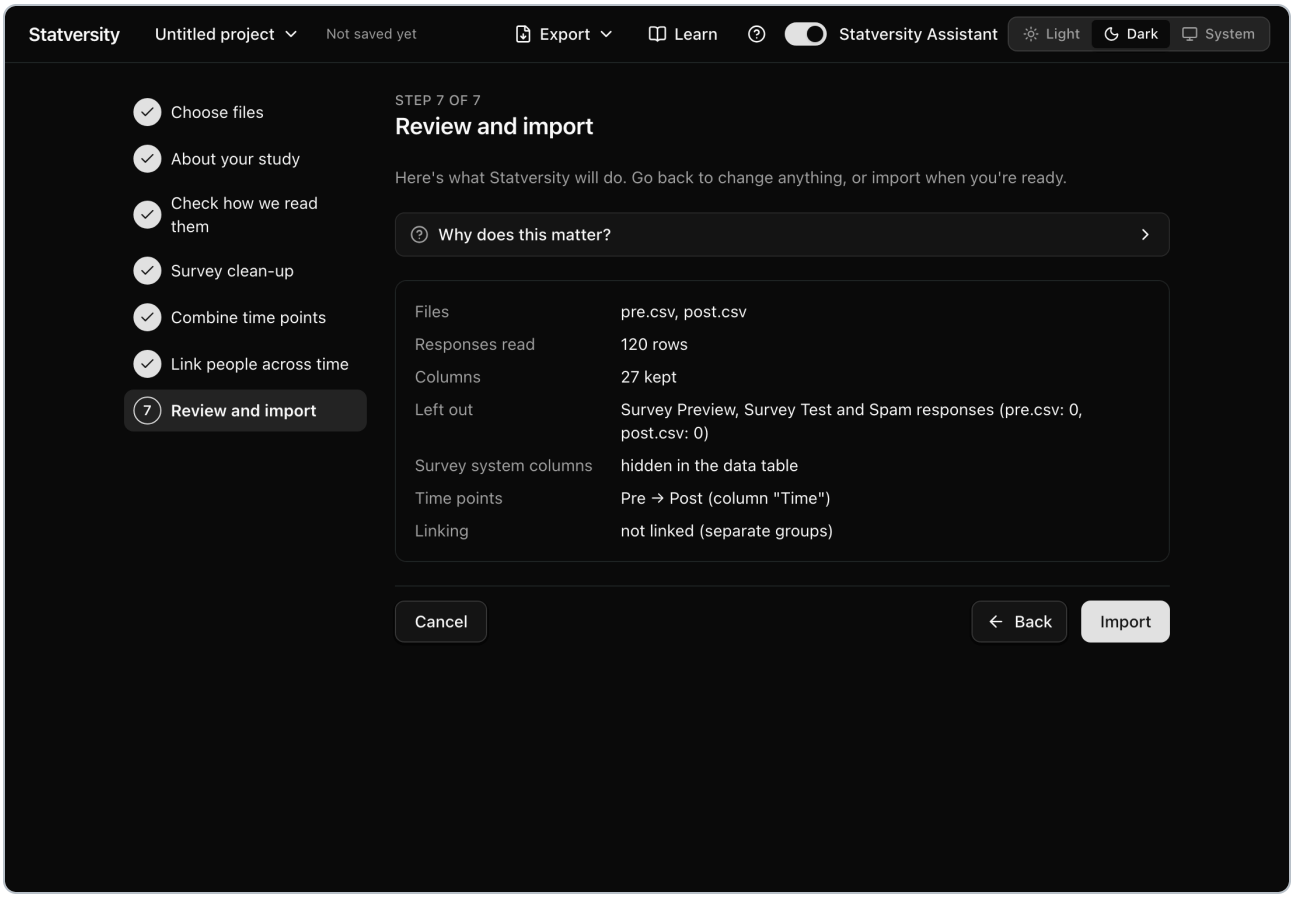


Figure 6. Step 7 of 7, Review and import, listing 120 rows and 27 columns.

7. The screen shows **Welcome**, the start of a short set of questions about what each survey item is. Click **Start**, then press **Continue** on each screen that follows (**What is each question?**, **Kinds of answers**, **Answer choices**, **Scales**, **Scale scores**, and **New groups**). On the final screen, **Summary**, click **Finish**. The screen shows **Variables**, your data's list of columns. Statversity has already combined the ten survey questions into one average score called **Q3 (average score)**.

8. Click **Analyze** in the row of tabs near the top. The **Test Advisor** appears. Its job is to pick the right statistical test for you, so you do not have to know the names of tests. Under **Which score or answer do you want to analyse?**, make sure **Q3 (average score)** is chosen. If it is not, pick it from the list.

9. Under **What do you want to know?**, choose **Did something change, or are some people different from others?**, then click **Continue**.

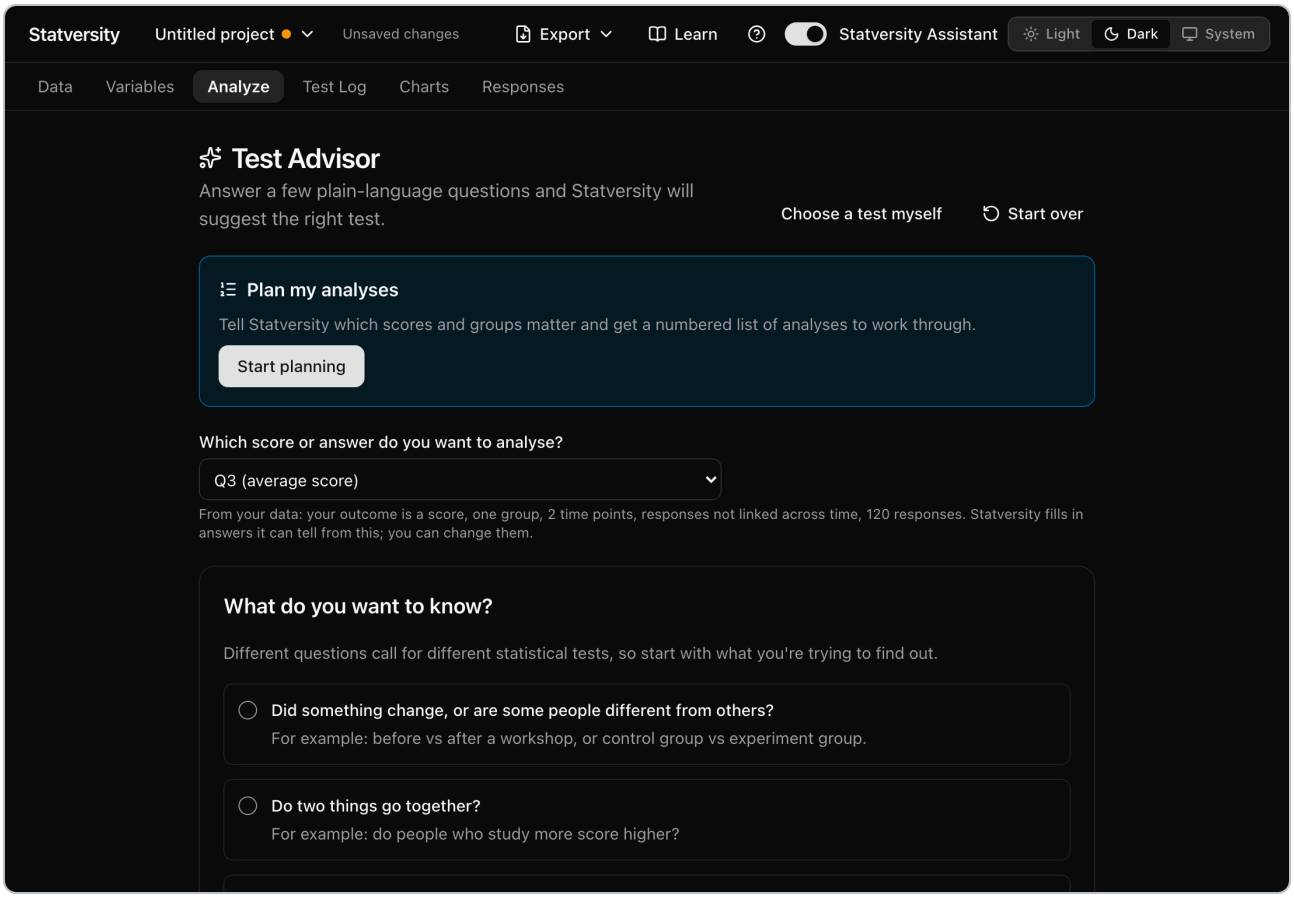


Figure 7. The Test Advisor with Q3 (average score) chosen and the first question, What do you want to know?, below it.

10. The next question is **What are you comparing?** Choose **The same group at different times, but I can't match each person at each time point.** Click **Continue**. This is the honest description of your files: the same group, measured twice, with no way to pair up individuals. Statversity fills in the other answers (the result is a score, and there are two time points) from your data.

11. The screen shows **Statversity recommends: Independent-samples t test**. A t test is a standard way to ask whether the averages of two groups differ by more than chance would explain. The orange note explains why Statversity chose this one: because the answers cannot be matched to the same people, Statversity treats Pre and Post as two separate groups. That is the honest choice, though less powerful than matching. Click **Set up this analysis**.

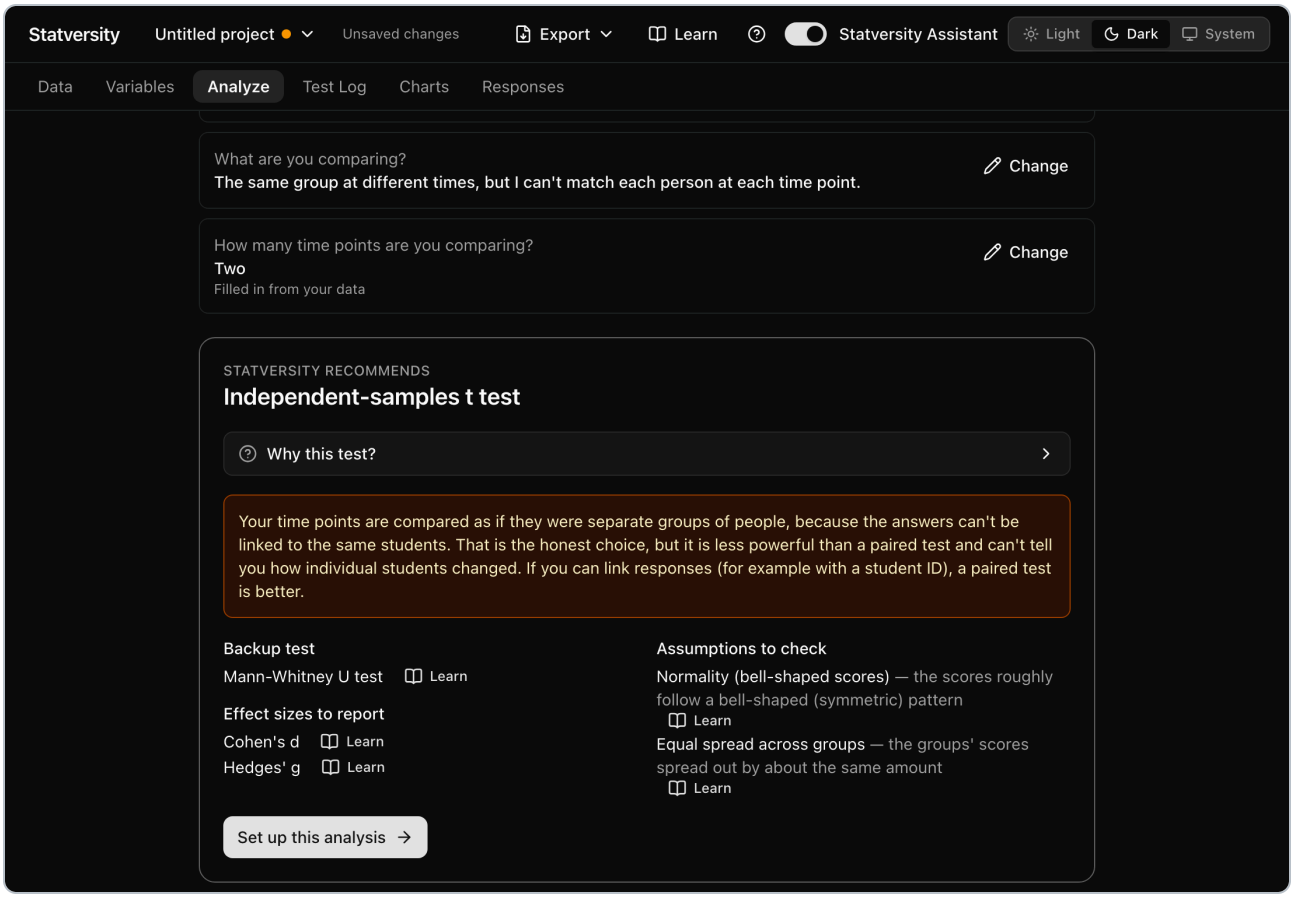


Figure 8. Statversity recommends an Independent-samples t test, with the reasons, the backup test, and the Set up this analysis button.

12. The screen shows **Choose your variables**. Statversity has already filled in **Outcome (scores)** with **Q3 (average score)** and **Groups to compare** with **Time point**. Do not change anything. Click **Check the assumptions**.

13. Statversity now checks whether your data meet the conditions the test depends on. Each check appears one at a time, and all three here say **Looks fine**. Click **Next check** twice to see them all, then click **Decide which test to use**.

14. The screen shows **Which test should you use?** and confirms the Independent-samples t test. Click **Run and see the results**.

Your result

The result opens with a box titled **What this means**. This is the part to read first, and for a beginner it is the only part you need.

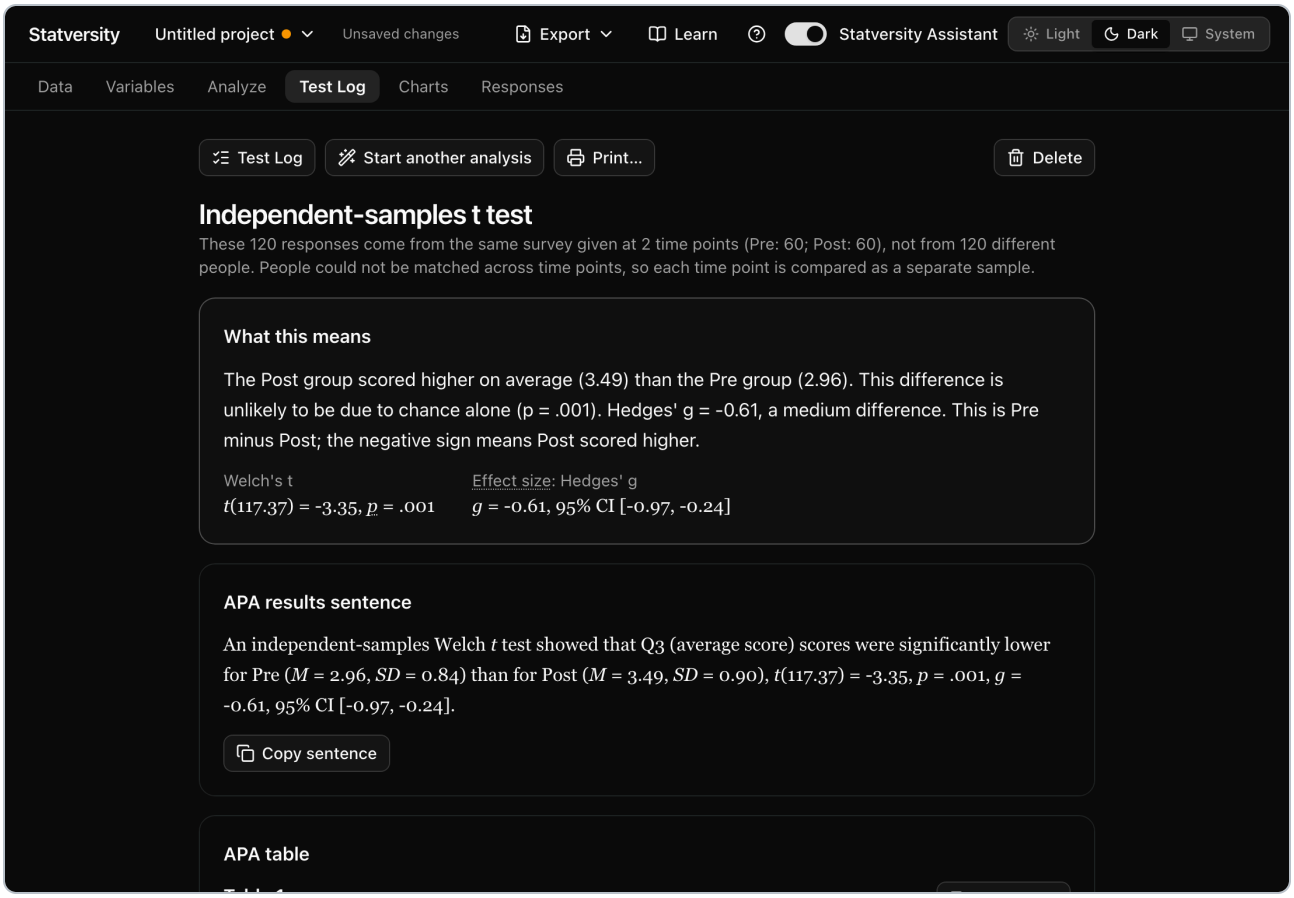


Figure 9. The result, with the What this means box at the top and the APA sentence and table below it.

The box reads: “The Post group scored higher on average (3.49) than the Pre group (2.96). This difference is unlikely to be due to chance alone (p = .001). Hedges’ g = -0.61, a medium difference. This is Pre minus Post; the negative sign means Post scored higher.”

Underneath, in smaller type, are the exact numbers: **Welch’s t**, t(117.37) = -3.35, p = .001, and **Effect size: Hedges’ g**, g = -0.61, with a 95% confidence interval of [-0.97, -0.24]. You will meet these again in the paper-ready sentence and table just below the box.

In plain English: the average score rose from 2.96 before the training to 3.49 after it, and that rise is too large to be written off as luck. It is also a medium-sized change, not a trivial one. You would still want to be careful about saying the training *caused* it, because other things may have changed between the two surveys.

IN PLAIN LANGUAGE

What do "p = .001" and "medium effect" mean? The p-value answers one narrow question: if the training did nothing at all, how often would you see a gap this big just by chance? About once in a thousand times, so chance is a poor explanation. The effect size answers a different question, how big the gap is: Hedges' g of 0.61 (ignore the minus sign, it only says which group came first) is a medium-sized difference by common rules of thumb. A small p-value tells you the difference is probably real, and the effect size tells you whether it is big enough to care about. Always read both.

What to do next

You have just done a complete analysis: import, questions, test choice, assumptions, and result. Chapter 11 repeats it screen by screen, explaining each choice. Chapter 23 explains how to read p-values, effect sizes, and confidence intervals when you are new to statistics, including what "significant" does and does not mean.

SECTION 2

Screen by screen

CHAPTER 4

The top bar and project basics

The top bar is the one part of Statversity that never goes away. It is on every screen, and it holds the controls that belong to the whole project rather than to one screen: the project name, saving, exporting, the Learn library, the tour, the Statversity Assistant switch and the color theme. Under it, once you have loaded data, a second row of tabs moves you between the main areas. This chapter describes both rows, the project menu, the saving and quitting dialogs, the welcome tour and the keyboard shortcuts.

The top bar

The screenshot below shows the top bar on a wide window with a project open and unsaved changes. The table under it explains each numbered badge.

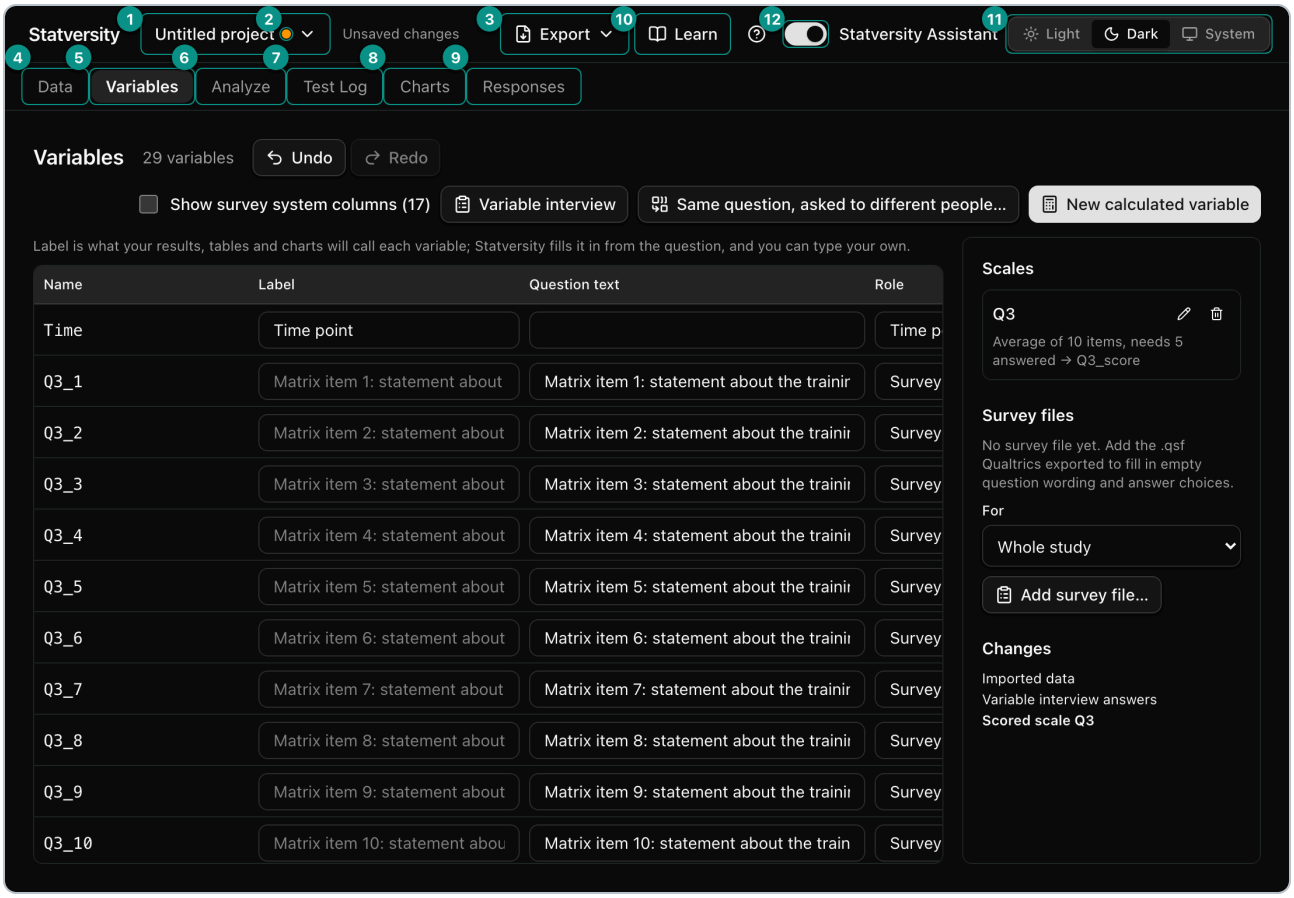


Figure 10. The top bar with numbered badges on each control.

Number	Control	What it does
1	Project name (menu button)	Shows the name of the open project, or **Project** if none is open. Click it to open the project menu (described below).
2	Unsaved changes dot (amber)	Appears beside the project name when you have work that is not saved to a file yet. It clears when you save.
3	**Export**	Opens the export menu: **Report…**, **Print report…**, **Data…**, **Codebook…** and **Test Log…**. It is disabled until a project exists.
4	**Data**	Opens the data table.
5	**Variables**	Opens the variable table.
6	**Analyze**	Opens the analysis area.
7	**Test Log**	Opens the list of finished tests.
8	**Charts**	Opens the chart area.
9	**Responses**	Opens the reader for written answers.
10	**Learn**	Opens the Learn library of plain-language guides.
11	**Theme** (**Light**, **Dark**, **System**)	Chooses the color theme. **System** follows your computer.
12	**Statversity Assistant** (switch)	Turns the Assistant on or off.

> **NOTE**
> The tab row (items 4 to 9) is part of the second row, not the top bar proper. It is drawn on the same screenshot because you use it together with the top bar. It appears only when a dataset is loaded, and it is hidden on Home, in the import wizard, and in Learn.

The controls that have no badge

A few controls are not numbered in the screenshot but you will use them.

- **Statversity** (the name at the far left). This is a button, also called **Home**. It takes you to the Home screen. It asks nothing, and your project stays open.
- **Autosave text.** Beside the project menu, a small line of text reports the save state. It says one of **All changes saved**, **Not saved yet**, **Saving a backup…**, **Backup failed. Save your project soon.** (followed by the reason), **Unsaved changes · backup at** a time, or **Unsaved changes**. You cannot click it. Statversity keeps automatic backups while you work, so a crash does not lose everything, but a backup is not the same as saving your project. Save it yourself.
- **Take the tour** (a question-mark icon). Starts the welcome tour (see below).

The Statversity Assistant switch

The **Statversity Assistant** switch decides how much Statversity guides you. When you hover over it, the tooltip explains: on means Statversity asks about your study and suggests tests; off means you go straight to your variables and pick tests yourself.

Turning it on or off changes the import wizard (the **About your study** step appears only when it is on), the Analyze tab (Test Advisor when on, **All tests** when off), and several hints. If you are sitting on the Advisor when you flip it, Statversity swaps you to **All tests**, and the other way round. Your data and saved results are untouched. This guide assumes the switch is on, and chapter 10 covers working with it off.

> **IN PLAIN LANGUAGE**
> Think of the Assistant as a patient colleague sitting next to you. On, they ask questions before you do anything. Off, they leave you alone and assume you know what you are doing.

The theme switch

Choose **Light**, **Dark**, or **System**. **System** follows your computer's own appearance setting. The whole app recolors at once, and Statversity remembers your choice. It changes colors only, not layout or results.

Export

Export opens a menu of five items. Each opens a dialog. They are covered in chapter 9, but in short: **Report…** builds a Word or PDF report from the tests you choose, **Print report…** opens the same dialog ready to print, **Data…** saves your data as Excel, CSV, SPSS or Stata, **Codebook…** saves a description of your variables, and **Test Log…** saves the list of your tests. If nothing is logged yet the report dialog says to log a test first.

The More actions menu on narrow windows

When the window is narrower than about 1,024 pixels, the right-hand group collapses into a single three-dots button called **More actions**. The **Export** button itself disappears below about 768 pixels, and its items move into the same menu. The screenshot below shows it open in a window 900 pixels wide.

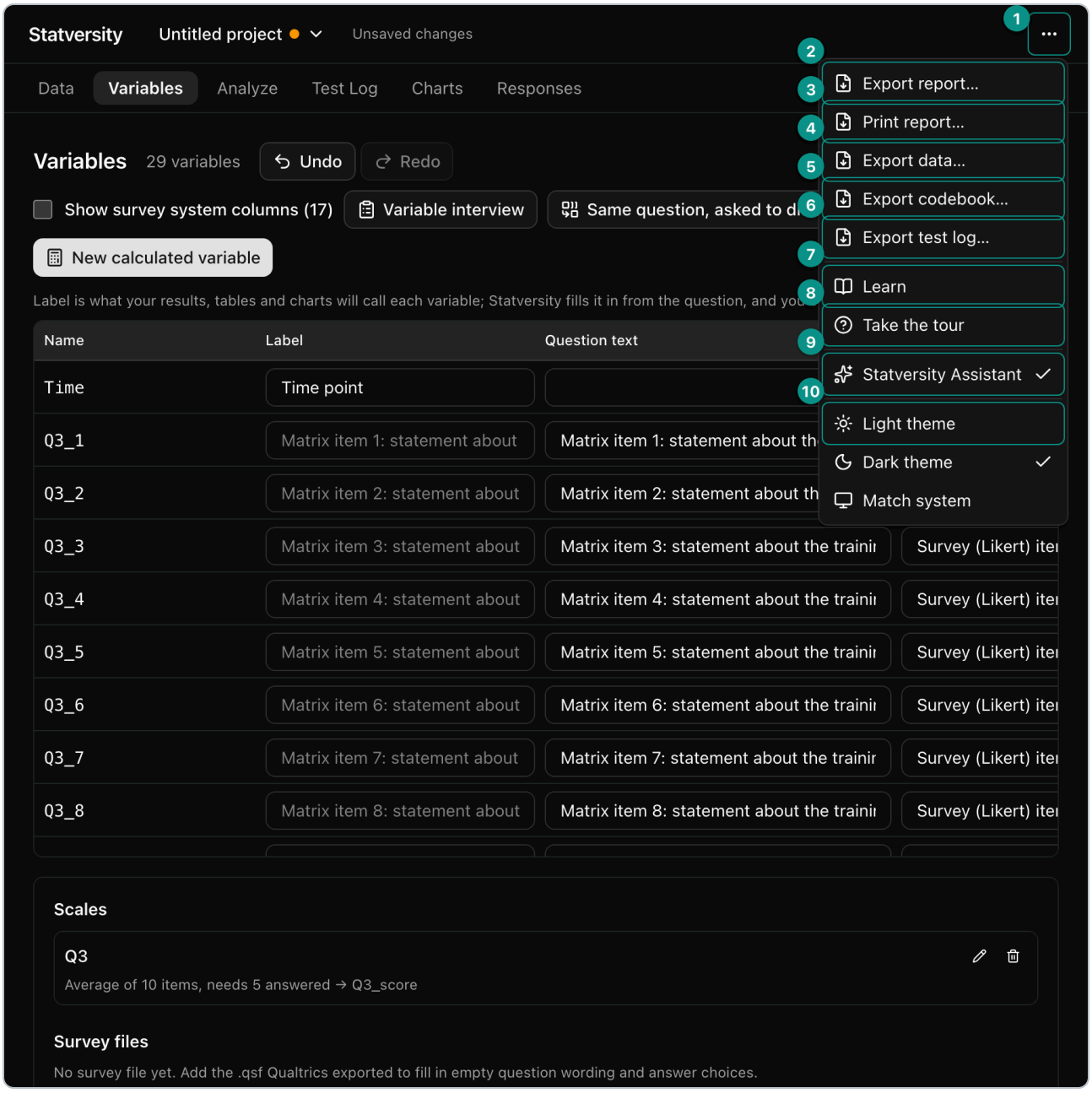


Figure 11. The More actions menu in a narrow window.

Number	Control	What it does
1	**More actions** (three dots)	Opens this menu.
2	**Export report…**	Opens the report dialog in export mode. Disabled without a project.
3	**Print report…**	Opens the same dialog in print mode.
4	**Export data…**	Opens the data export dialog.
5	**Export codebook…**	Opens the codebook export dialog.
6	**Export test log…**	Opens the Test Log export dialog.
7	**Learn**	Opens the Learn library.
8	**Take the tour**	Starts the welcome tour.
9	**Statversity Assistant**	Toggles the Assistant. A tick shows when it is on.
10	**Light theme**, **Dark theme**, **Match system**	Picks the theme. A tick marks the current one.

> **TIP**
> If you cannot find a button you saw in this guide, widen the window first. It may simply have moved into **More actions**.

The project menu

Click the project name to open the project menu. A project is one file with the extension `.stv`. It holds your data, your variable settings, your tests, your charts and your plans, so you can close Statversity and pick up later exactly where you left off.

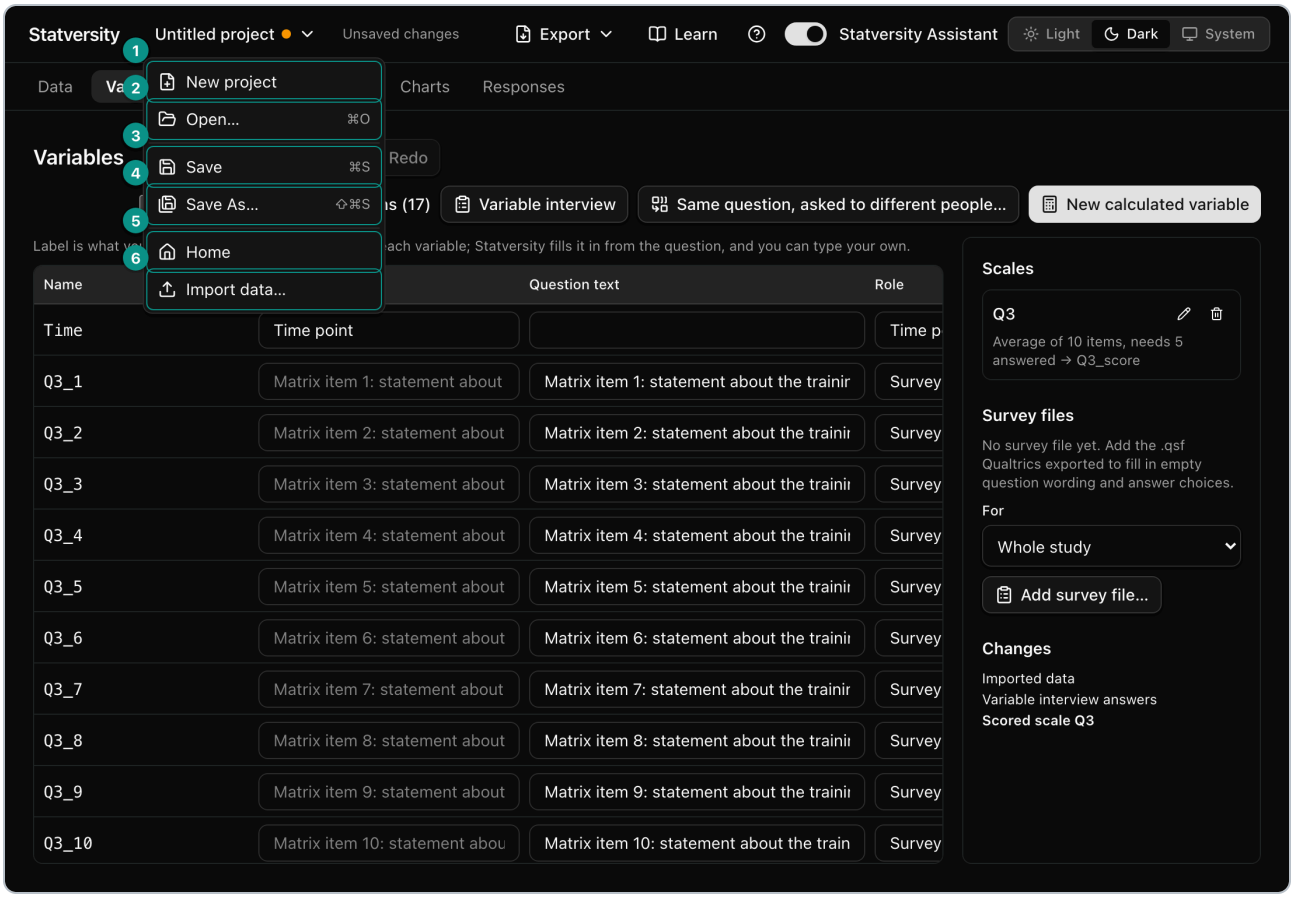


Figure 12. The project menu open.

Number	Control	What it does
1	**New project**	Starts a new, empty project and opens the import wizard. If you have unsaved work it asks about it first.
2	**Open…**	Opens a `.stv` project file (older `.statly` files still open). Shortcut: Cmd+O (Ctrl+O on Windows).
3	**Save**	Saves to the current file, or asks for a name the first time. Shortcut: Cmd+S. Disabled without a project.
4	**Save As…**	Saves a copy under a new name. Shortcut: Shift+Cmd+S. Disabled without a project.
5	**Home**	Goes to the Home screen.
6	**Import data…**	Opens the import wizard.

Importing when a project already has data

If you choose **Import data…** while the open project already has data, Statversity asks **Import into which project?** It says: “This project already has data. You can add new rows from another export of the same data, replace everything, or start a new project.” The buttons are:

- **Cancel**, which changes nothing.
- **Add new responses**, which opens **Combine data** on the Data tab with the **Add new responses** job already chosen. Your variable setup stays as it is.
- **Replace this project’s data**, which is permanent.
- **Start a new project**, the default.

> **CAUTION**
> Replacing can’t be undone: the data in this project and all of its variable setup (types, labels, scale groups, survey flow, tags) will be permanently discarded. To add rows from a new export of the same data, choose **Add new responses** instead.

If you choose **Replace this project’s data** and the new file contains every column of the current data, the import wizard stops you with a notice: “This file has the same columns as the data already in this project. It looks like a new export of the same data.” Its buttons are **Add new responses instead** and **Replace anyway**.

Saving, and the save-changes dialog

Press **Save** the first time and Statversity shows a save dialog where you name the file and choose a folder. After that, **Save** writes to the same file and the amber dot disappears. **Save As…** writes a copy and carries on with the new file.

When you try to leave unsaved work (by starting a new project, opening another, or quitting), Statversity asks “Save changes to [your project name]?” The buttons are **Cancel**, **Don’t save**, and **Save**. When you quit, the wording is **Do you want to save changes to** your project name **before quitting?** and the middle button reads **Don’t Save**. This also appears if you close the window, press Cmd+Q, choose Statversity > Quit, or use the Dock’s Quit.

If Statversity closed unexpectedly

If Statversity closed without saving, the Home screen shows a **Recover unsaved work** card. It lists automatic backups, named either **Untitled project (never saved)** or by file name, with **Backup from** a date and time. Each row has **Recover**, which re-opens that backup, and **Discard**, which deletes it.

The dataset tabs

Once a dataset is loaded, a second row appears under the top bar. The tabs move you between the main areas of the project. They are hidden on Home, in the import wizard and in Learn.

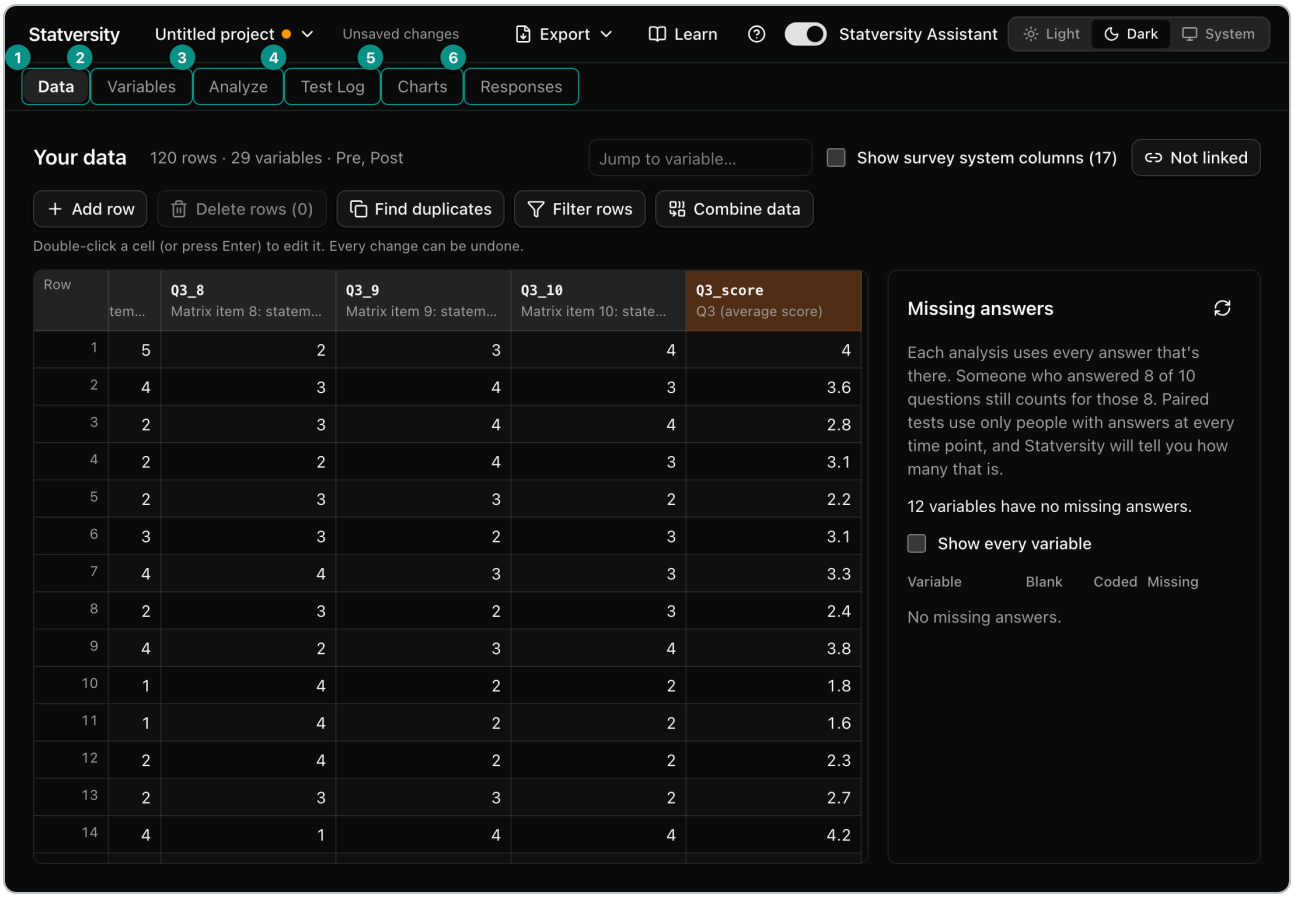


Figure 13. The row of tabs under the top bar, with data loaded.

The screenshot shows a single dataset (the one-group pretest and posttest practice data), so there is one tab per area and no per-dataset tabs or add button.

Number	Control	What it does
1	**Data**	Opens the data table.
2	**Variables**	Opens the variable table. It also stays highlighted while the Variable Interview is open.
3	**Analyze**	Opens the analysis area. With no saved plan, or while you are answering Advisor questions, it opens the Test Advisor (Assistant on) or **All tests** (Assistant off). With a plan for this dataset it opens the plan page. It is also highlighted while you work through roles, assumptions, plans and the test menu.
4	**Test Log**	Opens the list of finished tests. It is also highlighted while you view a result.
5	**Charts**	Opens the chart list.
6	**Responses**	Opens the reader for written answers.

The welcome tour

The first time Statversity runs, a six-step tour opens by itself. Afterward you can start it again from **Take the tour** in the top bar, in the **More actions** menu, or on the Home screen. The steps are:

1. **Welcome**: Welcome to Statversity.
2. **Import**: Import your data.
3. **Variable Interview**: Tell Statversity about your variables.
4. **Test Advisor**: Get a recommended test.
5. **Results & Test Log**: Review your results.
6. **Learn & Study Planner**: Learn as you go.

Three buttons control it. **Skip** closes the tour, and it will not reopen by itself. **Back** goes to the previous step (it is absent on step 1). **Next** moves ahead, and on the last step it reads **Done**.

Keyboard shortcuts

These work from anywhere. On Windows and Linux use Ctrl where the table says Cmd.

Shortcut	What it does
Cmd+S	Saves the project. Same as **Save**.
Shift+Cmd+S	Saves under a new name. Same as **Save As…**.
Cmd+O	Opens a project file. Same as **Open…**.
Cmd+Z	Undoes the last data or variable edit. Works only on the Data and Variables screens, and not while you are typing in a text box, which keeps its own undo.
Shift+Cmd+Z or Cmd+Y	Redoes. Same restriction.
Closing the window, Cmd+Q, Statversity > Quit, Dock Quit	Starts the save-before-quit check described above.

For keyboard users, a hidden **Skip to content** link appears when you first press Tab. It jumps past the header to the main area.

Messages and notifications

Short messages appear at the bottom right after an action. Success and information messages disappear after five seconds. Error messages stay until you click the **Dismiss** button, and some include a **Details** expander with more information.

When Statversity starts, it shows **Warming up the statistics engine…** and **This usually takes a few seconds.** If the engine fails to start, the screen says **The statistics engine didn’t start**, offers a **Technical details** expander, and gives a **Retry** button that tries again.

CHAPTER 5

Getting data in

Every analysis starts with getting your survey data into Statversity. This chapter covers the Home screen, every step of the import wizard, the file formats Statversity can read, and the tools for combining several files: combining time points, merging question versions, and matching questions by wording.

Home

Home is where Statversity opens. You can also get back to it by clicking **Statversity** in the top bar or by choosing **Home** in the project menu.

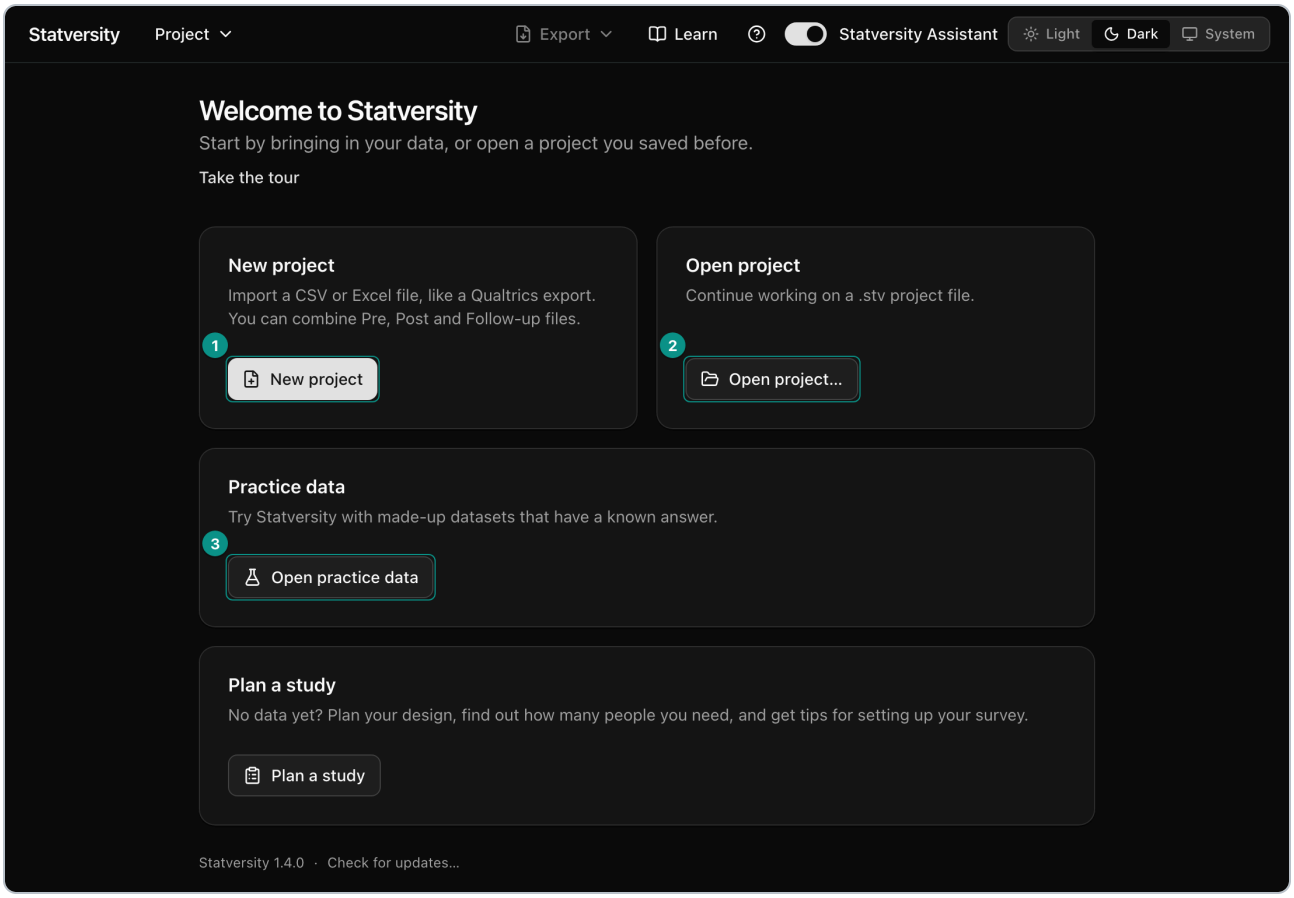


Figure 14. The Home screen with numbered badges.

Number	Control	What it does
1	**New project**	Starts a project and opens the import wizard at the **Choose files** step.
2	**Open project…**	Opens a file picker for a `.stv` project file (older `.statly` files still open). The project loads and usually lands on the Data screen.
3	**Open practice data**	Opens the Practice data screen, where you can import a sample dataset to learn on.

A **Recover unsaved work** card appears on Home only after a crash or an unsaved close. It lists automatic backups. **Recover** re-opens one; **Discard** deletes it. It is not in the screenshot because nothing needs recovering.

Two more things on Home are not numbered. **Take the tour** reopens the welcome tour. The **Plan a study** card, **Plan a study**, is for people with no data yet; it opens the Study Planner, which works out how many participants you need.

> **TIP**
> New to Statversity? Click **Open practice data**, pick a dataset, and click **Import this dataset**. You will have realistic data to try everything in this guide on.

The import wizard

The wizard carries you from files on your disk to a clean dataset. You reach it from Home > **New project**, from the project menu > **Import data…**, from the Data screen's empty state button **Import a file**, or from Practice data > **Import this dataset**.

A list of steps appears down the left side (numbered circles in a narrow window). The heading says **Step** and a number **of** the total. The steps are:

1. **Choose files**
2. **About your study** (only with the Assistant on)
3. **Check how we read them**
4. **Survey clean-up**
5. **Combine time points** (only with two or more files)
6. **Link people across time** (only with two or more files)
7. **Review and import**

You can click a finished step in the list to go back to it. Future steps are greyed out. At the bottom of the wizard, **Cancel** (bottom left) abandons the import and takes you to the Data screen if you already have data, or Home if not. **Back** goes to the previous step. The main button changes with the step: **Read files**,

Next question, **Continue**, and finally **Import**. If the main button is disabled, a grey line of text beside it explains why.

Step 1: Choose files

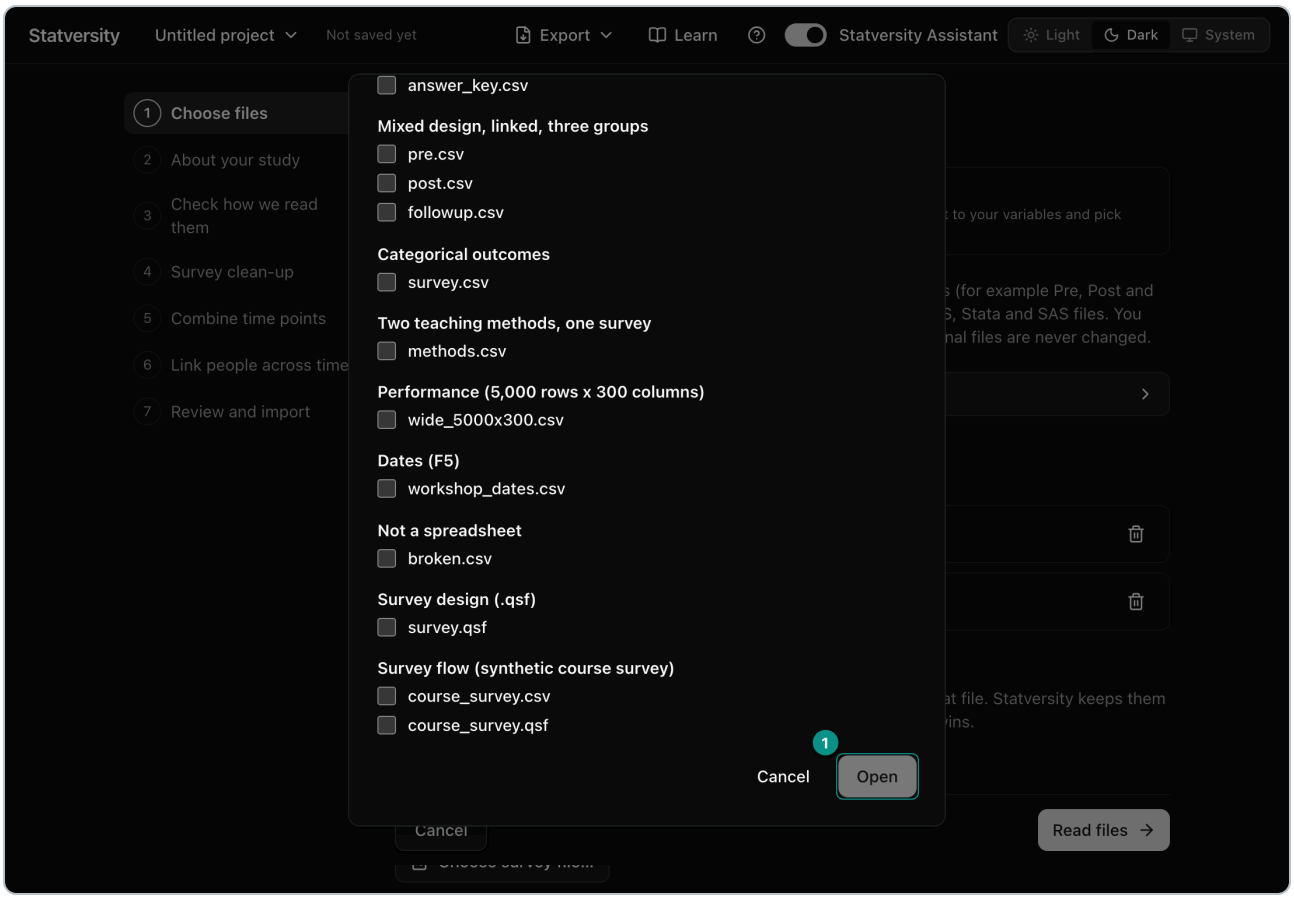


Figure 15. The Choose files step with the file picker open.

Number	Control	What it does
1	**Open** (in the file picker)	Confirms your choice in the picker and adds the files to the list. In the real app the picker is your operating system's own dialog; the screenshot shows a stand-in picker.

The picker covers the **Files to import** list in the screenshot. That list shows each chosen file with its name and a format badge such as **SPSS file** or **Stata file**. A small button **Remove** followed by the file name takes a file out of the list.

Other controls on this step:

- **Statversity Assistant** switch with a help line. It is the same switch as in the top bar, repeated here with an explanation. It adds or removes the **About your study** step.
- **Choose files…** opens a file picker. You can pick one or several CSV, Excel (.xlsx), SPSS, Stata or SAS files.

- **Paste from clipboard** imports a table you copied from Excel or Google Sheets. It adds a file called **Pasted data**. If nothing is on the clipboard, Statversity says to copy a table first.
- **Add your survey file (.qsf)** (a section that appears when one file is chosen) with **Choose survey file…**. A .qsf file is the design file Qualtrics can export for a survey. If you give it to Statversity, it fills in the full wording of each question and the names of the answer choices. It is optional. With several data files the section is called **Survey files (.qsf), one for each file**, with a **Survey for** file name row and **Choose survey file…** for each. Once chosen, the survey's name and question count appear with a trash button to remove it.
- **Qualtrics exports** is a drop-down with three choices: **Detect automatically (recommended)**, **These are Qualtrics exports**, and **These are not from Qualtrics**. It tells Statversity how to read the first rows of your file. Leave it on automatic unless Statversity guesses wrong on the next step.

Click **Read files** to move on. It is disabled until at least one file is chosen.

Step 2: About your study

This step appears only with the Assistant on. Statversity asks a few plain questions, one at a time, labelled **Question** number **of** total. Your answers are used later to make first guesses about how your variables and tests should be set up. They do not change your data. **Skip these questions** skips the rest and takes you to the next step. **Back** returns to the previous question and **Next question** moves on.

The questions are:

Question	Choices
What did you do?	**Everyone answered once**; **The same people answered twice: before and after something**; **The same people answered three or more times**; **Different groups answered once, and you compare them**; **Something else**
A question about groups (the wording changes with your design, for example **Do you want to compare groups of people who took the survey?**)	**No, look at everyone together** (or **No, everyone together**); **Yes, compare groups**
How many groups will you compare?	**2 groups**; **3 or more**; **Not sure yet**
What were the groups called?	An optional text box (for example Lecture, Workshop)
Can you match each person's answers across times?	**Yes, they entered an ID or code**; **No, answers are anonymous**; **Not sure**
What kinds of questions are in your survey?	Tick any: **Knowledge or test questions with right answers**; **Agree-disagree or rating questions**; **Background questions (year, age, program)**; **Written answers in their own words**
Did everyone see the same test questions?	**Yes**; **No, each person got a random set from a bigger pool**; **Not sure**

> **IN PLAIN LANGUAGE**
> “Matching each person’s answers across times” matters because a before-and-after study is much more powerful when you can compare each person with themselves. If the survey was anonymous, Statversity can still compare the groups as wholes, but it cannot pair them.

The screenshot below shows the first question.

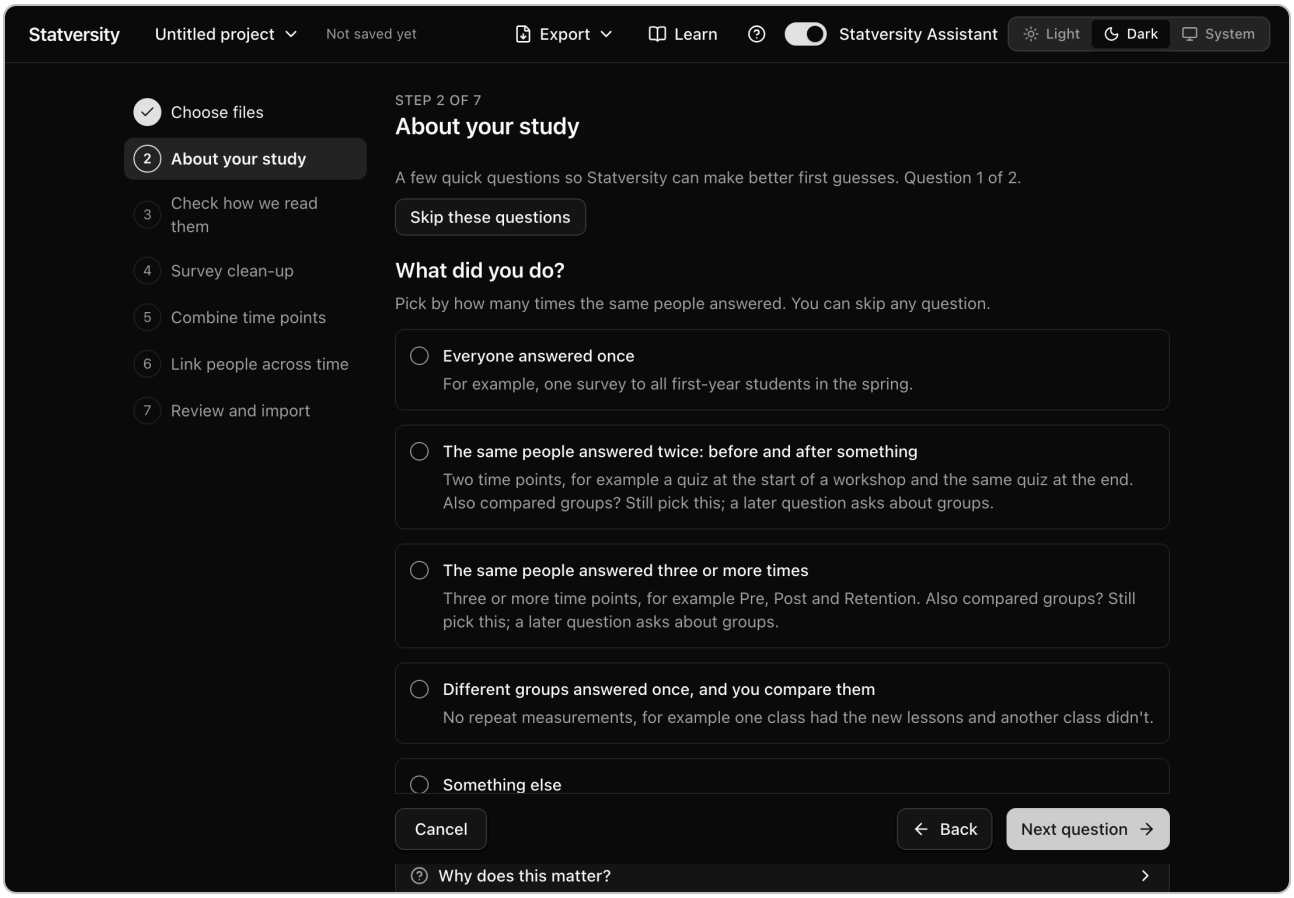


Figure 16. The About your study step.

Step 3: Check how we read them

Statversity now reads each file and shows you what it found. This step is mostly for checking; if everything looks right, click **Continue**.

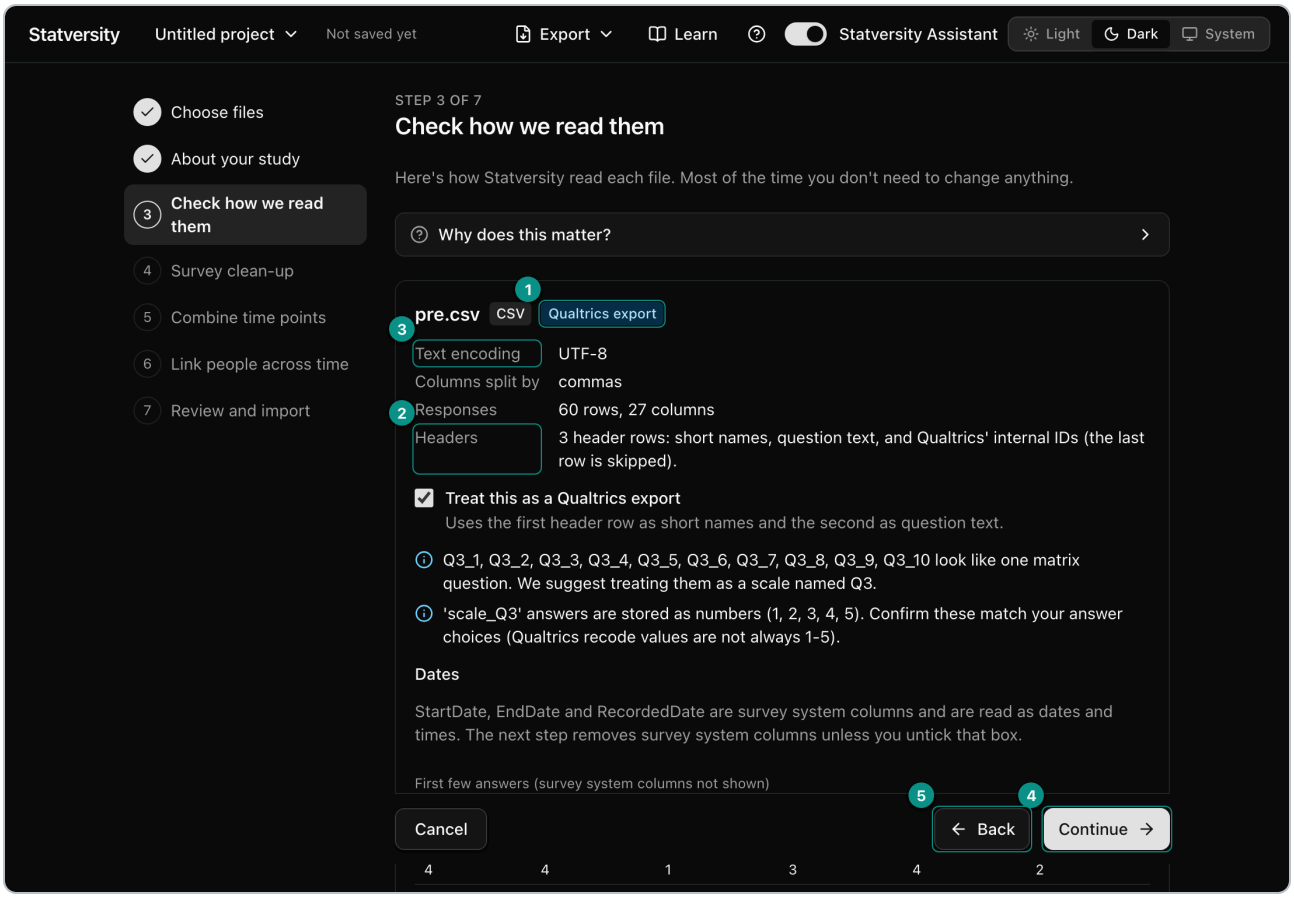


Figure 17. The Check how we read them step.

The screenshot shows a pretest file, pre.csv, from a Qualtrics export.

Number	Control	What it does
1	File type badge on the file card (and the **Qualtrics export** badge)	Shows the format detected: CSV, Excel, SPSS and so on, and whether the file looks like a Qualtrics export.
2	**Headers** line (and the **Treat this as a Qualtrics export** checkbox)	Shows how the header rows were read. When the checkbox is ticked, the first header row is read as short names and the second as the question text.
3	**Text encoding** line	Shows how the characters in the file were decoded. Wrong encodings show up as odd symbols in accented letters.
4	**Continue**	Accepts the reading and moves on.
5	**Back**	Goes to the previous step.

Each file card lists **Read from**, **Text encoding**, **Columns split by**, **Responses** (the count of rows) and **Headers**. Other controls on this step:

- **Which sheet holds the responses?** appears for Excel files with several sheets. Pick the right worksheet and Statversity re-reads the file.
- **Dates** appears when a column looks like dates. For each such column, if the day and month are ambiguous you choose **Month first** or **Day first**. A **Keep as text** checkbox leaves the column as it is instead of reading it as dates.
- **First few answers (survey system columns not shown)** is a read-only preview table of your data.
- **Answer labels only** appears when you gave a pair of files, one with numbers and one with words. The words file is used only to name the answer choices.

Step 4: Survey clean-up

This step handles quirks that survey tools add. Sections appear only when they apply to your file. If nothing applies, Statversity says **Nothing to clean up in this file. Continue to review your import.**

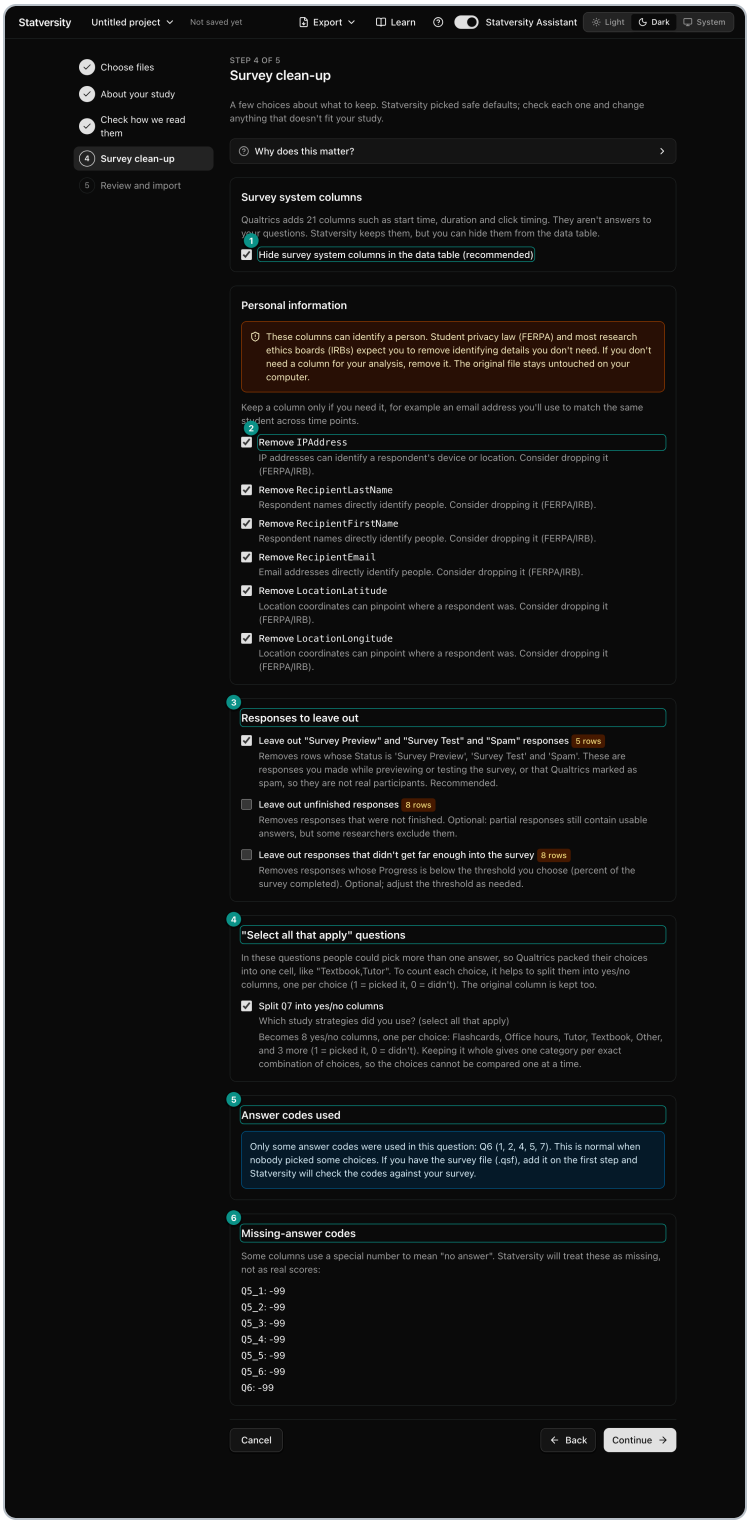


Figure 18. The Survey clean-up step.

The screenshot comes from a messy Qualtrics export (messy_3header.csv), so every clean-up option appears. It is a tall capture because this step scrolls. With only one file the wizard has five steps, so the step list reads “Step 4 of 5”.

Number	Control	What it does
1	**Survey system columns** checkbox: **Hide survey system columns in the data table (recommended)**	Hides the survey tool's bookkeeping columns.
2	**Personal information** checkboxes: **Remove** column name	Removes columns that look like names, emails or IP addresses.
3	**Responses to leave out** checkboxes and the progress box	Drops test, unfinished and too-short responses.
4	**"Select all that apply" questions**: **Split** column **into yes/no columns**	Turns a multiple-answer question into one yes/no column per choice.
5	**Answer codes used**	A read-only note listing which codes were used in a question when only some of the possible codes appear.
6	**Missing-answer codes**	A read-only list of columns where a special number such as -99 will be treated as "no answer".

Here is what each part means.

Survey system columns. Survey tools such as Qualtrics add columns that are about the survey process, not your questions: start and end times, an internal response ID, a progress percentage, the distribution channel, and so on. They are called system columns. The checkbox **Hide survey system columns in the data table (recommended)** keeps them in the file but hides them from the Data and Variables screens so they do not distract you. You can show them again later. **Default: ticked.** Leave it ticked unless you need those columns.

Personal information. Statversity spots columns that look like names, email addresses or IP addresses. Each gets a checkbox **Remove** and the column name. A ticked box means the column is not imported at all. If you do not need the column for your analysis, removing it makes your project safer to share. **Default: no change from what Statversity suggests;** read the list and decide.

Responses to leave out. Sometimes the file includes rows that are not real answers. The section offers checkboxes such as **Leave out** a value **responses** (for example, preview or test responses), **Leave out unfinished responses**, and **Leave out responses that didn't get far enough into the survey**, which shows a count. A number box, **Keep only responses with progress of at least**, with a percent sign, lets you set the cut-off. Rows you leave out are not imported. The count tells you how many that is, so you can tell whether the choice is reasonable.

"Select all that apply" questions. A question that lets people tick several answers is stored in one column, with the chosen choices run together (for example "Email, Phone"). A statistical test cannot use that directly. **Split** the column name **into yes/no columns** turns it into one column per choice, each

holding 1 if the person ticked that choice and 0 if not. After the split you can compare, say, the percent who ticked Email across groups. Statversity creates the new columns and keeps the original. **Default: ticked** when the question is detected.

Check the order of the answer choices. (Not in the screenshot, because it appears only when Statversity needs you to confirm an order.) Rating scales are stored as numbers, and the numbers matter: a 5 should mean "strongly agree" if higher should mean more agreement. For each answer choice you see a number box (the code), up and down buttons (**Move** the choice **up** and **down**), and a checkbox **This order and numbering are correct**. If the order is wrong, move choices or change the codes. You must confirm the order before you can continue.

> **CAUTION**
> If a scale runs the wrong way (1 = strongly agree), every average you compute from it will be reversed. Look at this list carefully.

Unusual answer codes. If the codes have gaps that Statversity cannot explain, it asks you to tick **I've checked these codes** before continuing (not in the screenshot). Two read-only notes are numbered 5 and 6 in the screenshot: **Answer codes used** lists the codes found in a question where only some codes appear (here Q6 used 1, 2, 4, 5 and 7), and **Missing-answer codes** lists columns where Statversity will treat a code such as -99 as missing.

> **IN PLAIN LANGUAGE**
> **Reverse-worded items** are survey questions that are worded the opposite way from the rest, for example "I often feel anxious" in a confidence scale. For these, a high agreement score means low confidence, so the score has to be flipped before the questions are averaged. That is not done in this step. You mark such questions later, on the Variables screen (chapter 6), where the option is called **Reverse-coded**.

Step 5: Combine time points

This step appears only when you chose two or more files, which you do when the same people answered at different times (a pretest file and a posttest file, for example). Statversity stacks the files into one dataset and adds a column recording which time point each row belongs to.

The step has these controls: **Name of the new time column** (names the column that records the time point, for example Time); a **Time label for** file name box per file (Pre, Post and so on); **Move** file **earlier** and **Move** file **later** buttons, which appear only if Statversity cannot work out the order from the labels; and the column-matching sections below. Click **Continue** to accept.

The column-matching sections tell you what Statversity found:

- **Possibly renamed**: columns that look like the same question under different names. For each one, choose **Same question: combine them** or **Different questions: keep separate**.
- **Only in some files**: columns present in some files but not others.
- **Same name, different questions**: columns that share a short name but ask different things. They are kept apart.
- **Matched in every file** (an expander): the columns Statversity matched without any doubt.

Step 6: Link people across time

Also only with two or more files. Choose **Don't link (default)** if the survey was anonymous, or **Link people by an ID** if each person entered an ID or code. When you link, pick the **ID column** (starting at **Choose a column…**) and optionally tick **Ignore extra spaces** and **Ignore capital letters**, which treat "ab12 " and "AB12" as the same person. A report, **How the IDs matched**, shows how many matched, and **Show unmatched IDs** lists the ones that did not.

Step 7: Review and import

The last step shows a summary list stating what will happen: **Responses read**, **Left out**, **Survey system columns**, **Answer choices**, **Split into yes/no columns**, **Time points**, **Survey file** and **Answer labels**. Nothing has changed yet. If it looks wrong, click a step in the list on the left to go back.

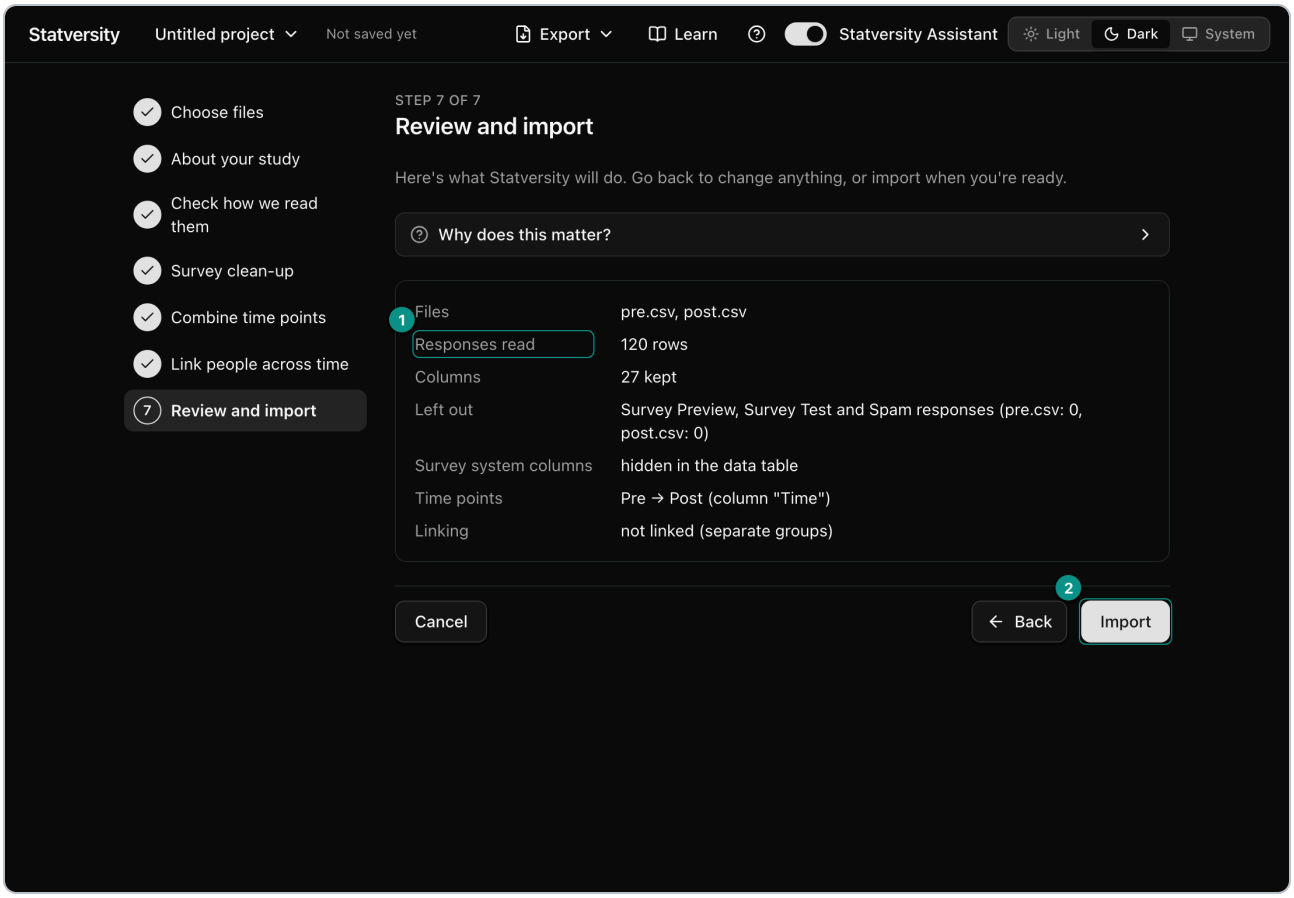


Figure 19. The Review and import step.

Number	Control	What it does
1	Summary list	States what the import will do. Read-only.
2	**Import**	Runs the import.

After the import, an **Import complete** banner appears. If Statversity finds repeated respondents it adds a **Duplicates** notice with **Review**, which opens the Data screen with the Duplicates panel, and a **Dismiss** button. The main button becomes **Set up variables** with the Assistant on, which opens the Variable Interview, or **Go to variables** with it off, which opens the Variables screen.

Supported file formats

Statversity reads:

- **CSV** (comma-separated text) files.
- **Excel** (.xlsx) workbooks. When a workbook has several sheets, **Which sheet holds the responses?** lets you choose.
- **SPSS** files. The format badge reads **SPSS file** or **SPSS portable file**.

- **Stata** files (**Stata file**).
- **SAS** files (**SAS file** or **SAS transport file**).
- Tables pasted from the clipboard.

Statversity also reads Qualtrics exports in all of these forms. For SPSS, Stata and SAS files it carries over the variable labels and answer labels stored in the file. You can also write data back out: from **Export** > **Data…** choose Excel, CSV, SPSS or Stata.

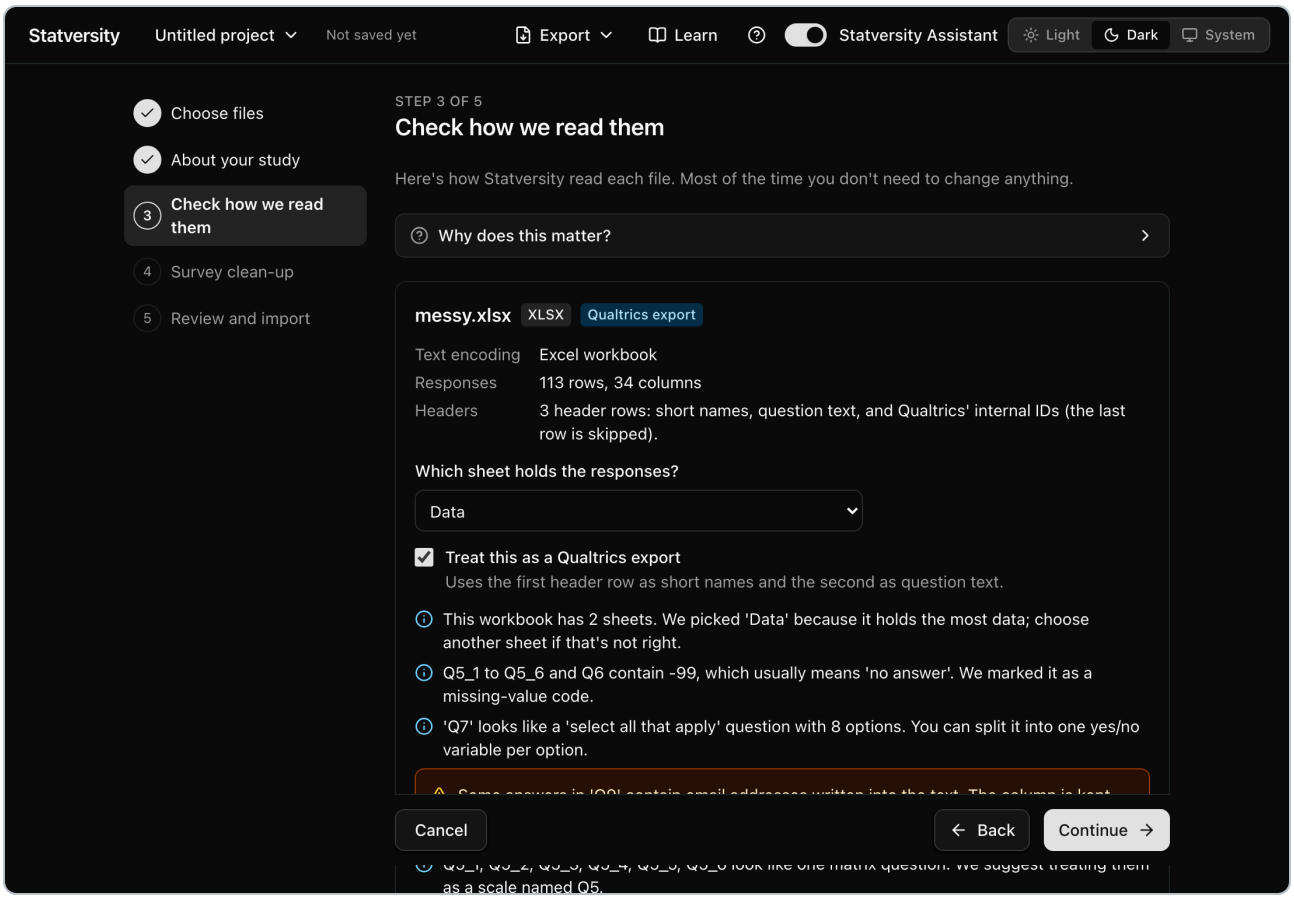


Figure 20. An Excel file (messy.xlsx) on the detection screen.

The screenshot shows an Excel workbook, messy.xlsx, on the detection screen. The format badge reads **XLSX**, **Text encoding** reads **Excel workbook**, and **Which sheet holds the responses?** offers the two sheets in the workbook. SPSS, Stata and SAS files look the same, with their own format badges.

Combining data

Combining is how you add to a dataset after it exists. On the Data screen, the **Combine data** toggle opens a side panel with one card per job. Each card has a **Start** button.

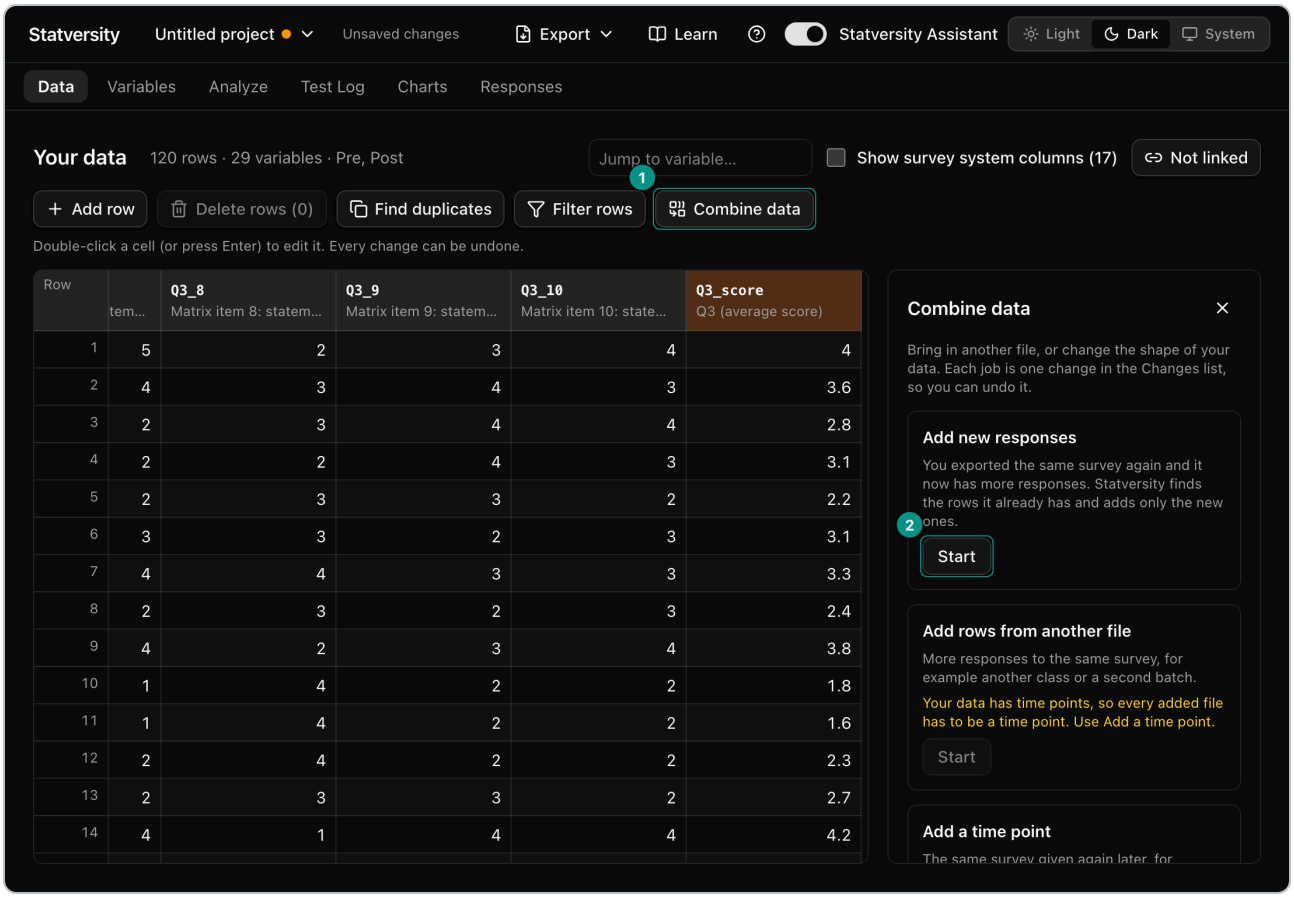


Figure 21. The Combine data side panel.

Number	Control	What it does
1	**Combine data** (toggle)	Opens and closes the side panel.
2	**Start** on a job card	Opens the Combine data screen for that job.

The jobs are:

- **Add new responses**: add rows from a new export of the same survey.
- **Add rows from another file**: stack another file under yours.
- **Add a time point**: add a later survey wave.
- **Add columns from another file (match by ID)**: bring in extra columns for the same people.
- **Split repeated columns into time points (wide to long)**: turn one column per time point into one row per time point.
- **One row per participant (long to wide)**: the reverse.

If you choose **Import data…** while a project already has data, the **Add new responses** button in that dialog takes you straight to this job (chapter 4).

Some jobs show a reason why they are unavailable, such as “Link people across time first.”

The Combine data screen walks through steps called **Choose file**, **Check**, **Confirm**, **Link people across time** and **Done**. Its buttons are **Cancel**, **Back**, and a main button whose label depends on the step: **Read files**, **Continue**, **Link people** or **Finish**. **Skip for now** appears on the link step. The confirm step applies the change as one step that you can undo. **Back to your data** returns to the Data screen.

Same question, asked to different people

Some surveys show different people different versions of a question, for example one version for each classroom. In the data, each version sits in its own column. To analyze them together you merge the versions into one column. This is done on the Variables screen with **Same question, asked to different people…**, or in the Variable Interview (chapter 6).

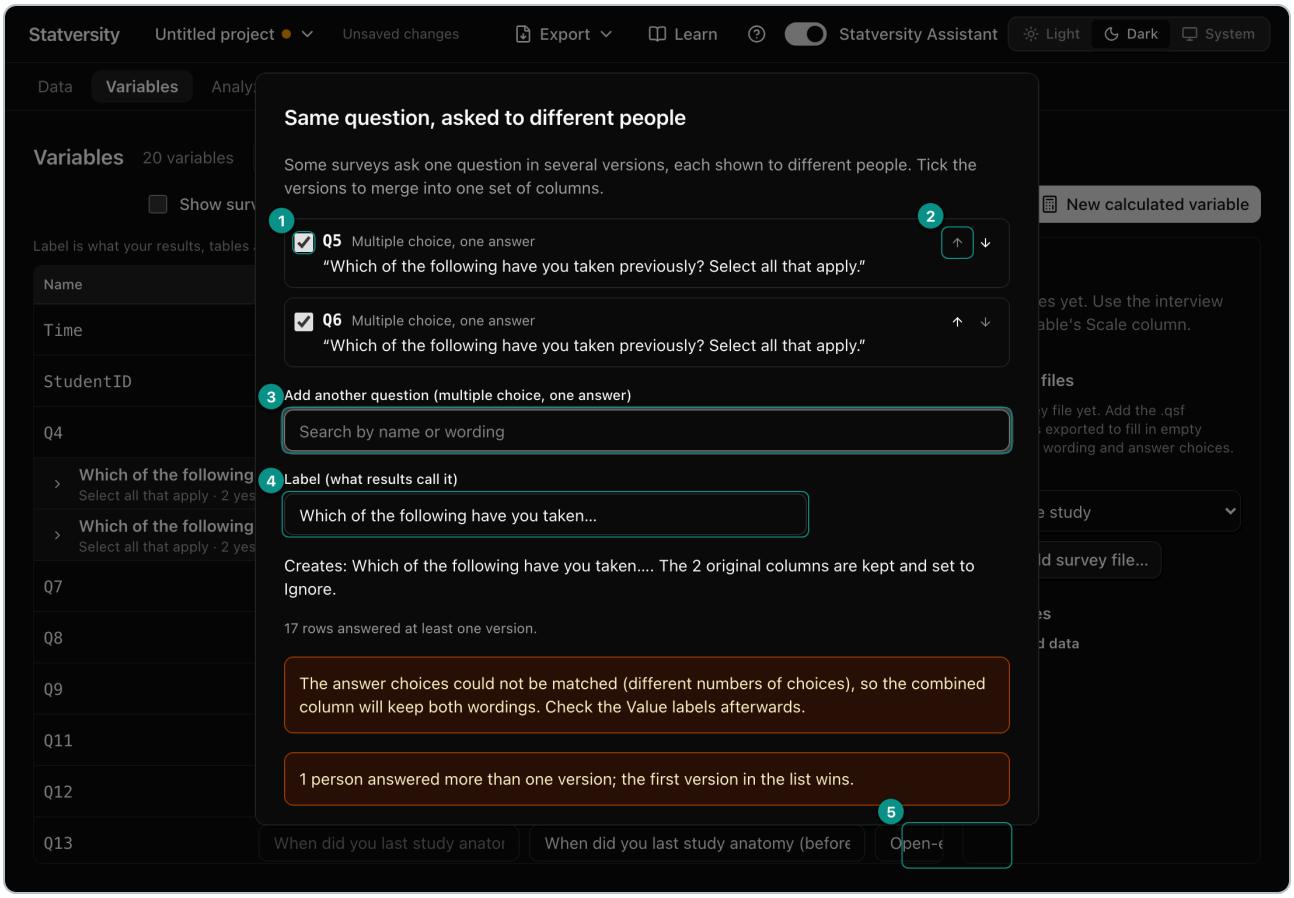


Figure 22. The combine-versions dialog, opened from the Variables screen with Q5 and Q6 added by search.

Number	Control	What it does
1	**Include** checkbox for each question	Chooses which versions to merge.
2	**Move** question **up** / **down** buttons	Set the order of versions.
3	**Search by name or wording** and **Add**	Add another version to the merge.
4	**Label (what results call it)**	Names the merged variable. **Use the suggested label** fills in a suggestion.
5	**Combine**	Creates the merged column.

The dialog also has a **Cancel** button. In the screenshot Q5 and Q6 were added with the search box.

Wording match

When you import several files, such as a pretest, posttest and retention test, the same question may be called Q1 in one file and something else in another. Statversity matches questions by their wording, using the survey file when you supplied one. On the **Combine time points** step the summary line reports how many questions matched by wording and how many columns share a short name but ask different questions. Under **Matched in every file** each row is tagged **matched by wording**.

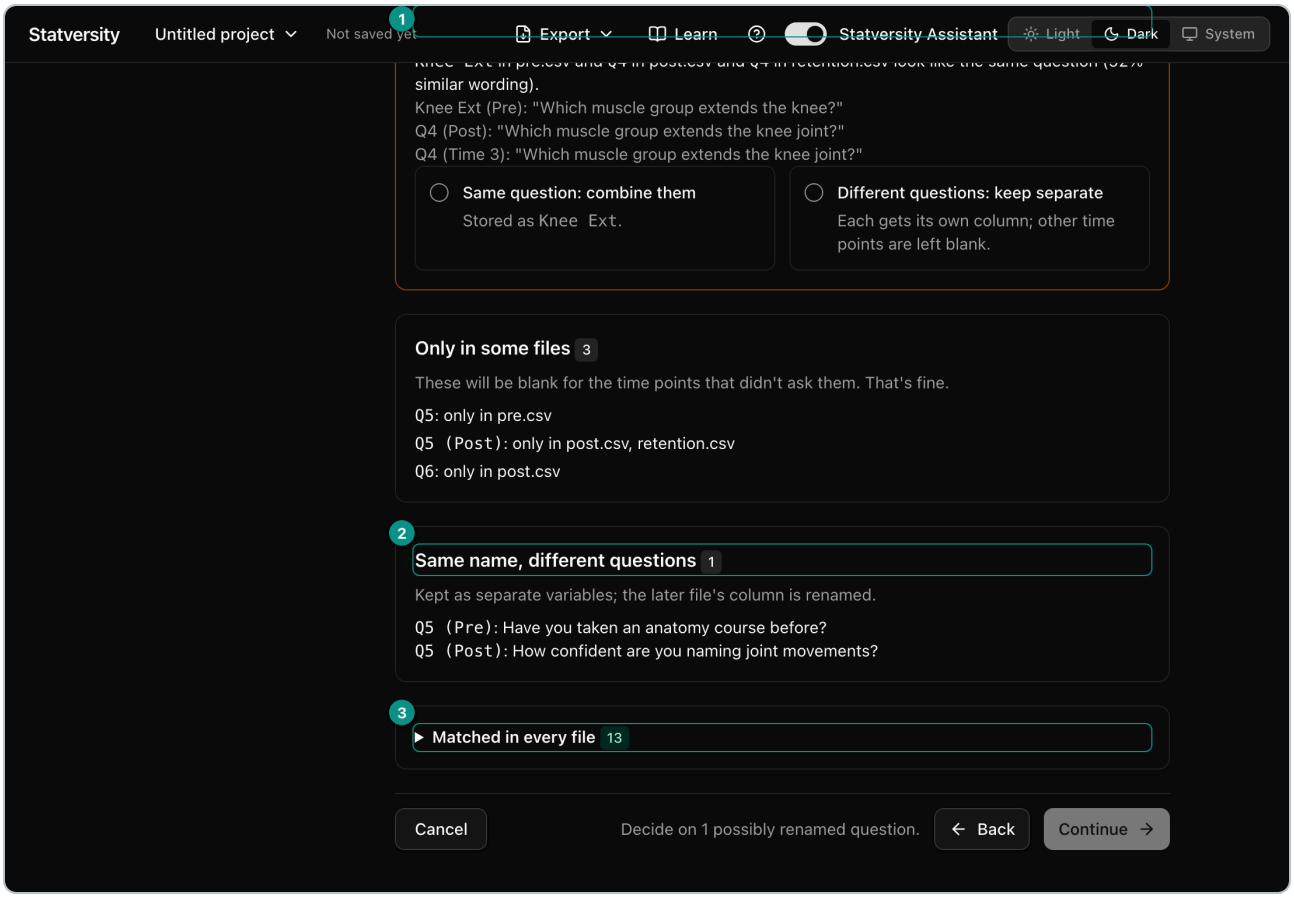


Figure 23. The wording match results on the Combine time points step.

Number	Control	What it does
1	**Possibly renamed**	Lists questions Statversity thinks are the same question under another name; you choose **Same question: combine them** or **Different questions: keep separate**.
2	**Same name, different questions**	Lists columns with the same short name but different wording. They are kept separate and the later file's column is renamed.
3	**Matched in every file**	Expander showing the questions matched in every file.

The screenshot is the Combine time points step (not a dialog), scrolled to the column-matching sections for a pretest, posttest and retention file.

CHAPTER 6

Variables

A variable is a column in your data: one question, one measurement, or one piece of information about each person. The Variables screen is where you tell Statversity what each column is, because the right test depends on it. A column of 1-to-5 ratings and a column of student ID numbers are both numbers, but you would never average the IDs. This chapter covers the Variables screen, what the variable types mean, the Variable Interview and scale scoring, calculated variables, answer groups, the Data grid and editing, Duplicates, and Filter rows.

The Variables screen

Open it with the **Variables** tab. It also opens automatically after import when the Assistant is off. If there is no data, it shows **No data yet**.

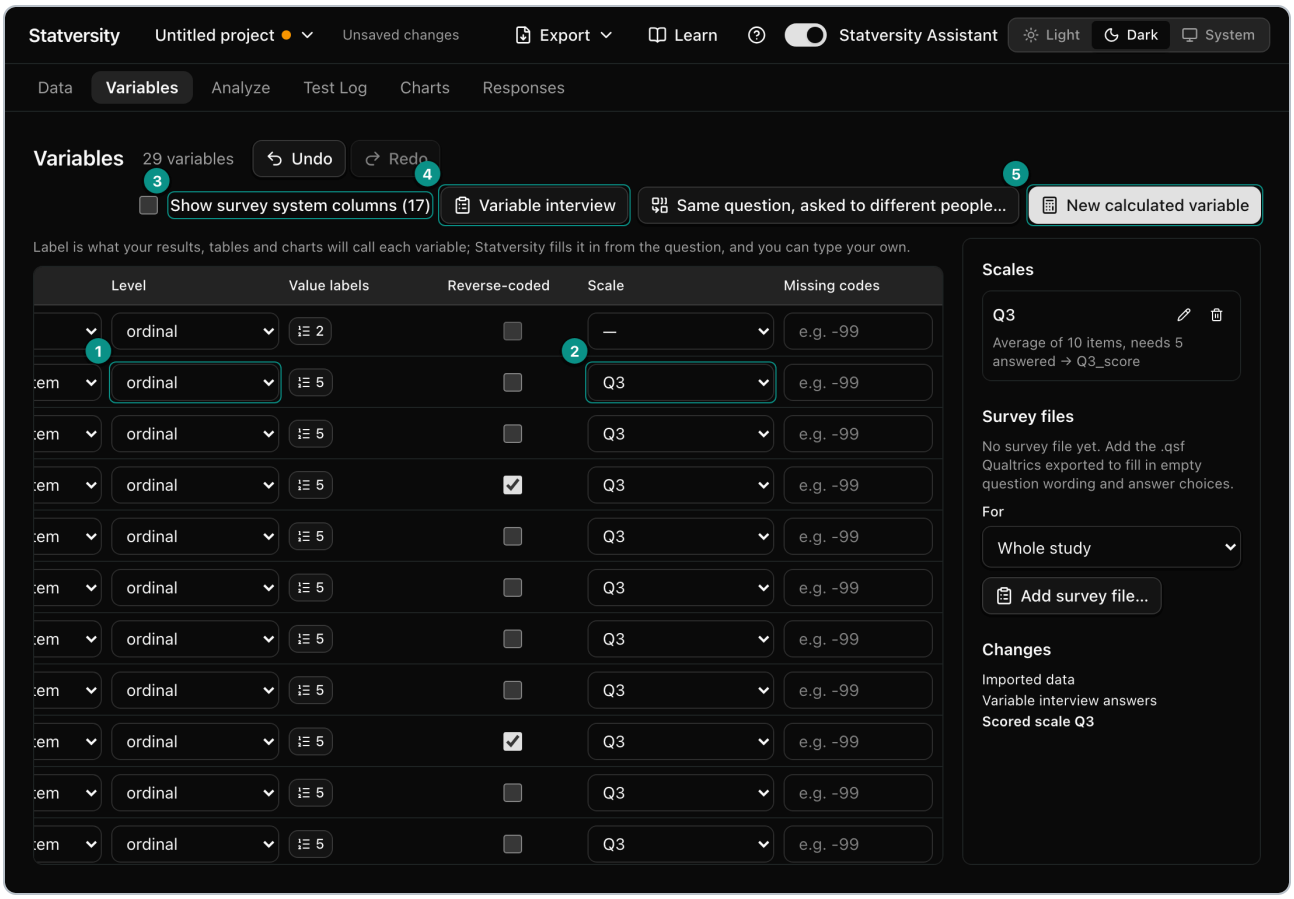


Figure 24. The Variables screen with numbered badges.

Number	Control	What it does
1	**Level** drop-down on a row	Sets whether the variable is **nominal**, **ordinal** or **continuous**.
2	**Scale** drop-down on a row (and the **Scales** side panel)	Puts a question into a scale, or shows **—** for none.
3	**Show survey system columns** checkbox (with a count)	Shows the survey tool's bookkeeping columns, which are hidden by default.
4	**Variable interview**	Re-opens the Variable Interview in review mode.
5	**New calculated variable**	Opens the builder for new columns.

The screenshot is the one-group pretest and posttest practice data after the Variable Interview, so a scale Q3 and a score column already exist. **Groups from answers**, which makes a category variable out of answers, is not a button on this screen; you reach it inside the builder (see Calculated variables below). The toolbar can also show **Show ignored columns**; it is not in this screenshot.

The title reads **Variables** with a count such as "12 variables". The toolbar has **Undo** and **Redo** buttons (their tooltips name the action and shortcut), a **Show survey system columns** checkbox with a count, **Show ignored columns** with a count, **Variable interview**, **Same question, asked to different people…** and **New calculated variable**.

The table

Each row is a variable. The columns are:

Column	What you can do
Name	The column name in your file.
Label	A friendly name that results will use. Type in it to change it.
Question text	The full wording of the question, filled in from a Qualtrics survey file if you gave one.
Role	What the variable is for (see below).
Level	**nominal**, **ordinal** or **continuous**.
Value labels	A button showing the number of answer labels, or **None**. It opens the **Answer choices for** the variable dialog.
Reverse-coded	A checkbox. Tick it for a question worded the opposite way from the others in its scale.
Scale	A drop-down that puts the question in a scale.
Missing codes	A box where you type codes that mean "no answer", for example -99.

Sets of related columns, such as the items of one matrix question, fold into group rows. Use **Expand** or **Collapse** with the group name to open and close them. Other per-row items:

- **Read as dates…** converts a text column into dates. A dialog, **Read** the name **as dates**, asks how the dates are written, with **Cancel** and **Read as dates**.
- **Keep as text** stops Statversity treating a column as dates.
- A **calculated** badge marks columns you made, and a **Remove calculated variable** button deletes one.

Every edit can be undone. Drag the **Resize side panel** edge or a **Resize** column edge to change widths.

Roles

The **Role** drop-down says what the variable is:

- **Identifier (ID)**: a code for each person.
- **Group or cohort**: which class, program or condition someone was in.
- **Time point**: when the answer was given (pre, post).
- **Test question**: a question with a right answer.
- **Test total score**: the sum of the test questions.
- **Other score or measurement**: any other number.
- **Survey (Likert) item**: an agree-disagree or rating question.
- **Background / demographic**: year, age, program.
- **Multiple answers (select all that apply)**
- **Open-ended text**: written answers.
- **Ignore**: leave it out. When you choose it a note says **Moved to ignored columns. Tick 'Show ignored columns' to see it.**

Two more values appear on their own: **Scale score** (set automatically for scale columns) and **Not set**.

Side panels

- **Groups to compare**: ticked questions are offered first whenever an analysis or chart needs groups.
- **Scales**: lists each scale as **Average of** a number **items** or **Sum of** a number **items**. Each has **Edit scale** and **Delete scale** buttons. **Edit scale** opens a dialog with **Scale name**, a choice of **Average of the answered items** or **Sum of the items**, and **Minimum answered items**, then **Cancel** and **Save**. When empty it says **No scales yet. Use the interview or the table's Scale column.**
- **Survey files**: lists the Qualtrics survey files with **Replace…**, a **For** drop-down (**Whole study** or a time point), **Add survey file…** and **Question blocks…**, which opens **Question blocks from your survey**.
- **Changes**: a history list; the current step is marked **(current)**.

Variable types in plain language

Statversity asks for a **Level** for each variable. The word is the same as “level of measurement” in a statistics class, and it decides which summaries and tests make sense.

> **IN PLAIN LANGUAGE**
>
> **Nominal** (categories, **Categories (nominal)** in the Interview). The answers are names with no order. Examples: program (Nursing, Education), gender, yes/no, which class someone was in. You can count them and compare percentages, but an “average program” means nothing.
>
> **Ordinal** (**Ordered categories (ordinal)**). The answers have an order, but the gaps between them are not necessarily equal. Examples: strongly disagree to strongly agree, year in school (first, second, third), a “low, medium, high” rating. You can rank and compare them. Many researchers treat a multi-item scale of ratings as if it were numbers, but a single rating item is ordinal.
>
> **Continuous** (**Numbers (continuous)**, shown as **continuous** on the table; also called scale in some programs). The answers are numbers where equal gaps mean equal differences. Examples: score out of 100, age in years, minutes spent. You can average these.

If you are unsure, ask yourself: could I put these in a sensible order? If not, nominal. Could I say that the step from 2 to 3 is the same size as from 4 to 5? If yes, continuous. Otherwise ordinal.

> **CAUTION**
>
> Do not mistake codes for numbers. If you coded program as 1 for Nursing and 2 for Education, the column is nominal, even though it is stored as numbers.

The Variable Interview and scale scoring

With the Assistant on, the Variable Interview opens right after import. It walks you through your variables a step at a time. You can reopen it any time with **Variable interview** on the Variables screen; it then opens in review mode. The list of sections (**Interview sections**) down the side shows only those that apply: **Welcome**, **Same questions, different groups**, **What is each question?**, **Kinds of answers**, **Groups to compare**, **Knowledge questions**, **Answer choices**, **Test scoring**, **Question blocks from your survey**, **Scales**, **Scale scores**, **New groups** and **Summary**. Completed sections show a check mark and can be clicked.

Along the bottom are **Back**, **Start** (first step), **Continue** and **Finish**. **Skip for now** leaves the interview, and in review mode it is called **Close** and asks whether to discard unapplied changes. **Start over** resets your answers. In review mode, **Apply changes** applies the edits so far without leaving. **Finish** applies everything and lands on Variables.

What is each question?

For each question you choose a role from the same list as above in a radio list called **What is this?** and may edit the **Label (what results call it)**. A badge tells you if Statversity guessed from earlier projects, from your answers about your study, or from your survey. If many questions look alike, Statversity offers **Yes, tag all** a number, **No, one at a time** or **Review the list**.

Scales and reverse-worded items

A scale is a score built from several related questions. A “classroom confidence” scale might average five rating questions. The **Scales** step lets you build them.

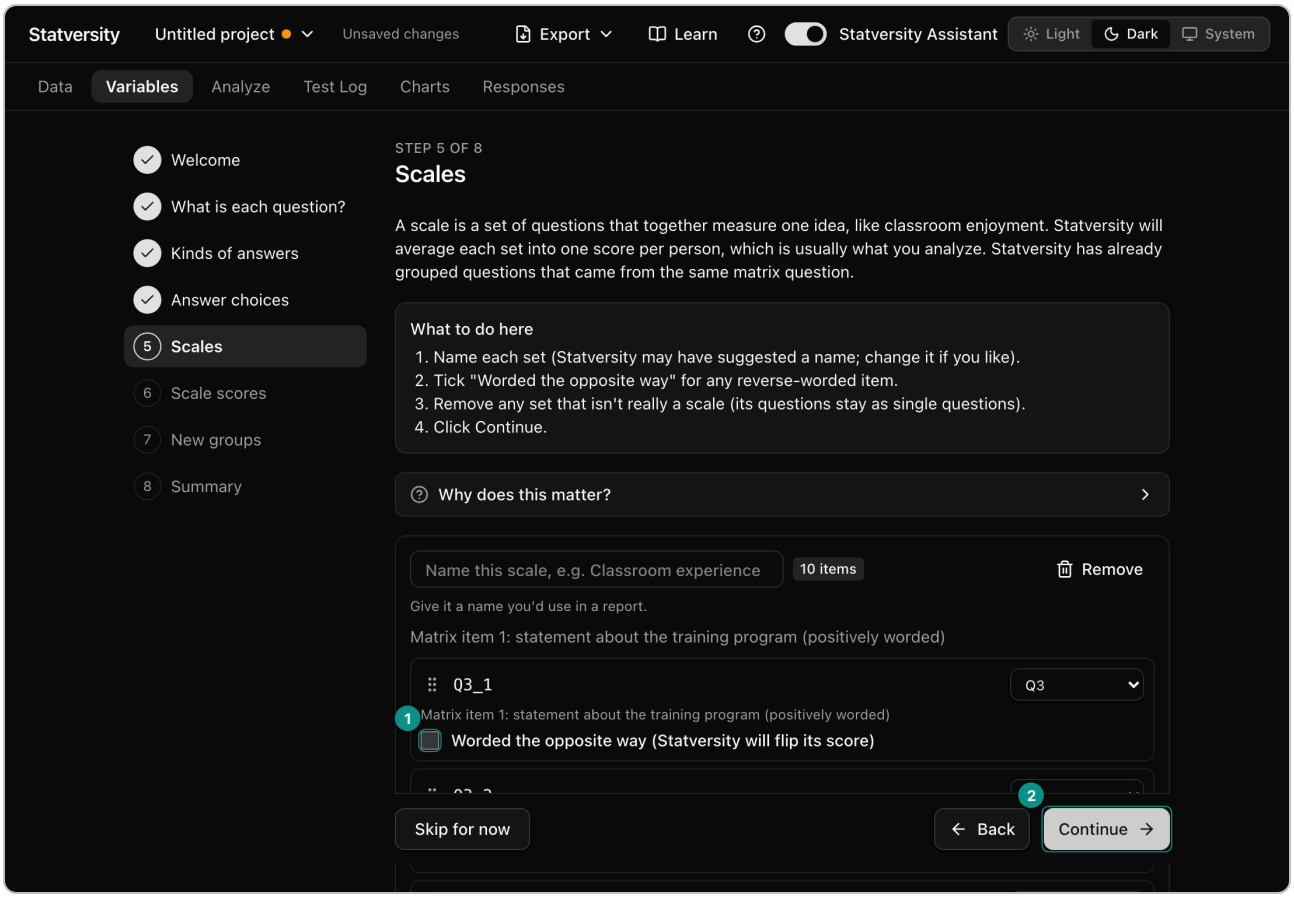


Figure 25. The interview step for building scales.

Number	Control	What it does
1	**Worded the opposite way (Statversity will flip its score)**	Marks a question as reverse-worded. Its scores are flipped before they are averaged.
2	**Continue**	Moves to the next step.

The screenshot is the **Scales** step (step 5 of 8). The choice between **Average of the answered questions (recommended)** and **Sum of the answers** is on the next step, **Scale scores**, so it is not numbered here.

On the **Scales** step each scale is a box with **Scale name**, **Remove**, and a **Scale for** question drop-down (**Not in a scale**) on every question. You can also drag a question into a scale and use **Add a scale**. On **Scale scores**, **Minimum answered questions for** the scale lets you require a minimum number of answered items; someone who answers fewer than that gets a blank score.

> **IN PLAIN LANGUAGE**
>
> A **reverse-worded** question says the opposite of the others. If four questions say “I feel confident” and one says “I feel nervous”, a high rating on the last one means low confidence. Without flipping it, adding or averaging would partly cancel out. Flipping a 5-point item turns 1 into 5, 2 into 4, and so on.

Average is the usual choice because it keeps the scale in the original 1-to-5 units and still works when someone skipped a question. **Sum** is natural for tests where you add up points.

Other interview steps cover **Kinds of answers** (set the level, with preset buttons under **Use a common scale:**), **Answer choices** (a sortable list), **Groups to compare** (**Which of these do you want to compare groups by? Tick all that apply.** with **None for now**), **Knowledge questions** and **Test scoring** (answer keys, using **Download answer key template**, **Load answer key…** or **Use the practice answer key**), and **Question blocks from your survey**. The **Summary** shows everything and has a card **Want Statversity to plan your analyses?** with **Plan my analyses** and **Not now, finish**.

Calculated variables

A calculated variable is a new column Statversity computes from existing ones. Click **New calculated variable**.

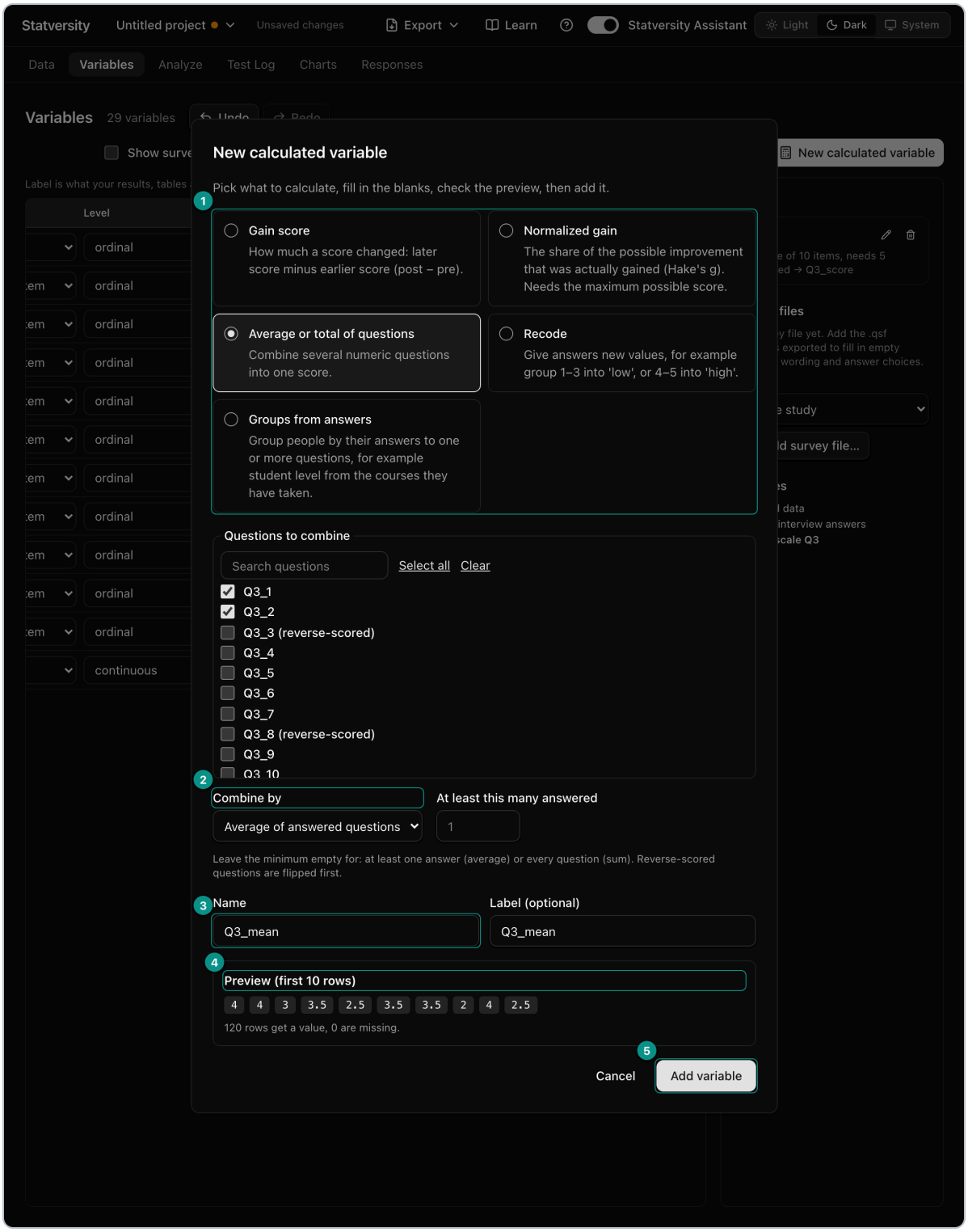


Figure 26. The calculated variable builder.

Number	Control	What it does
1	**What to calculate**	Chooses the kind of new column.
2	**Combine by** (and **At least this many answered**)	Fields for the chosen kind; here the average-or-total kind is shown, with Q3_1 and Q3_2 ticked. They change with the kind (see below).
3	**Name** and **Label (optional)**	Name the new column.
4	Preview list	Shows the first values (**Calculating…** while loading).
5	**Add variable**	Creates the column. **Cancel** beside it closes without creating.

The kinds are:

- **Gain score**: the later score minus the earlier one. Choose **One score measured at two time points** (with **Score**, **Later time point** and **Earlier time point**) or **Two columns in the same row** (**Later score (post)** and **Earlier score (pre)**).
- **Normalized gain**: the same inputs, plus **Maximum possible score**. It shows how much of the possible improvement someone achieved.
- **Average or total of questions**: pick questions in a picker with **Search questions**, **Select all** and **Clear**; **Combine by** is **Average of answered questions** or **Sum**; **At least this many answered** sets the minimum.
- **Recode**: pick the **Variable to recode**, then rules with **Add a rule** (each has old values, a new value and a **Remove rule** button). **Answers no rule covers** is **Keep as they are** or **Make them missing**.
- **Groups from answers**: see the next section.
- **Time between two dates**: **Later date**, **Earlier date** (**Another column** or **A fixed date**), **Time (optional)**, **Unit** (**Minutes**, **Hours**, **Days**, **Weeks**, **Months**, **Years**) and **Whole numbers only**.
- **Part of a date**: **Source** and **Part** (**Year**, **Month (1–12)**, **Month name**, **Day of month**, **Day of week**, **Hour**, **Week of year**).

The new column appears on Variables marked **calculated**.

Answer groups

Answer groups let you build categories out of answers. For example you could make a “Heavy user” group out of several usage questions. Choose **Groups from answers** in the builder, or use the **New groups** step of the interview (**Make a group variable**, **Make another group variable**, **Skip, no new groups**).

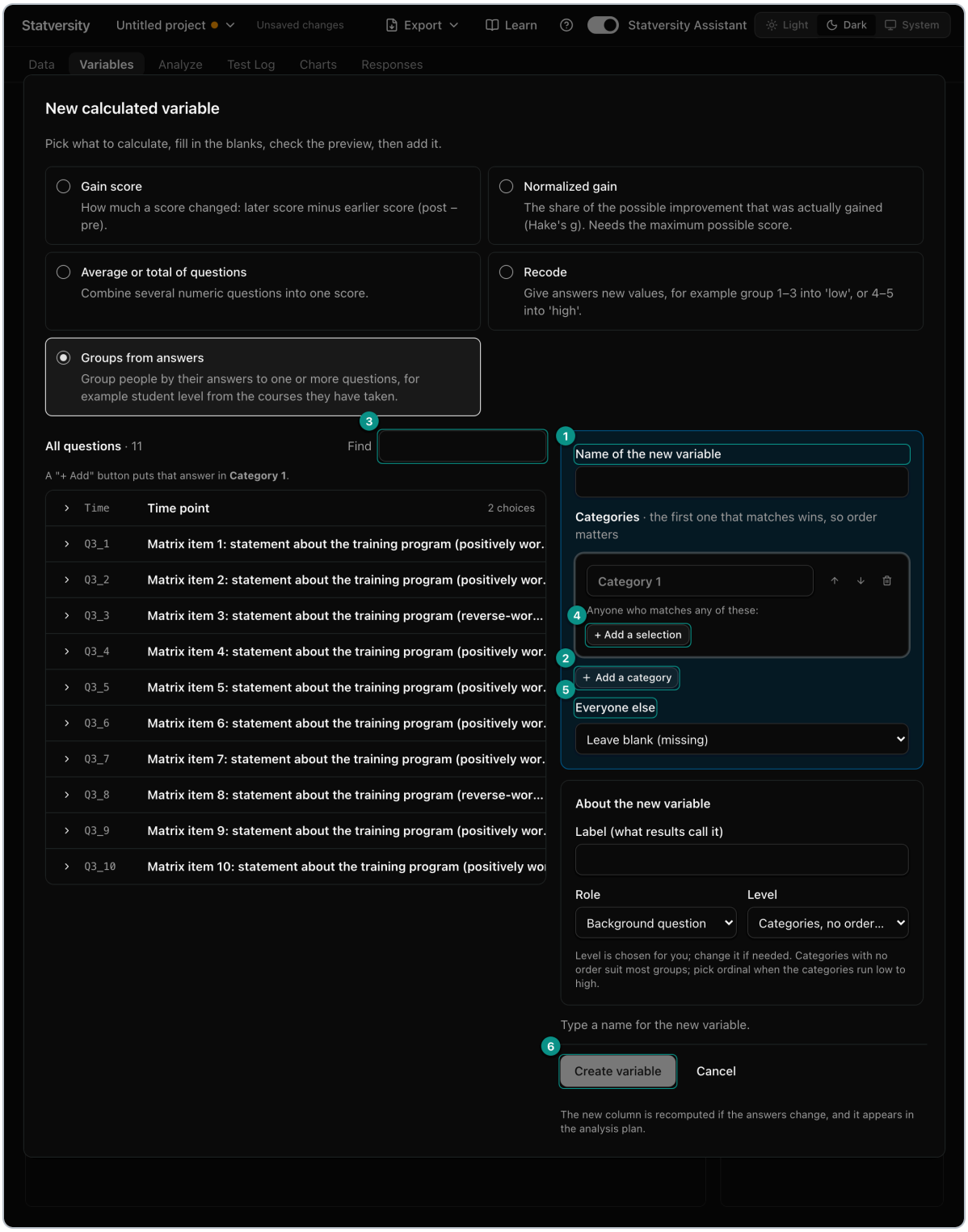


Figure 27. The answer groups builder.

Number	Control	What it does
1	**Name of the new variable**	Names the group variable.
2	**Category** number **name** boxes and **Add a category**	Define the categories. The first category that matches wins. **Move** up and down and **Remove** buttons reorder or delete.
3	**All questions** list with **Find a question**	Lists questions and their answers; **+ Add** puts an answer in a category.
4	**Add a selection to** the category	Opens a picker with a **Question** drop-down and answer ticks; **Add selections** confirms.
5	**Everyone else**	Says where people who match nothing go: **Leave blank (missing)** or a category.
6	**Create** the variable name	Creates the variable. **Cancel** closes.

You can also set the **Label (what results call it)**, **Role** and **Level** of the new variable.

Survey flow

When you imported a Qualtrics survey file with display logic (questions shown only to some students), the toolbar also has a **Survey flow** button. It opens a diagram of which answer leads to which question, a search across every question and answer choice, and a quicker way to build a group: **Make a group** collects the same answer from every branch it appears in, such as “took a course at another school”, into one yes/no column. The Variable Interview shows the same information on **What is each question?**, in a **Where this question leads** box on each gate question. Chapter 17 walks through both.

The Data grid and editing

The **Data** tab shows your data as a table. The title is **Your data** with a line such as “120 rows · 18 variables”. When a filter is in use the line reads “k of n rows · Filter: name” with **Clear** beside it. If columns were flagged as personal information you see “n column(s) marked PII”.

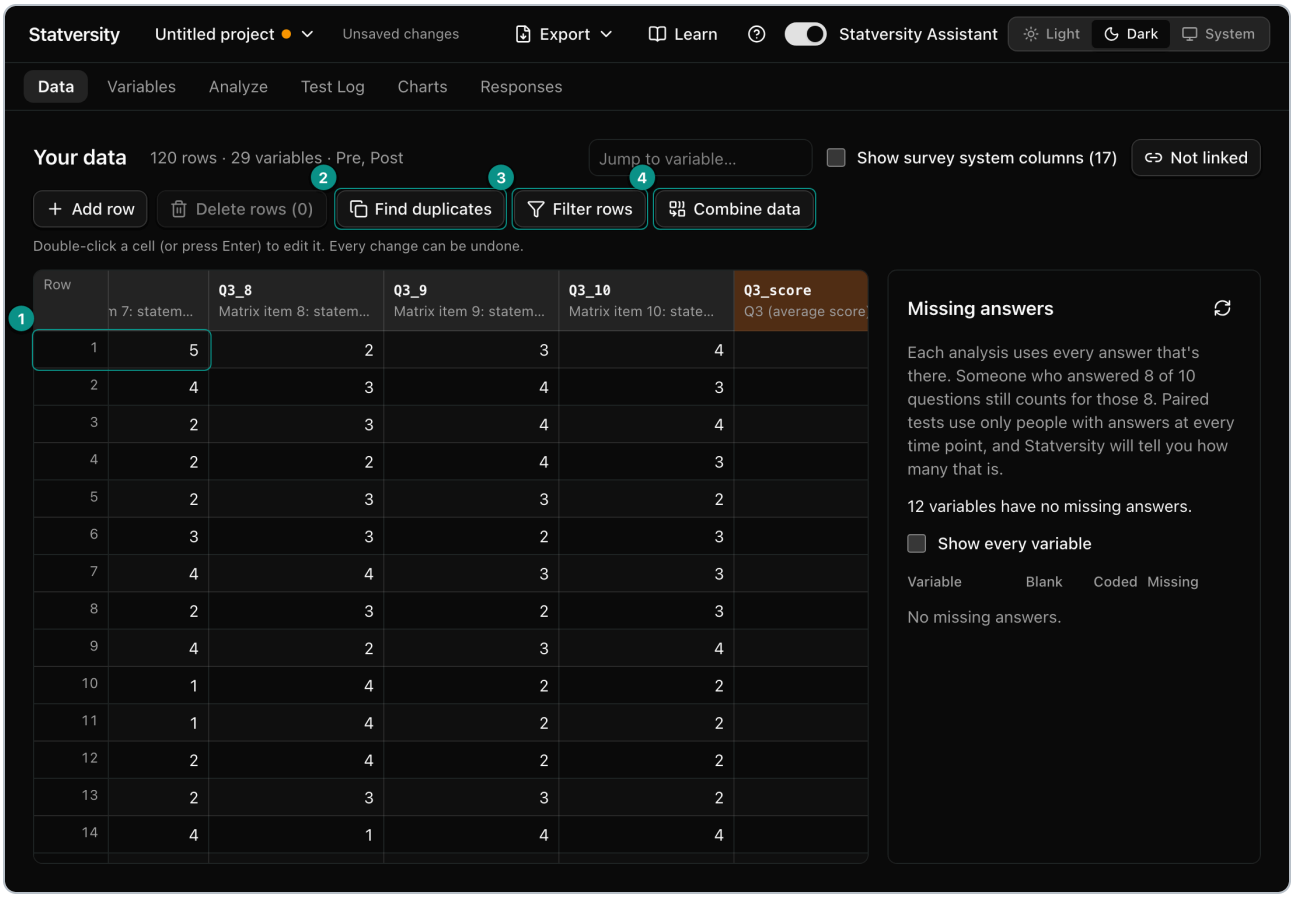


Figure 28. The Data screen with numbered badges.

Number	Control	What it does
1	A cell in the grid	Double-click it (or press Enter or F2) to edit.
2	**Find duplicates**	Opens or closes the Duplicates panel.
3	**Filter rows**	Opens or closes the Filter panel.
4	**Combine data**	Opens or closes the Combine panel (chapter 5).

Editing

The hint above the grid says **Double-click a cell (or press Enter) to edit it. Every change can be undone.** While editing, Enter saves, Tab saves and moves right, and Esc cancels. Cells that have answer labels edit through a drop-down. Missing-value codes are flagged with a tooltip. Click a row number to select the row; Shift selects a range, and Cmd (Ctrl on Windows) adds or removes. The toolbar buttons **Add row** and **Delete rows** with a count handle adding and deleting. **Delete rows** is disabled until you select rows. Cmd+Z undoes any of these.

Other controls on this screen:

- **Jump to variable…**: type a name and press Enter to scroll to that column.
- **Show survey system columns** with a count.
- A link button, **Not linked** or **Linked by** an ID column, appears when time points exist. It opens **Link people across time**, with **Don't link** or **Link people by an ID**, **ID column**, **Ignore extra spaces**, **Ignore capital letters**, the link report, **Close** and **Apply**.
- The default side panel is **Missing answers**: a table of **Variable**, **Blank**, **Coded** and **Missing** for each variable, with a **Show every variable** checkbox and a **Refresh missing-data summary** button.

> **TIP**
> Fix typos in the data rather than in your head. Every edit is recorded in the Changes list, and **Undo** is always there if you change your mind.

Duplicates

A duplicate is the same person appearing twice, usually because they submitted the survey twice. Open **Find duplicates**.

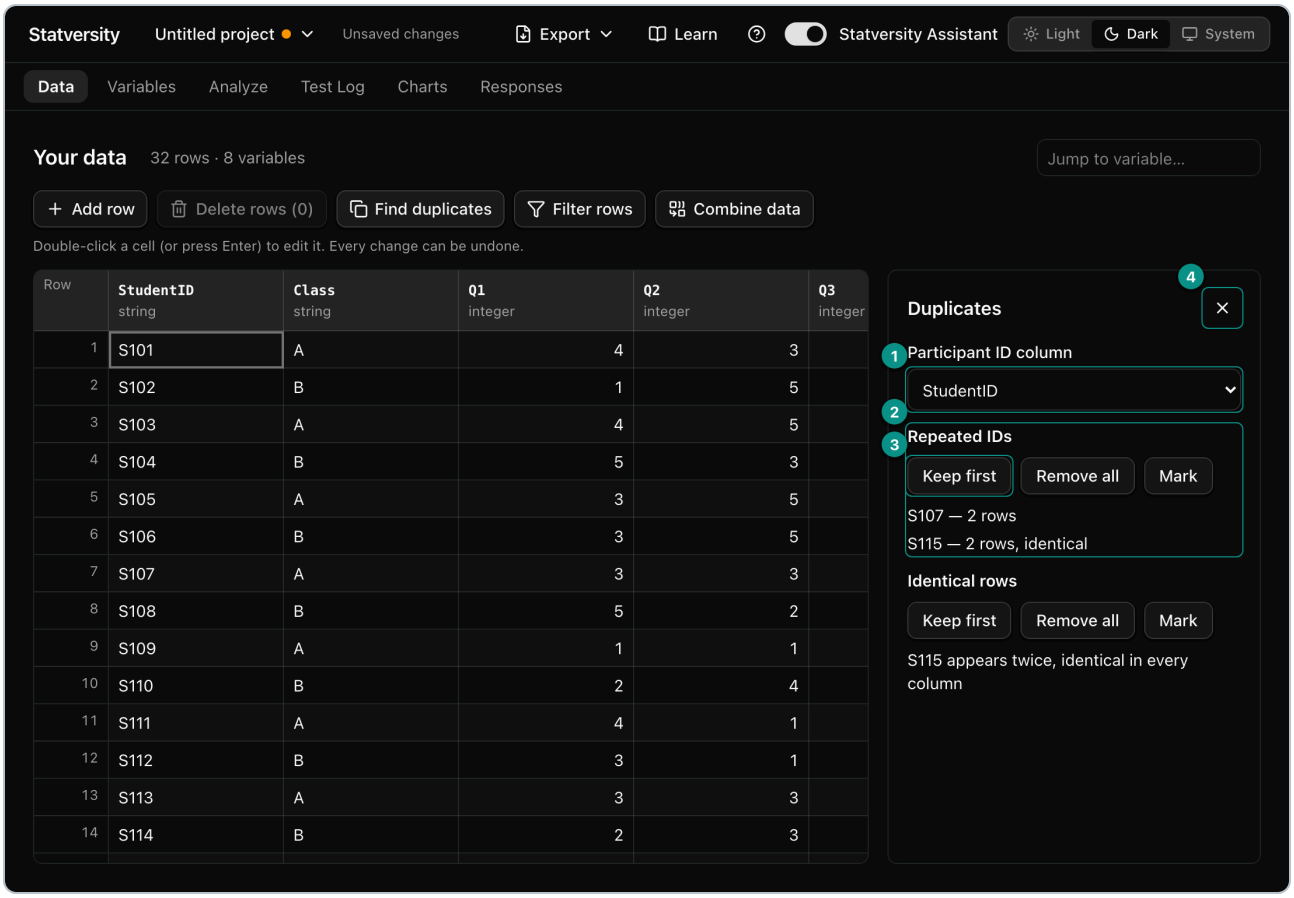


Figure 29. The Duplicates panel.

Number	Control	What it does
1	**Participant ID column**	Picks the ID used to spot repeated people (**No ID variable** or a column).
2	**Repeated IDs**	Section listing groups of rows that share an ID. Clicking a group selects its rows in the grid. An **Identical rows** section below it lists rows that match in every column.
3	**Keep first**, **Remove all**, **Mark** (for repeated IDs)	Keep one copy, remove every copy, or mark them.
4	**Close duplicates panel**	Closes the panel.

When there is nothing to report the panel says **No repeated IDs and no identical rows.** All actions can be undone.

> **CAUTION**
> **Remove all** removes every copy, including the first. If you want to keep one response per person, choose **Keep first**.

Filter rows

A filter keeps only some rows for your analyses, for example only students who finished the survey, or only one class. Filtering does not delete anything; the rows are just not used. Click **Filter rows** to open the panel, then **New filter**.

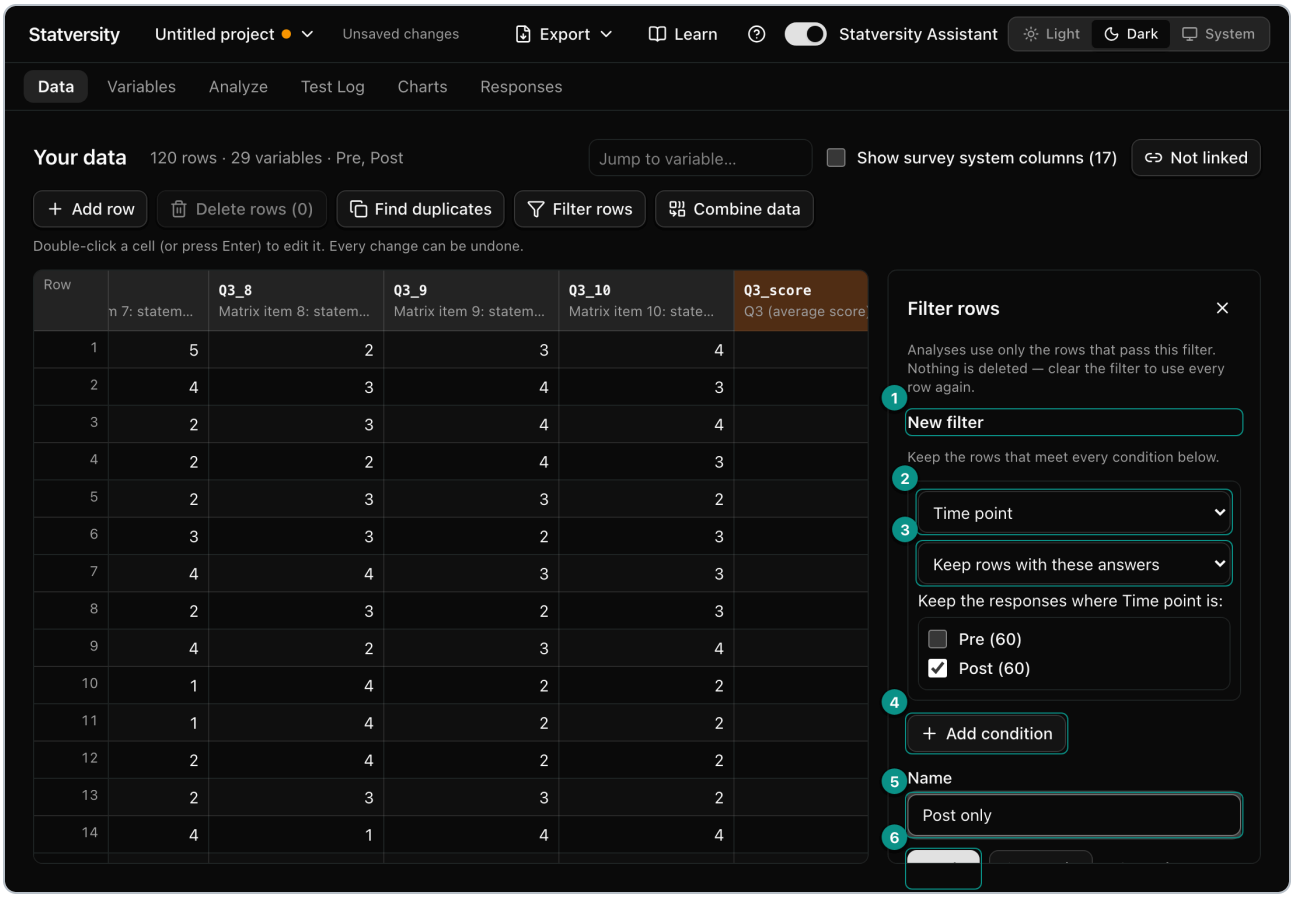


Figure 30. The Filter rows panel with one rule.

Number	Control	What it does
1	**New filter**	Starts a filter.
2	**Column to filter by**	Picks the column the rule applies to.
3	**matches** drop-down	**Keep rows with these answers** or **Keep rows without these answers**, with answer ticks.
4	**Add condition**	Adds another rule.
5	**Name** box	Names the filter.
6	**Apply**, **Save only**, **Cancel**	Use the filter now, save it for later, or close.

For number columns you can instead choose **at least**, **at most** or **between**, and for dates **on or after**, **on or before** or **between**. **Remove this condition** deletes one rule. The saved list shows **Use** or **Stop using**, **Edit**, **Delete** and **In use: k of n rows**; with none it says **No saved filters yet.** A filter that is in use limits every later analysis, and you will see it in the Data screen's counts line.

Statversity can also run a test separately for each group (**Run separately for each group** on the Choose your variables screen), covered in chapter 7. There is no row-splitting control on the Data screen.

CHAPTER 7

Analyze

The **Analyze** tab is where you decide which statistical test fits your question, tell Statversity which columns play which part, and check that the test is a fair choice for your data. This chapter walks through that area in the order you will meet it: planning, the Variable Interview, the Test Advisor, the recommendation, **Choose your variables**, and the assumption checks.

> **IN PLAIN LANGUAGE**
>
> A **statistical test** is a fixed recipe for answering one kind of question, such as "did scores go up?" or "do these two groups differ?". Different questions and different kinds of data need different recipes. Choosing the wrong one gives an answer that looks official but does not mean what you think. Statversity's job on this tab is to help you choose well.

This chapter assumes the **Statversity Assistant** switch in the top bar is on, which is the default. If you turn it off, the Analyze tab changes; chapter 10 covers that.

Two ways into an analysis

When your data are loaded and you have no saved plan, clicking **Analyze** opens the Test Advisor, where you answer a few questions and Statversity suggests a test. If you have built a plan (next section), **Analyze** opens a landing page that tracks your plan instead. Either way you can reach the other route from there: the landing has an **Open the Test Advisor** button, and the Advisor has a **Choose a test myself** button.

Plan my analyses

Plan my analyses builds a to-do list of analyses from what you tell Statversity about your study. It is the best starting point if you have a real study with several questions, because you do not have to remember what to run. You reach it from the **Plan my analyses** button at the end of the Variable Interview, from the **Start planning** button on the plan card that appears on Analyze, the Advisor and the All tests screens, or from **Edit** or **Change the plan** later.

The screenshot below shows the plan screen for the three-groups, three-times practice data (pretest, posttest and follow-up), where Statversity built three analyses.

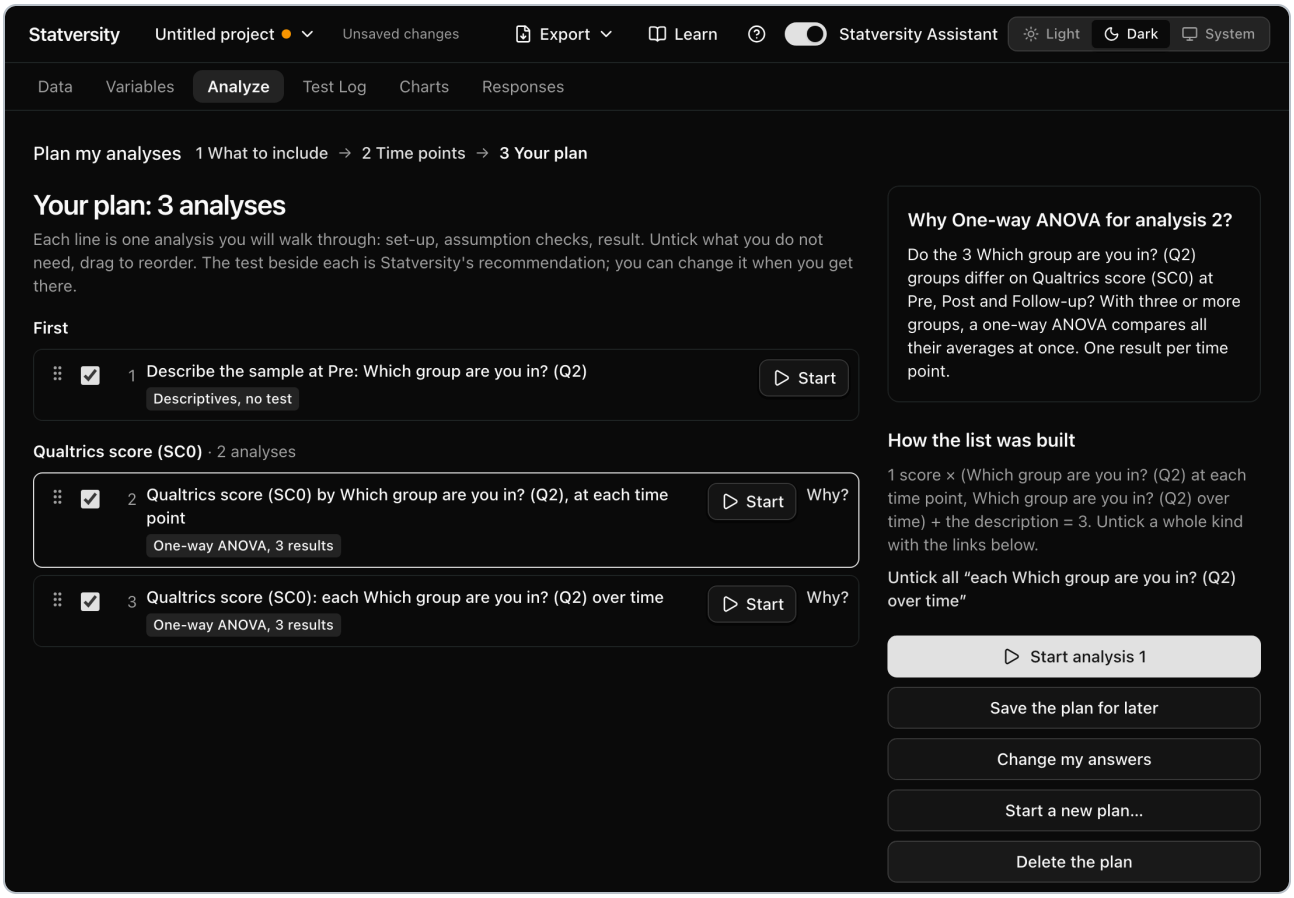


Figure 31. The Plan my analyses screen.

The page has a step list on one side: **What to include**, **Time points** (only when your data have time points), and **Your plan**. The layout switch at the top, **All on one page** or **One question at a time**, controls whether you see everything at once or one question per page. In **What to include** you tick what you want analyzed under the sections **Background questions**, **Groups to compare**, **Scales**, **Knowledge questions** and **Other scores**. Anything you do not tick is listed on a **Left out:** line so nothing disappears silently.

If your survey shows different questions to different students, a comparison you care about (for example “has ever taken a course at another school”) may be spread over several branch questions. Make it one group first, from the **Where this question leads** box in the interview or from the **Survey flow** screen on Variables; it then appears here under **Groups to compare** like any other grouping question. Chapter 17 shows how.

Next to each item the **How to use {label}** menu decides its job. **Describe only** reports it (averages, counts) without testing anything. **Compare scores by it** (for a grouping question) or **Relate to scores** (for a number) uses it in tests. **Describe and compare** and **Describe and relate** do both. If you have several grouping questions, **Make the main comparison** marks the one that matters most; it then carries a **Main comparison** badge.

When your data have time points, the **Time points** step asks what you want to compare. The choices are the groups against each other at each time point, each group over time, **Both** (recommended), or one pooled result across all time points. You can also say whether to compare scores by your background questions and whether to describe the sample first.

> **TIP**
> Not sure what to tick? Leave the defaults and click **Build my plan**. You can untick any analysis from the list afterwards, and nothing runs until you press **Start**.

Click **Build my plan** to produce the list titled **Your plan: {n} analyses**. Each row has a checkbox (**Include analysis {n}: {title}**), a drag handle for reordering, **Start**, and **Why?**, which explains in plain words why Statversity added that analysis. Rows you have run offer **Run again** and **Redo this step**. Below the list are **Start analysis {n}**, **Save the plan for later**, **Change my answers**, **Start a new plan…** and **Delete the plan**. The last two ask you to confirm first, because they discard your list.

The Analyze landing, plan sidebar and cards

Once a plan exists, **Analyze** opens a landing page. Its tab title reads **Analyze**, or **Plan complete · {done} of {total}** when everything is finished.

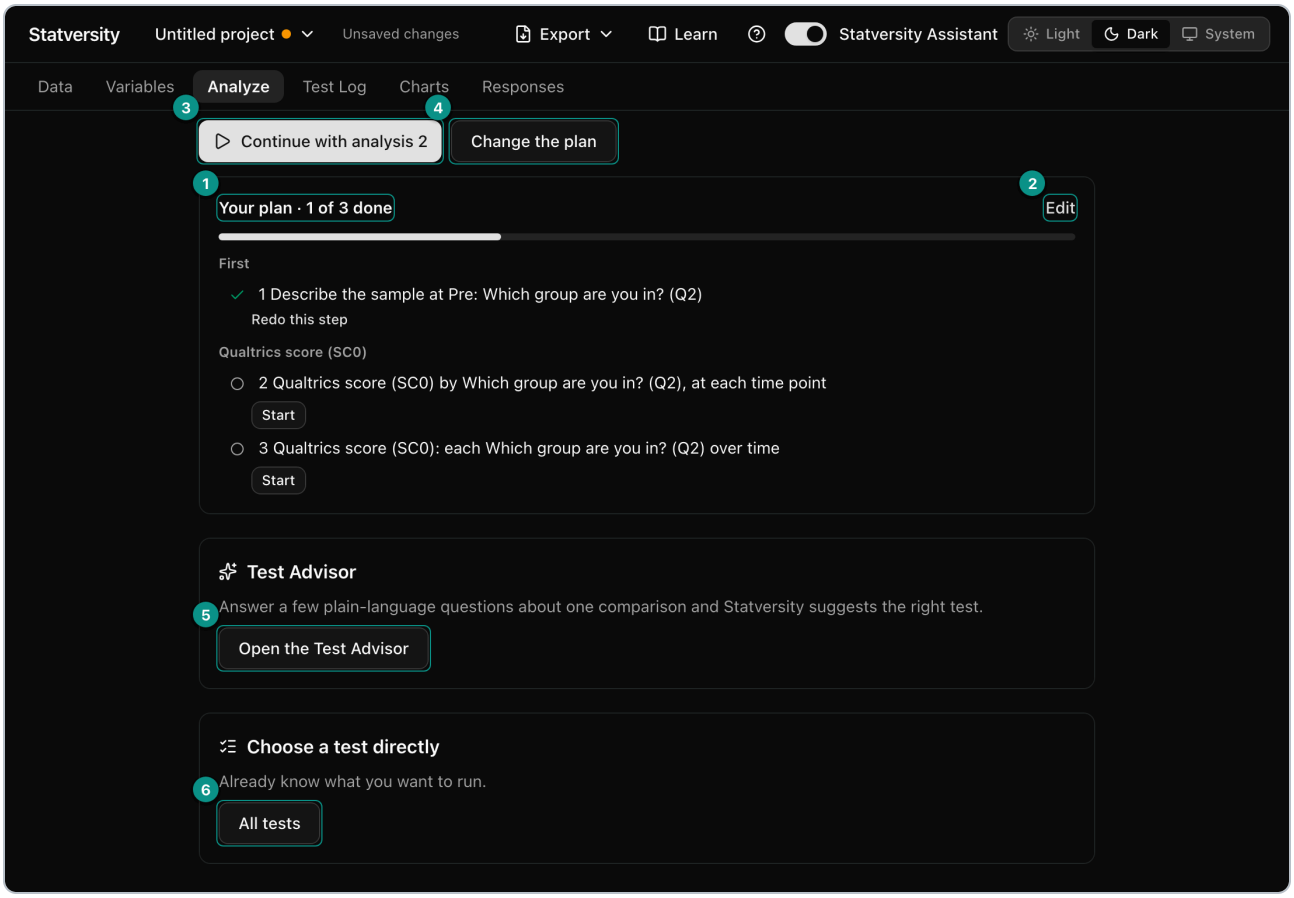


Figure 32. The Analyze landing with the plan sidebar and cards.

The screenshot uses the same three-groups practice data. Analysis 1 (describe the sample) was run and then **Stop for now** was clicked, so the landing shows **1 of 3 done** and **Continue with analysis 2**. In this view the plan appears as a card on the page. The numbered legend below follows the badges in the screenshot.

Number	Control	What it does
1	**Your plan · {done} of {total} done** (sidebar)	Lists each planned analysis with a prefix showing its state: **Done:**, **Current:**, **Skipped:** or **To do:**. Clicking a done item reopens its result.
2	**Edit**	Goes back to the plan screen to change what is included.
3	**Start analysis {n}** or **Continue with analysis {n}**	Starts the next unfinished item and opens **Choose your variables**, already filled in from your plan.
4	**Change the plan**	Opens the plan screen.
5	**Open the Test Advisor**	Opens the question-by-question Advisor instead of your plan.
6	**All tests**	Opens the searchable menu of every test (chapter 10).

A **Hide the plan** / **Show the plan** button collapses or restores the plan sidebar where one is shown; it is not on this landing view. The plan items also offer **Start**, **Skip this one**, **Run again** and **Redo this step**. An **Out of date** badge on an item means the data changed after that analysis ran; see chapter 8 for why. When there are many items, a line such as **and {n} more… show all** expands the list. Between analyses the footer shows **Next: {n} of {total} · {title}** with **Next analysis** and **Stop for now** (or **Back to the plan**).

The Variable Interview

The **Variable Interview** is a short guided conversation that runs right after you import data (with the Assistant on). You can also reopen it at any time from the Variables screen with **Variable interview**. It is not part of choosing a test, but everything on the Analyze tab depends on its answers, so this section explains what each question is really asking.

> **IN PLAIN LANGUAGE**
> Statversity cannot tell from a column of numbers whether it is a student ID, a test score, a rating from 1 to 5, or a group code. A column of 1s and 2s could be a score or “control” and “treatment”. The interview is how you tell it. Tests behave very differently depending on these labels, so a wrong answer here produces a wrong recommendation later.

The left side lists the sections under **Interview sections**. Which ones appear depends on your data: **Welcome**, **Same questions, different groups**, **What is each question?**, **Kinds of answers**, **Groups to compare**, **Knowledge questions**, **Answer choices**, **Test scoring**, **Question blocks from your survey**, **Scales**, **Scale scores**, **New groups** and **Summary**. Completed sections get a check mark and can be clicked to revisit. The buttons along the bottom are **Back**, **Start** (first step) / **Continue** / **Finish**, **Start over**, and **Skip for now**.

What each section is asking:

- **Welcome** shows your earlier answers under **About your study:** (use **Change** to revise them) and flags suspected duplicate rows. Your study answers (design, groups, whether people can be matched) are used to guess later answers, so they save you clicks.
- **Same questions, different groups** appears when the same question was asked in separate columns to different people, for example two survey branches. For each set choose **Merge** to treat them as one question or **Keep separate**. Merge when the wording is identical, because one combined column gives one test with more people.
- **What is each question?** assigns each column a role through the **What is this?** choices, such as **Identifier (ID)**, **Group or cohort**, **Time point**, **Test question**, **Test total score**, **Survey (Likert) item**, **Background / demographic**, **Open-ended text** or **Ignore**. A small badge may say that Statversity guessed from your earlier projects, from your study answers, or filled it in from your survey file. A guess is only a starting point. Use **Jump to question…** to skip around, and the

Label (what results call it) box to give the column a readable name that will appear in tables and charts. When many look-alike questions appear, Statversity offers **Yes, tag all {n}**, **No, one at a time** or **Review the list**.

- **Kinds of answers** asks whether answers are categories, ordered categories, or numbers. This is the most important question for choosing a test. **Categories (nominal)** are labels with no order (major, yes/no). **Ordered categories (ordinal)** have an order but unequal gaps (strongly disagree to strongly agree). **Numbers (continuous)** are measured amounts (points on an exam). Averages only make full sense for numbers.
- **Groups to compare** asks which grouping questions you usually want to compare by. Ticked ones are offered first later; **None for now** skips this.
- **Knowledge questions**, **Answer choices**, **Test scoring** and **Question blocks from your survey** apply to quizzes with right answers. You mark which questions have a correct answer, load or build an answer key, and choose whether to add up 1s and 0s into a total. Without a key Statversity cannot score the quiz.
- **Scales** and **Scale scores** group several rating questions that measure one idea (say, five items about math anxiety) and combine them into a single score, either **Average of the answered questions (recommended)** or **Sum of the answers**. If an item is worded the opposite way, tick **Worded the opposite way (Statversity will flip its score)** so high always means the same thing.
- **New groups** lets you make a grouping variable from answers, for example “first-generation versus not”, with **Make a group variable**.
- **Summary** lists everything for review (**Change** next to any question) and offers the card **Want Statversity to plan your analyses?** with **Plan my analyses** or **Not now, finish**.

> **CAUTION**
>
> **Finish** applies all your answers at once. In review mode (opened from the Variables screen) the button reads **Apply changes**, and leaving with edits you have not applied asks whether to discard them.

The Test Advisor

The **Test Advisor** asks a few plain-language questions and then recommends a test. You reach it from **Analyze** (when you have no plan), from **Open the Test Advisor**, or from **Start another analysis** on a result.

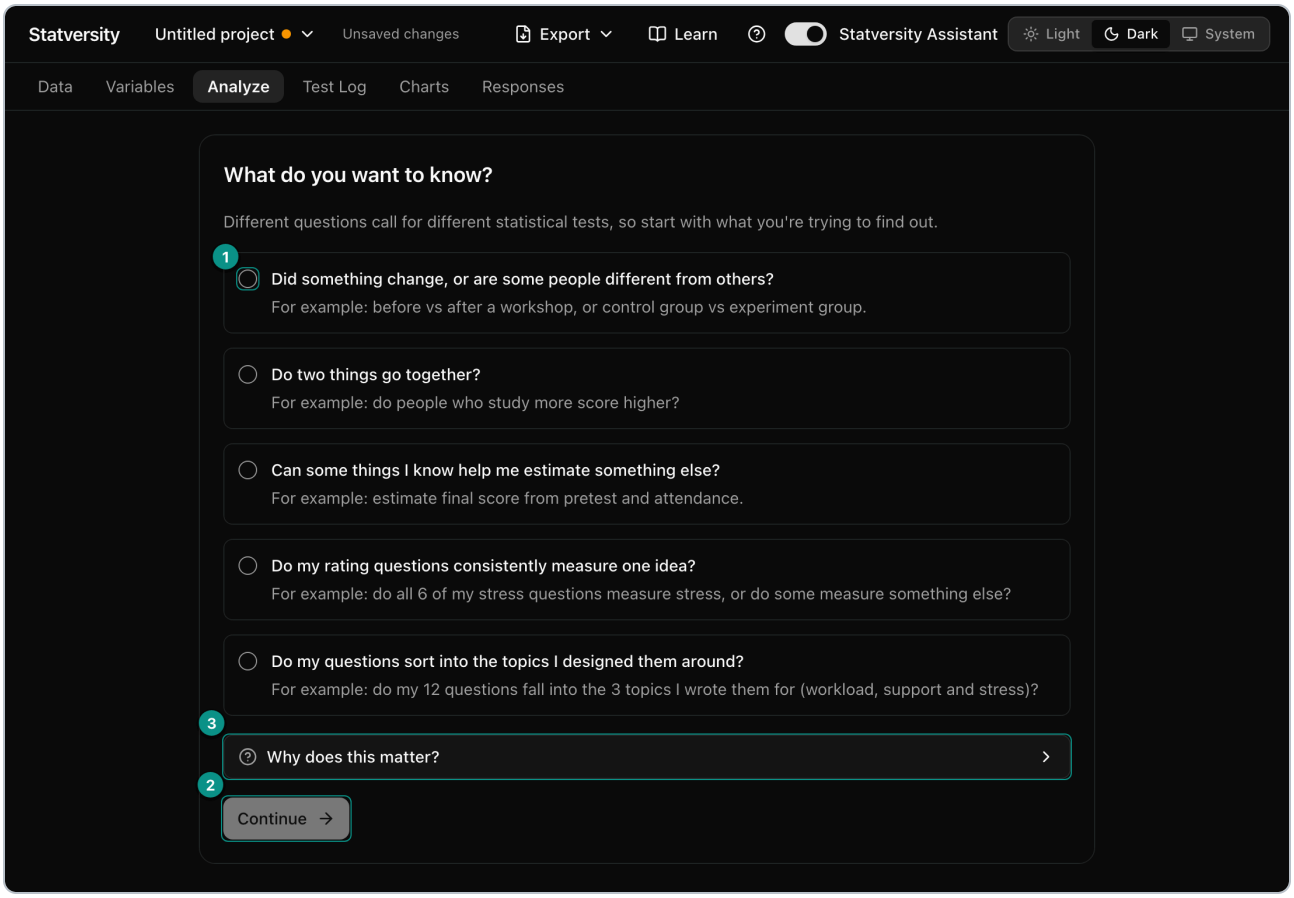


Figure 33. The first Test Advisor question.

Above the questions two optional menus help Statversity fill answers in for you. **Which score or answer do you want to analyse?** sets your outcome; picking a variable lets Statversity pre-answer questions from your data and shows a **From your data: …** summary. Leave it at **I'm not sure yet** if you prefer to answer everything yourself. For relationship questions a second menu, **Which other variable?**, appears with a **Change** link.

Number	Control	What it does
1	Answer option cards	Each option is one possible answer to the question. A tag **Suggested from your data** or **Statversity's usual choice** shows when Statversity pre-selected one.
2	**Continue**	Accepts the selected answer and moves to the next question (the button reads **Use this answer** when you are editing an earlier answer).
3	**Why does this matter?**	A collapsible panel that explains why the question matters for choosing a test.

The screenshot shows the first question, **What do you want to know?**, for the one-group pretest and posttest practice data with **Q3 (average score)** as the outcome. There is no **Back** button on the first question; it appears from the second question on.

Other controls on the screen: **Choose a test myself** jumps to the All tests menu, **Start over** clears the answers, and the **Your answers so far** list shows every answer with a **Change** button so you can fix an earlier one without starting over. If Statversity used its default for a question, a line **Statversity used its usual choice for…** with **Show the details** says which.

The kinds of questions you will be asked

The decision tree lives in a content file, so the exact wording can change with your data. In general the questions come in these families:

- **What do you want to know?** The first question. The choices are along the lines of “Did something change, or are some people different from others?”, “Do two things go together?”, “Can some things I know help me estimate something else?”, “Do my rating questions consistently measure one idea?” and “Do my questions sort into the topics I designed them around?”. This picks the family of test.
- **What kind of outcome do you have?** Whether you are analyzing a category (such as pass/fail), a single rating-scale item, or a score. Categories lead to counting tests, single rating items usually lead to rank-based tests, and scores lead to tests about averages.
- **Who was measured?** The next screenshot shows the question that settles this: whether you are comparing to a known value, different groups of people, the same people at different times whose answers you can match, or the same group at different times without matching.

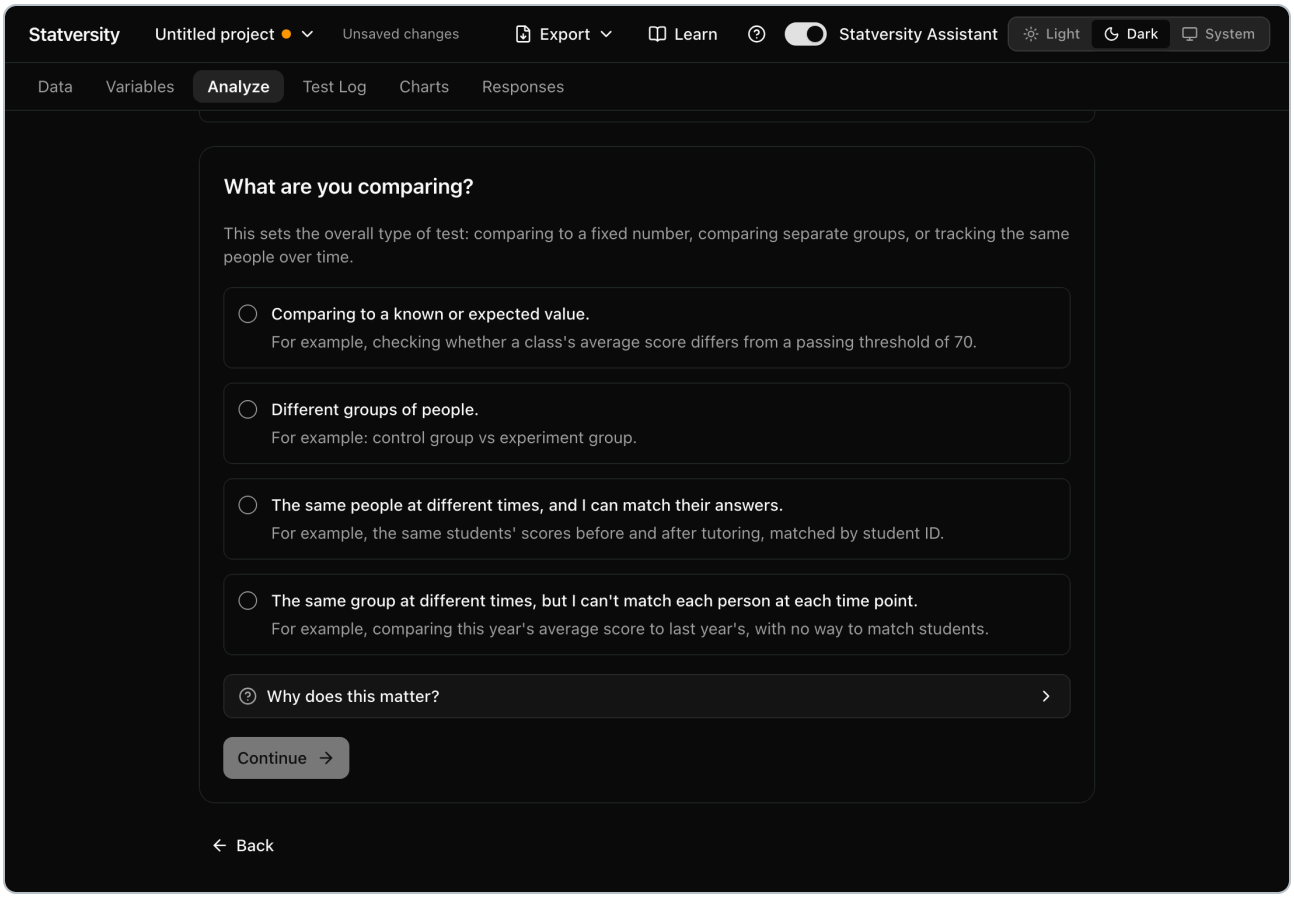


Figure 34. The linked or independent question.

- **How many groups or time points?** Two, or three or more. This chooses between, for example, a t test and an ANOVA, or between their rank-based alternatives.
- **Anything else to account for?** Whether you want to control for a third thing (a pretest score), have more than one grouping factor, or several outcomes together.
- **For relationships:** whether both are scores, one or both are ratings, or one is a yes/no, and whether the link depends on a third thing or works through one.

> **IN PLAIN LANGUAGE**
> "Linked" means Statversity can match each person's answer at one time to the same person's answer at another time, usually through an ID. Linked data allow a **paired** test, which compares each person to themselves. That is more sensitive than comparing two separate crowds, because it removes differences between people. If answers are anonymous, you must use a test for independent groups even if it was the same class both times.

The recommendation screen

When Statversity has enough answers it shows a card titled **Statversity recommends**.

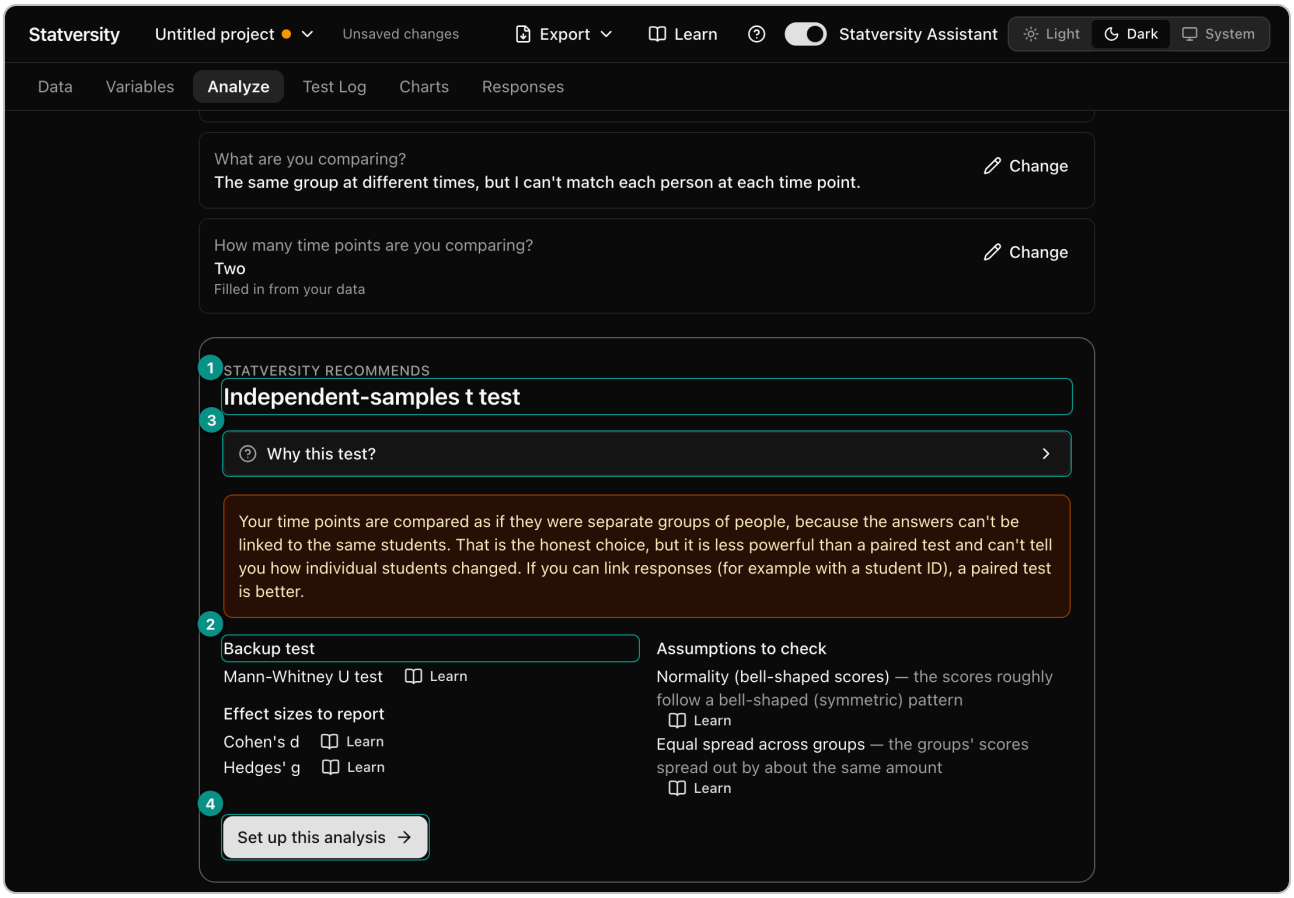


Figure 35. The recommendation card with alternatives.

Number	Control	What it does
1	Recommended test name	The test Statversity suggests for your question and data.
2	**Backup test**	The nonparametric alternative, used if the data fail the assumption checks (explained below).
3	**Why this test?**	A plain-language explanation of why this test matches your answers.
4	**Set up this analysis**	Opens **Choose your variables** for the recommended test.

The card also has rows for **Effect sizes to report** (a number saying how big the difference is, not just whether one exists), **Assumptions to check** (what Statversity will test before you rely on the result) and **Follow-up (post hoc) tests** (what to run if a comparison of three or more groups is significant, to find out which groups differ). Each row has a **Learn** button, and **Read the {page} guide** opens a longer explanation.

> **IN PLAIN LANGUAGE**
>
> A **nonparametric** test makes fewer demands on the shape of your data, usually by working with ranks (first, second, third) rather than raw scores. It is the safer fallback when scores are skewed or the sample is small. The cost is usually a little less sensitivity when the data were in fact well behaved.

Choose your variables

Choose your variables tells the test which column plays which part. You reach it from **Set up this analysis**, from **Start** on a plan item, or by picking a test directly. A back button at the top returns to where you came from (**Test Advisor**, **All tests** or **Your plan**).

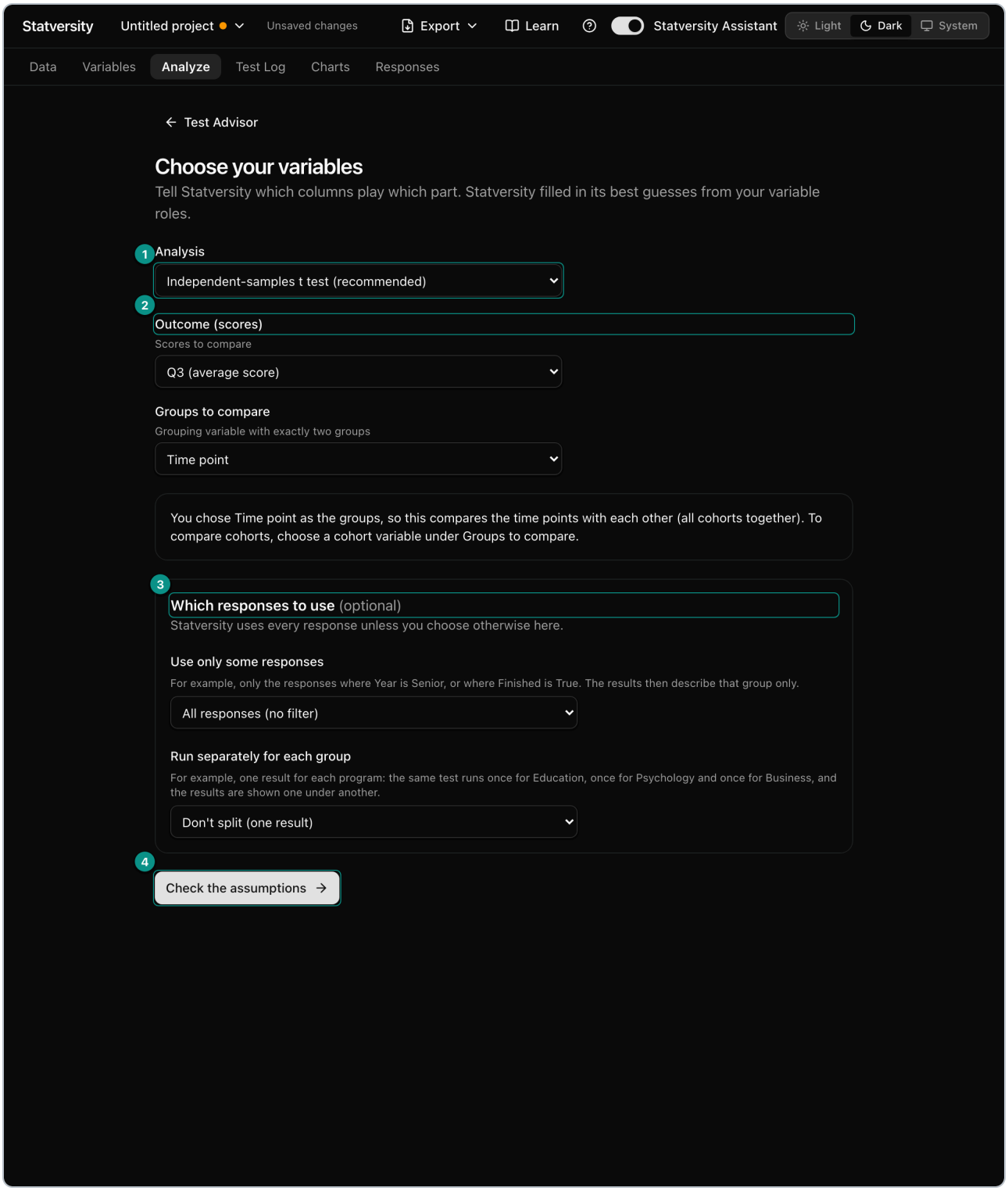


Figure 36. The Choose your variables screen.

Number	Control	What it does
1	**Analysis**	Switches the test. Options are tagged **(recommended)** or **(nonparametric alternative)**.
2	Role slots (for example the outcome and the groups)	Each slot takes one or more columns. Pick from the list that starts with **Choose a variable…**, or tick items and use the **Find** box.
3	**Which responses to use (optional)**	Limits the run to some rows with **Use only some responses**. Also holds **Run separately for each group**.
4	**Check the assumptions**	Moves on to the assumption checks.

The screenshot is the independent-samples t test for Q3 (average score) with **Time point** as the groups, reached from the Advisor recommendation, so it shows the **Outcome (scores)** and **Groups to compare** slots. A **How is your data laid out?** question (**One column per time point** or **One row per time point (stacked)**) appears only for repeated-measures layouts, so it is not shown here.

If you came from a plan, a notice says **Filled in from your plan; change anything you like.**

What a role means

A role is a job description for a column. The exact slot names come from the engine for each test, so they vary from one test to the next.

Test family	Typical roles	What each means
Paired or one-group before/after	Two measurements (before and after), or one score plus time	The two numbers measured on the same people. Statversity subtracts them person by person.
Independent-samples t test, Mann-Whitney	Outcome, Groups	The outcome is the number you compare; Groups is the column that says who belongs to which of two groups.
One-way ANOVA, Kruskal-Wallis	Outcome, Groups	Same, with three or more groups.
Repeated-measures and mixed ANOVA	Outcome (one column per time point), optionally a between-person grouping	The score at each time, plus the group each person belongs to.
Correlation	Two variables	The two columns you want to relate. Order does not matter.
Regression	Outcome, Predictors	The outcome is what you want to estimate; predictors are the things you use to estimate it.
Chi-square	Two categorical variables	The two category columns whose counts you cross-tabulate.
Reliability (Cronbach's alpha)	Items	The set of questions that should measure one idea.

Two special inputs appear for some tests: **Compare the average against this value** for one-sample tests, and **Highest possible score** for normalized gain.

With time points you also see **What do you want to compare?**, with the same four choices as in the plan. Under **Run separately for each group** the setting **Don't split (one result)** is the default; choosing a grouping variable instead runs the same test once per group and creates one Test Log entry for each. Links such as **Read more about running several tests** and **About moderation analysis** open the Learn pages.

Assumption checks and the decision

Most tests produce trustworthy answers only if the data meet certain conditions, called **assumptions**. Clicking **Check the assumptions** walks you through them one at a time.



Figure 37. The assumption checks screen.

Number	Control	What it does
1	Check heading (**{check}: {what it applies to}**)	Names the check and the data it examines, such as a group or the pre/post differences.
2	Explanation blocks (**What it is**, **Why it matters**, **What Statversity checked**, **What it means for you**)	Four short sections explaining the check in everyday language.
3	Verdict badge (**Looks fine**, **Caution**, **Concern**)	Summarizes the result at a glance.
4	**Next check** / **Decide which test to use**	Moves to the next check, and from the last check to the decision.
5	**Check them all and choose a test**	Skips ahead to the decision screen, which lists every check.

The screenshot is the first of three checks, **Normality: Pre**, for the independent-samples t test in the previous screenshot.

Use **Back** to return to a previous check, and **Read more about {topic}** for a Learn page.

The common checks, one sentence each

- **Normality (Shapiro-Wilk, with Q-Q plots and histograms):** are the scores roughly bell-shaped, in each group, or in the before/after differences for a paired test?
- **Equal spread (Levene's test):** do the groups vary about as much as one another around their own averages?
- **Sphericity (Mauchly's test):** in repeated measures, are the differences between each pair of time points about equally variable?
- **Outliers:** are any scores so far from the rest that they could pull the result by themselves?
- **Linearity and multicollinearity (regression):** is the relationship roughly a straight line, and are your predictors so similar to each other that their separate effects cannot be told apart?
- **Other design checks:** such as equal slopes for ANCOVA or equal covariance patterns for MANOVA, which appear only for those tests.

> **CAUTION**
> Statistical tests for normality become over-sensitive with large samples: they flag tiny, harmless departures. Statversity tells you this and shows the plots, so look at the picture as well as the verdict.

The decision

After the last check, the decision screen **Which test should you use?** shows a list titled **Assumption checks** with each verdict and explanation, and two options: **Use the recommended test: {name}** and **Use the nonparametric alternative: {name}**. The one Statversity suggests carries a **Suggested** badge.

You make the final choice; Statversity does not run anything by itself.

What Statversity does when a check fails:

- It does not stop you. A **Caution** or **Concern** badge changes which option is suggested, usually to the nonparametric alternative, and the explanation says why.
- You can still pick the recommended test. Statversity then shows the result and records the assumption verdicts in the Test Log and in the **Assumption checks** section of the result, so a reader of your report can see them.
- When a correction exists (for example a Greenhouse-Geisser correction for a sphericity concern, or a Welch version for unequal spread), Statversity reports it with the result.

Click **Run and see the results** to run the chosen test and open the Results screen. **Back** returns to the last check.

> **TRY IT YOURSELF**
> Open the practice data from chapter 3, click **Analyze**, answer the Advisor with the suggested answers, and stop at the recommendation. Read **Why this test?** and write one sentence in your own words saying why Statversity did not suggest a different test.

CHAPTER 8

Results

The **Results** screen is where an analysis ends and your write-up begins. It is a long page, built to be read from top to bottom: first a plain-language summary, then text and a table you can paste into a paper, then a chart, then the supporting numbers. This chapter goes through it in that order, then covers the Test Log, which keeps every result you have ever run.

You reach a result by finishing an analysis, by clicking a line in the Test Log, or by clicking a finished item in the plan sidebar. If you open Results with nothing selected, you see **No analysis selected.**

> **IN PLAIN LANGUAGE**
> A statistical result is a set of numbers plus a decision rule. The numbers describe your data; the rule (such as “p below .05”) tells you whether the pattern is bigger than chance would usually produce. Statversity prints both the numbers and a sentence that says what they mean, so you never have to translate a table of symbols on your own. Chapter 23 explains p-values and effect sizes in depth.

The top of the page

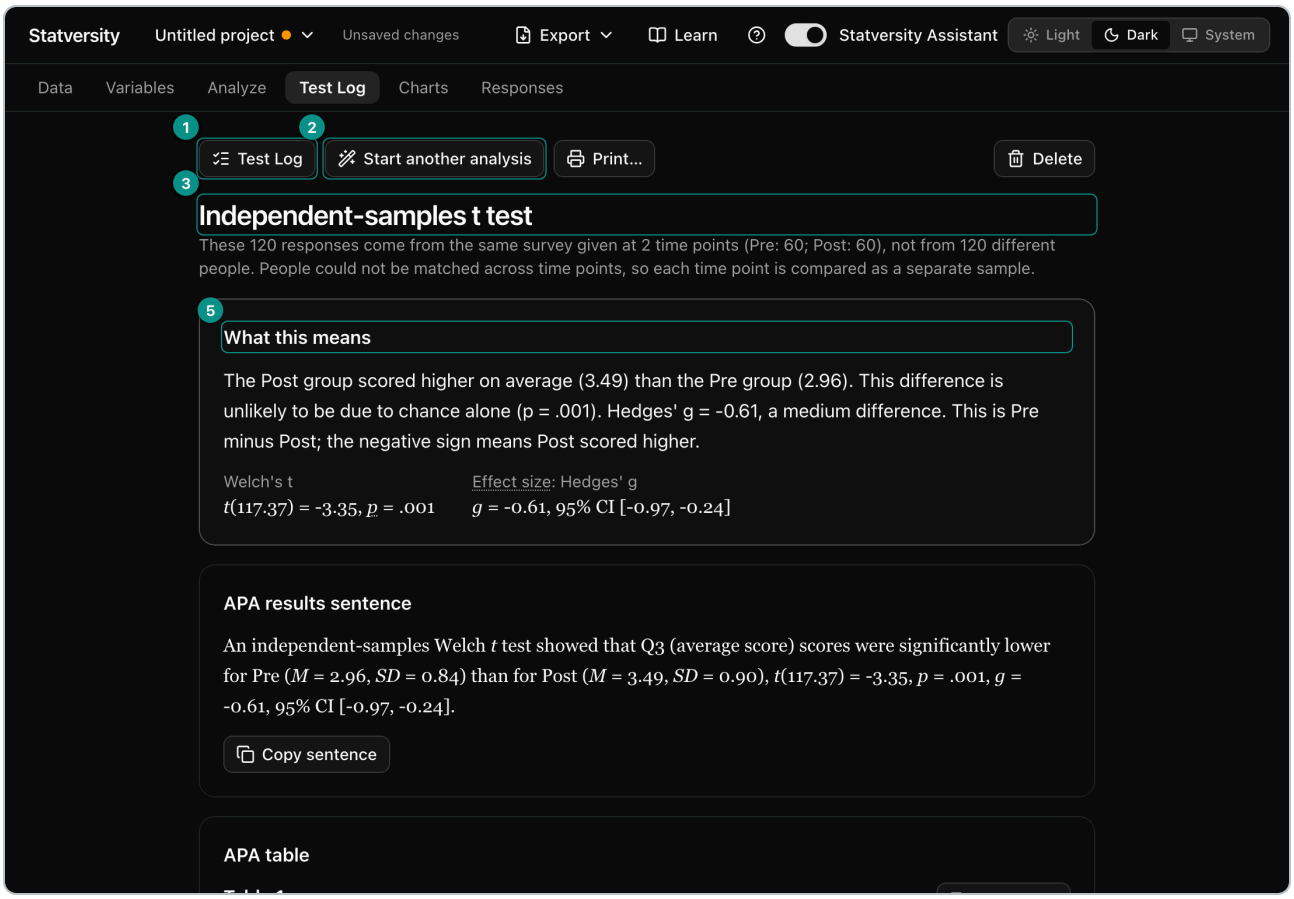


Figure 38. The top of a Results page.

Number	Control	What it does
1	**Test Log**	Goes to the list of every finished test.
2	**Start another analysis**	Opens the Test Advisor (Assistant on) or the All tests menu (Assistant off).
3	Title	Names the test that was run, for example the type of t test or ANOVA, so you always know which analysis you are reading.
4	N line	Says how many people (rows) the test used, and sometimes how many were left out for missing answers.
5	**What this means**	A plain-language box explaining the finding.

The screenshot is an independent-samples t test of Q3 (average score), Pre against Post, with the Assistant on. The N line reads “These 120 responses come from the same survey given at 2 time points (Pre: 60; Post: 60), not from 120 different people. People could not be matched across time points, so

each time point is compared as a separate sample." Statversity says this because the Pre and Post files were stacked without a student ID, so it cannot tell which Pre answer belongs with which Post answer; it counts responses, not people. When the dataset is not stacked by time, or the comparison is not by time point, the line simply reads "n responses analysed" with the per-group counts in parentheses.

Beside those, the top right of the page also has **Print…**, which prints this single result, and a red **Delete**, which removes the test from the Test Log after a confirmation (**Delete this test?** with **Cancel** and **Delete**).

The **N line** matters because a test's power depends on how many people it used. If your data had 40 rows but only 32 people answered both questions, the test used 32. Check that number against what you expected. A much smaller N than expected usually means missing answers or a filter you forgot.

Notices above the result

Depending on how the test ran, a colored notice may appear above the sections:

- **Your data has changed since this analysis was run.** with a **Run again** button. The result is stale; see "Out of date" below.
- **Open the newer run** when a later re-run of the same analysis exists.
- **See every {split variable}** when this is only one group of a test you ran separately for each group.
- **Run {backup} instead** (Assistant off only) and **Run Fisher's exact test instead**, which suggest a more suitable test. These are covered in their own section at the end of this chapter.

What this means

The **What this means** box is the part to read first. It states in everyday words what Statversity found: which groups differed or did not, in which direction, and roughly how big the difference is. It avoids symbols on purpose.

> **IN PLAIN LANGUAGE**
> Read it as a translation, not as the final word. It tells you what the numbers say about your sample. It does not tell you why the difference exists, whether it matters in practice for your students, or whether it will repeat with a new sample. Those are judgments for you and your reader. In particular, "not significant" does not mean "no difference"; it means this study could not rule out chance as the explanation.

Below it you may see a **How to report this** section with guidance on what to write, and a **Learn more about the {test}** button that opens the full Learn page for the test.

APA sentence and APA table

Scientific papers in psychology, education and related fields follow the APA style for reporting statistics. Statversity writes both pieces for you.

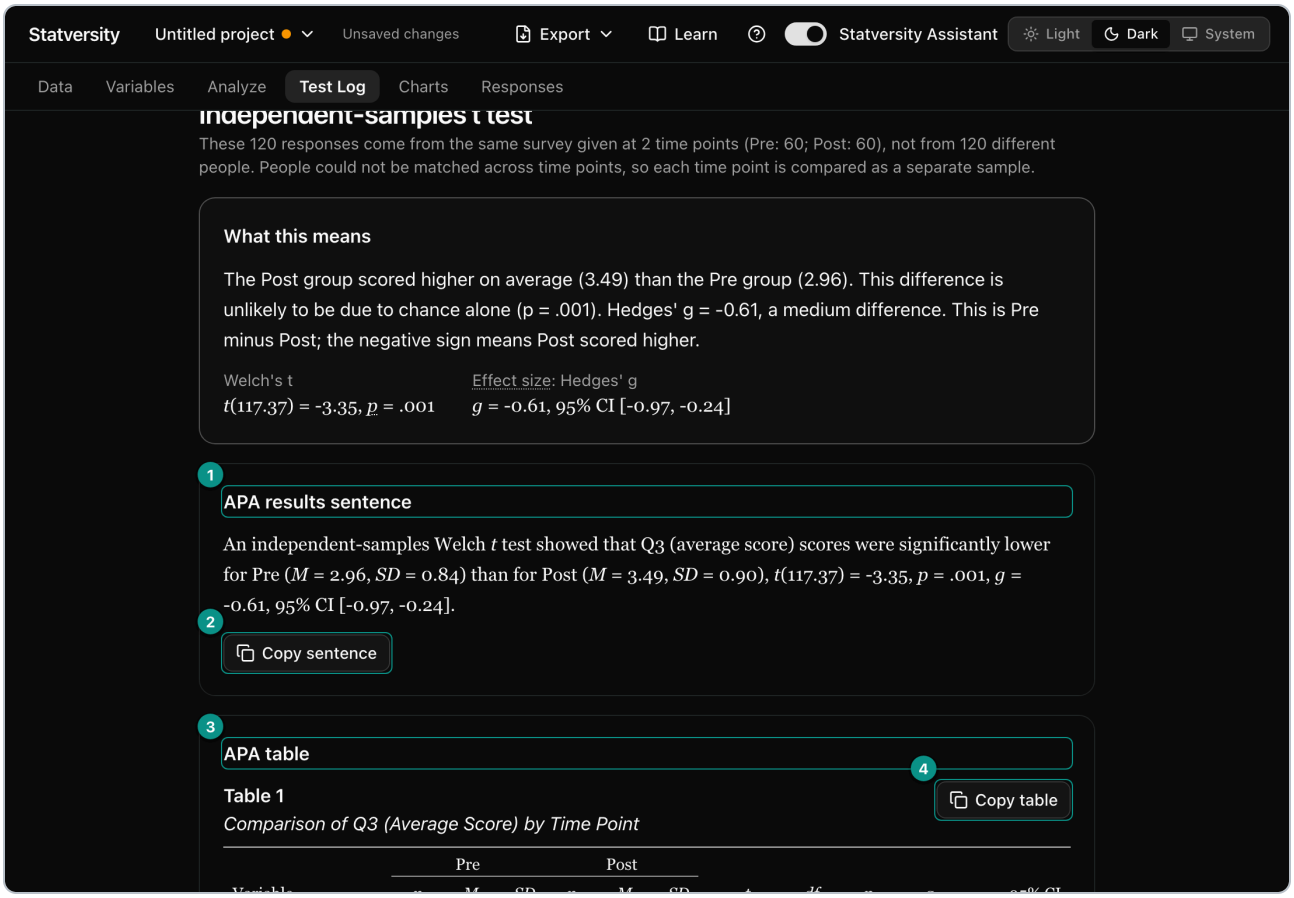


Figure 39. The APA sentence and APA table with their copy buttons.

Number	Control	What it does
1	**APA results sentence**	One sentence in APA format with the test statistic, degrees of freedom, p value and effect size, ready to drop into a Results paragraph.
2	**Copy sentence**	Copies that sentence to the clipboard. The button changes to **Copied** for a moment.
3	**APA table**	The same results laid out as an APA table.
4	**Copy table**	Copies the table to the clipboard, ready to paste into Word, Excel or Google Docs.

Paste the sentence into your document and check the italics: APA uses italic letters for symbols such as *t*, *F* and *p*, and some programs drop the italics when you paste. Whatever you paste, make sure the numbers still match the screen.

> **CAUTION**
> Statversity writes the sentence for the test you ran. If you change the data afterward or rerun with different settings, copy it again. A sentence copied from an out-of-date result does not describe your current data.

Where a test has follow-up comparisons (for example which groups differ after an ANOVA), a follow-up section appears with its own table. An **About {follow-up method}** button opens a guide to that method.

The comparison figure

Every t test, ANOVA and rank test that compares groups or time points draws a **Comparison figure** under the result. It is a bar chart of the groups with a **bracket** over each pair that differs significantly, so a reader sees at a glance which groups are different, not just that “something” is.



Figure 40. The Comparison figure card for a one-way ANOVA with three groups, numbered 1 to 5.

Number	Control	What it does
1	**Save figure…**	Saves the figure as a PNG, SVG or PDF, with the brackets and the note included.
2	**Edit chart**	Opens the chart editor for this figure (see below).
3	A bracket	Joins the two bars it compares. The stars above it say how strong the evidence is. Here Control against Intervention B has three stars, the strongest.
4	The figure note	Says what the error bars show, what the stars mean and which adjustment was used for the p values (here Tukey).
5	**Show the numbers**	Expands a table of the plotted values.

What the bars show depends on the test:

- **Tests about means** (t tests, ANOVA): each bar is the group mean, with an error bar. The default is the 95% confidence interval.
- **Rank tests** (Mann-Whitney, Wilcoxon, Kruskal-Wallis): each bar is the group **median**, with the **interquartile range** (25th to 75th percentile) as its error bar. Rank tests compare ranks, not means, so the figure shows the middle score and spread that match what was tested.

The star key is the same everywhere: one star (*) means $p < .05$, two (**) mean $p < .01$ and three (***) mean $p < .001$. When a test compares more than two groups, the stars come from the adjusted p value of the follow-up comparisons (Tukey, Games-Howell, Holm, Dunn and so on), never from the plain p value of each pair. A test that compares only two groups gets one bracket, labelled from the test's own p.

Two-factor results (a factorial ANOVA with two between factors, or a mixed ANOVA with groups and time points) show **two** comparison figures, each under its own follow-up section. **Comparison figure: within each {factor}** brackets the changes across the other factor inside each group, and **Comparison figure: at each {factor}** brackets the groups against each other at each point. The headings carry your own variable names, for example "within each Group" and "at each Time point". Chapter 20 walks through both for a mixed ANOVA.

By default only the pairs that differ significantly get a bracket. A figure full of "ns" labels would bury the finding. If you want to show every pair, use the option below.

> **IN PLAIN LANGUAGE**
> A bracket is Statversity's way of saying "these two bars are really different, not just different by chance". No bracket means the data did not show a clear difference for that pair. It does not prove the groups are the same.

Edit chart opens the chart editor. Two options under **Chart options** are specific to this figure:

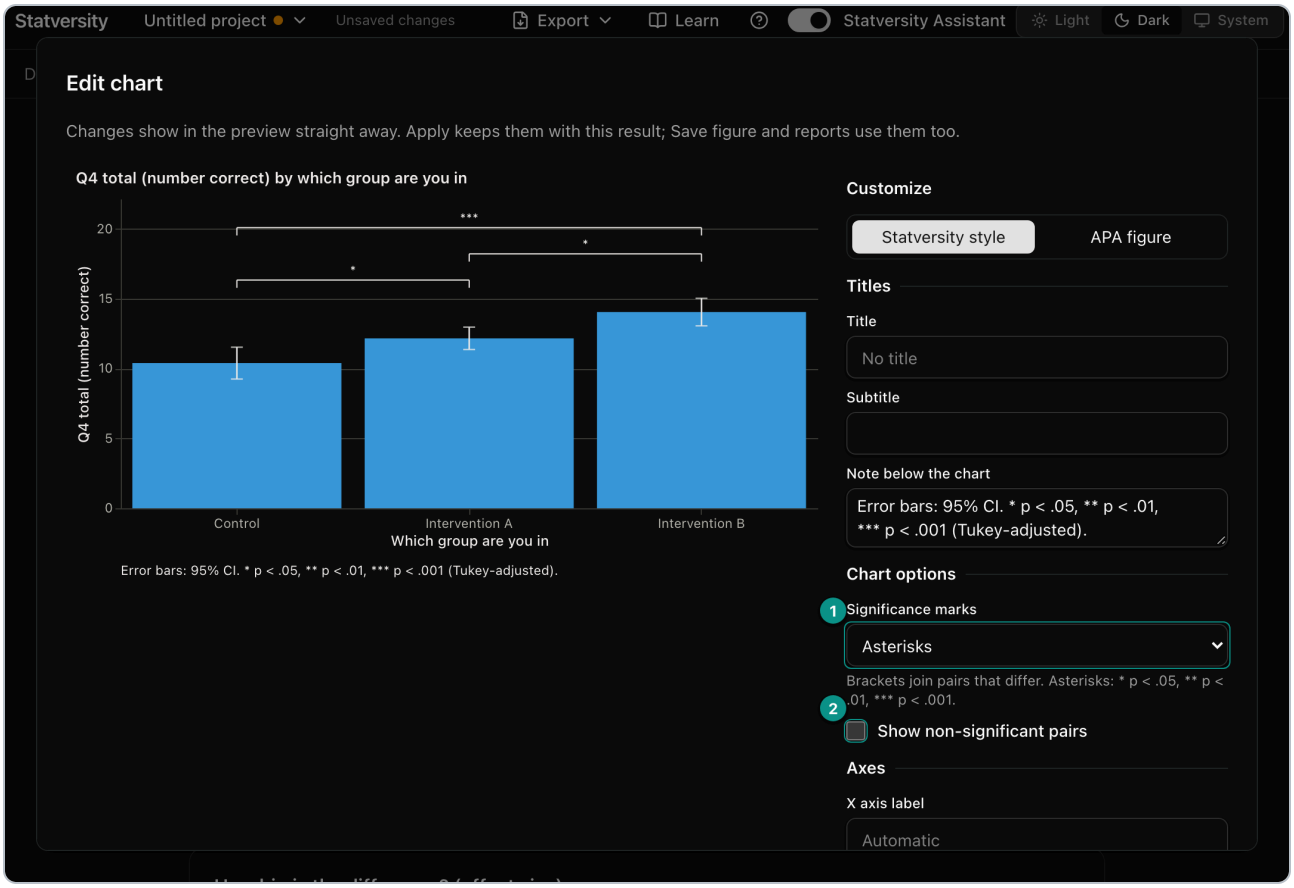


Figure 41. The Edit chart dialog with the Significance marks select and the Show non-significant pairs checkbox.

Number	Control	What it does
1	**Significance marks**	**Asterisks** (default), **Exact p-values** (such as p = .003 or p < .001) or **Off** (no brackets).
2	**Show non-significant pairs**	Also draws the pairs that do not differ, labelled **ns**. Unavailable when marks are **Off**.

The other options are the usual ones from chapter 9 (title, note, error bars, colors, size). **Save figure…**, exported reports and printing all use the figure as you edited it, and reports and print include it automatically.

The chart and Edit chart

Many tests draw a chart that matches the result, such as a bar chart with error bars, a box plot, an interaction plot or a path diagram for mediation. The independent-samples t test used in this chapter's screenshots shows its figures inside **Assumption checks** instead: a Q-Q plot and a histogram for each group.

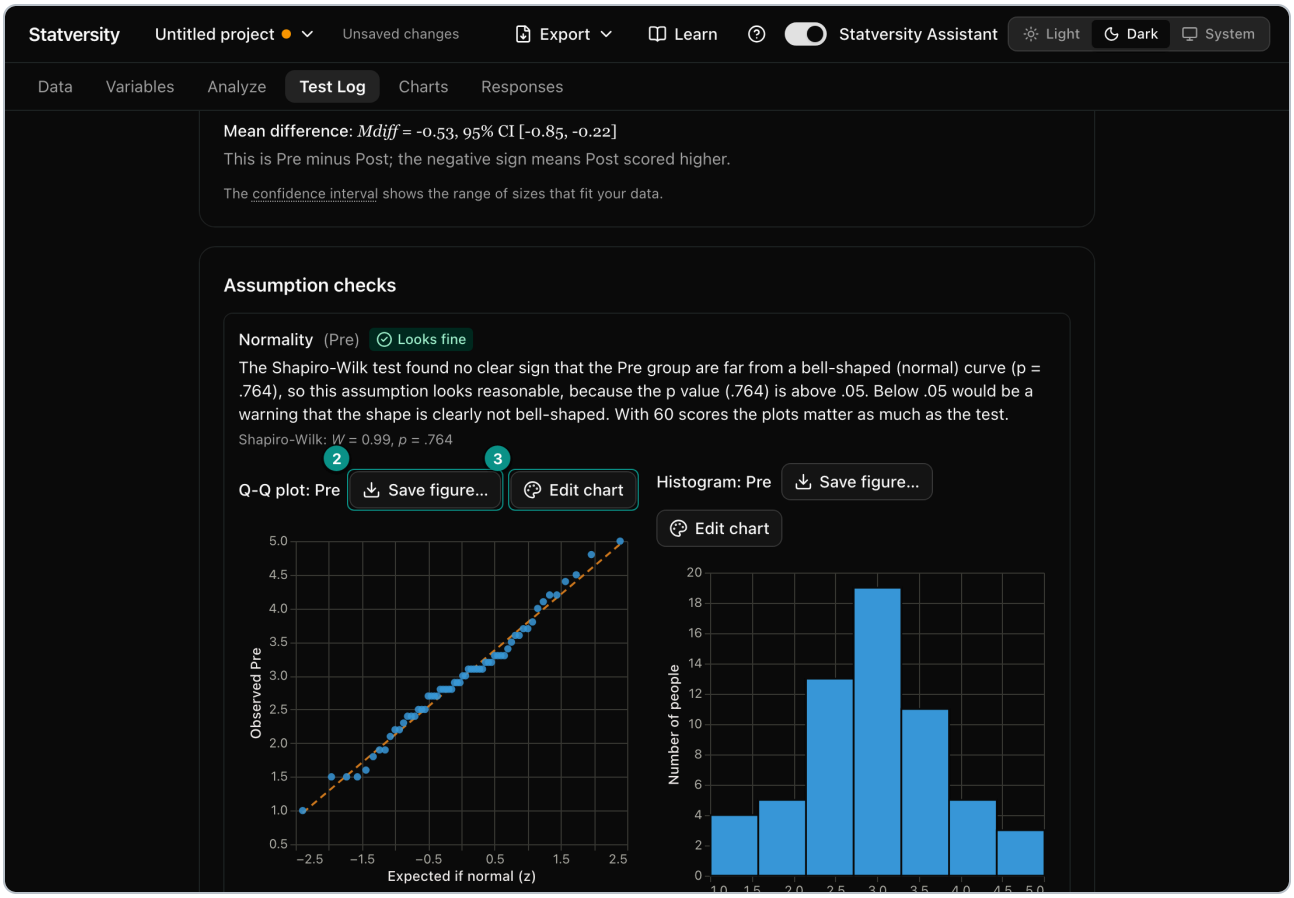


Figure 42. A Q-Q plot from the Assumption checks section, with its Save figure and Edit chart buttons.

Number	Control	What it does
1	The chart	Here, the Q-Q plot for the Pre scores. A chart is a picture of the same data the test used, drawn in APA figure style by default.
2	**Save figure…**	Opens a menu to save the picture as a PNG at three sizes, as an SVG (vector), or as a PDF.
3	**Edit chart**	Opens the chart editor for this result. A tag **Edited** appears once you have changed something.

The chart editor has the same look-and-feel options as the Chart builder, which chapter 9 lists, but without the data options: you can change the title, notes, error bars, colors, font, size and more. **Reset to default** undoes your edits, **Cancel** closes without keeping them, and **Apply** keeps the look with this result. **Save figure…** and exported reports use the edited version.

Detailed statistics

Below the chart come the sections that support the sentence and table. Which ones you see depends on the test.



Figure 43. The detailed statistics sections of a result.

Number	Control	What it does
1	Effect-size section (here **How big is the difference? (effect size)**)	The size of the effect with a confidence interval; here Hedges' g, Cohen's d and others for the Pre and Post difference. The section title varies with the test. An **About {effect size}** button explains it.
2	**Assumption checks**	A summary of the checks that were run and their verdicts, so a reader can see you checked them. Here: normality for Pre and Post, and equal spread.
3	**Descriptive statistics**	The basic numbers for each group or variable: how many people, the average, and the spread.
4	**How to report this**	A short note on how to write the result up.

The screenshot is a tall capture of the same t test, so it shows the sections in page order, which is the order of the numbers.

> **IN PLAIN LANGUAGE**
>
> An **effect size** says how big a difference or relationship is, in a way that does not depend on sample size. A very large sample can make a tiny, useless difference "significant", and a very small sample can hide a large one. Reporting both the p value and an effect size gives a fuller picture. The **confidence interval** is a range of values that are plausible for the true effect, given your data.

For tests run separately for each group (**Run separately for each group** on **Choose your variables**), the page shows a section per group. When time points are involved, a button such as **Now run: each group over time** appears so you can run the next comparison in the same family. A **What changed since run # {n}** section appears after a re-run and compares the earlier and later numbers in a table.

The Test Log

The **Test Log** (the tab or the button at the top of a result) is your record of every test you have run. It is saved with the project, so closing and reopening the project does not lose it. Its intro reads: "Every test you run is kept here, with its inputs and results, and saved with the project."

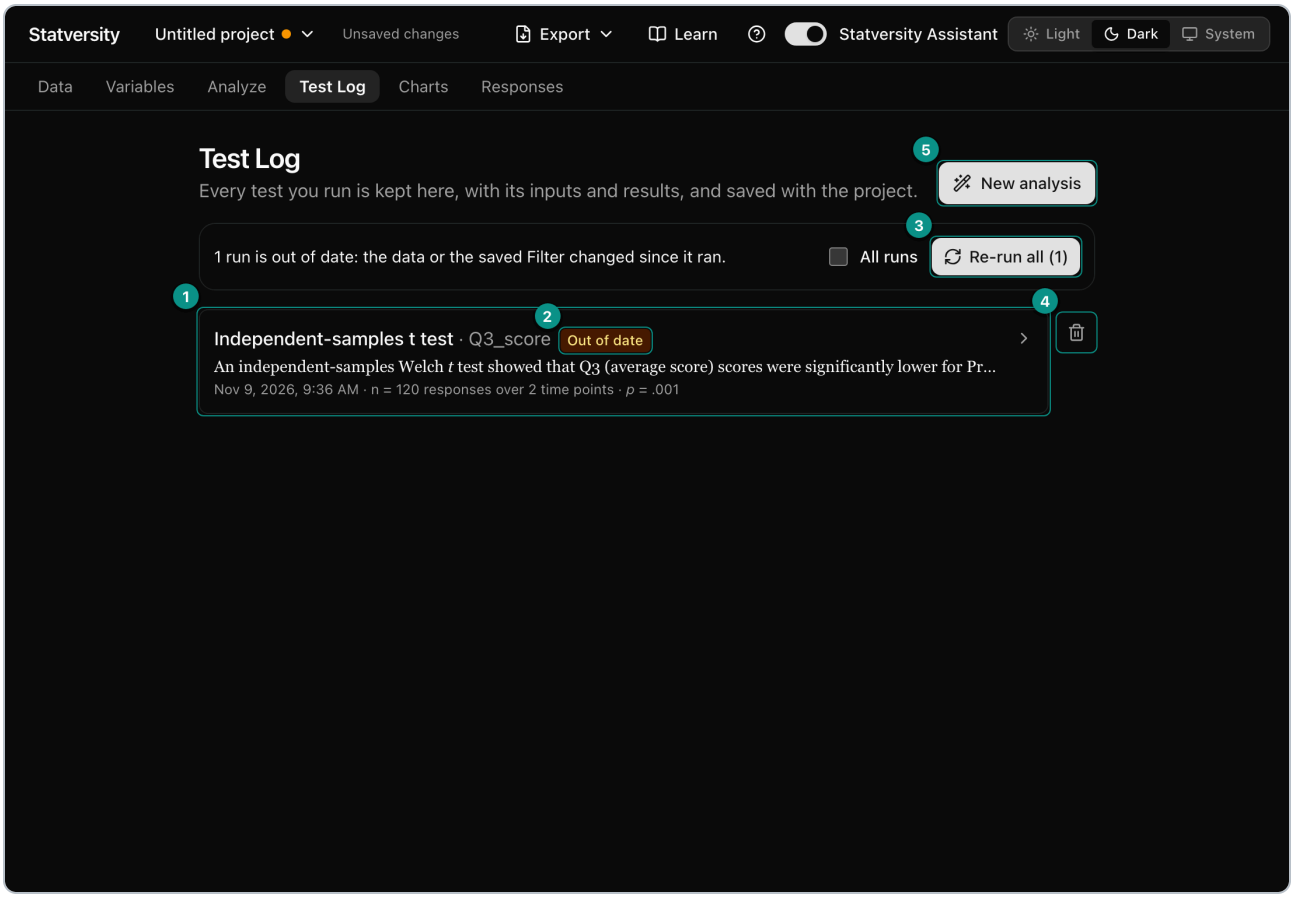


Figure 44. The Test Log with a row, rerun controls and an Out of date badge.

Number	Control	What it does
1	A log row	Shows the test and what it was run on. Clicking it opens that result.
2	**Out of date** badge	Marks a run whose data have changed since it ran.
3	**Re-run all ({n})**	Re-runs every out-of-date test in one go.
4	**Delete this test** (icon on the row)	Removes that test after a confirmation (**Delete this test?**).
5	**New analysis**	Opens the Advisor or All tests.

The screenshot was taken after one data cell was edited, which made the single run out of date. The **Group tests** button, which starts a test family, appears only once the log holds more than one test, so it is not shown here. There is no per-row re-run button; re-running is done with **Re-run all** or **Run again** on a result.

Other Test Log controls: the expander **Running several tests? Why grouping them matters**; the checkbox **All runs** in the re-run bar and **Cancel** (or **Stopping…**) while it runs; **Show superseded runs ({n})** / **Hide superseded runs**, which reveals old versions that a re-run replaced; and family cards with

a **Correction** menu (**Holm**, **Bonferroni**, **Benjamini-Hochberg (FDR)**), **Edit** and **Ungroup**. If you have no tests yet the page says **No analyses yet. The Test Advisor will help you choose one.**

> **IN PLAIN LANGUAGE**
> Every test carries a small chance of a false alarm. The more tests you run on one dataset, the more likely at least one will look significant by luck alone. Grouping related tests into a **family** lets Statversity adjust their p values so the family as a whole keeps a fair false-alarm rate. The dialog **Group tests into a family** asks for a **Family name**, which tests belong, and a **Correction method**.

Out of date, and why it appears

The **Out of date** badge means the data or the saved Filter changed after the test ran, so the result no longer describes what is in the project. Typical causes are editing a cell, deleting duplicate rows, adding a new calculated variable that a test uses, or switching a Filter on or off. The bar at the top reads **Every run is up to date with your data.** when all is well, or says how many runs are out of date.

You do not lose anything. The old result stays as a record. To bring it up to date, use **Run again** on the result or its notice, or **Re-run all ({n})** on the Test Log. The new run appears as a re-run of the earlier one, the older run is hidden unless you choose **Show superseded runs ({n})**, and the result shows a **What changed since run #{n}** table comparing old and new.

> **TIP**
> If you make a lot of data edits, wait until you are done and use **Re-run all** once, rather than rerunning each test as you go.

Start another analysis

Start another analysis at the top of a result leads back to the place where you choose a test: the Test Advisor with the Assistant on, the All tests menu with it off. When you are working through a plan, the footer of the result instead shows **Next analysis** and **Stop for now** (or **Back to the plan**), and the result is marked done in your plan sidebar.

Run {backup} instead

When the Assistant is off, Statversity does not walk you through the assumption checks before you run. Instead, if the assumptions look doubtful after the run, the results page shows a notice with a button named **Run {backup} instead**, where {backup} is the name of the more suitable test, such as a rank-based alternative. Clicking it runs that test on the same variables and opens its result as a new Test Log entry; your first result stays in the log.

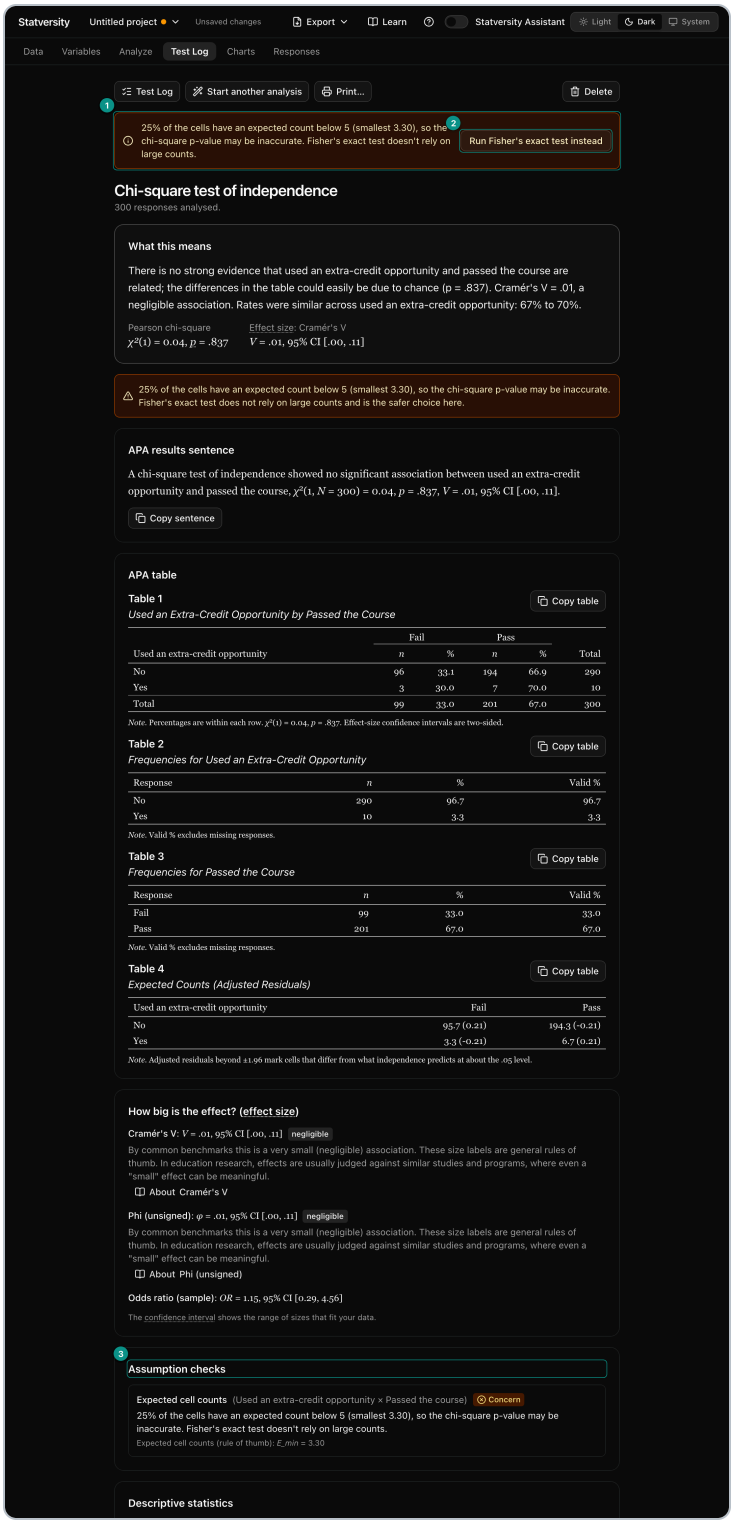


Figure 45. A result with the Fisher's exact test notice, from categorical_outcomes with the Assistant off.

Number	Control	What it does
1	The doubtful-assumption notice	Explains in plain words which assumption looked doubtful.
2	**Run {backup} instead**	Runs the suggested backup test on the same variables.
3	**Assumption checks**	Shows the verdicts behind the notice.

The screenshot is a chi-square test of independence on the categorical_outcomes practice data (Q4 by Q3) with the Assistant off. Because a quarter of the cells have small expected counts, the button reads **Run Fisher's exact test instead**. For other tests the button reads **Run {backup} instead**.

> **CAUTION**
>
> The backup is a suggestion, not a verdict. If you run both tests and they agree, report the one that suits your data and say you checked. If they disagree, read the assumption checks and the "What this means" box of both before deciding, and report your choice and why.

Chapter 10 returns to this button when it explains working without the Assistant.

CHAPTER 9

Charts, Responses, Study Planner, Export and Print

This chapter covers the parts of Statversity that sit beside the analysis flow: building your own charts, reading written answers, planning a study before you have data, and getting results out of Statversity as files or paper.

Charts

The **Charts** tab holds figures you build yourself, separate from the chart that appears automatically on each result.

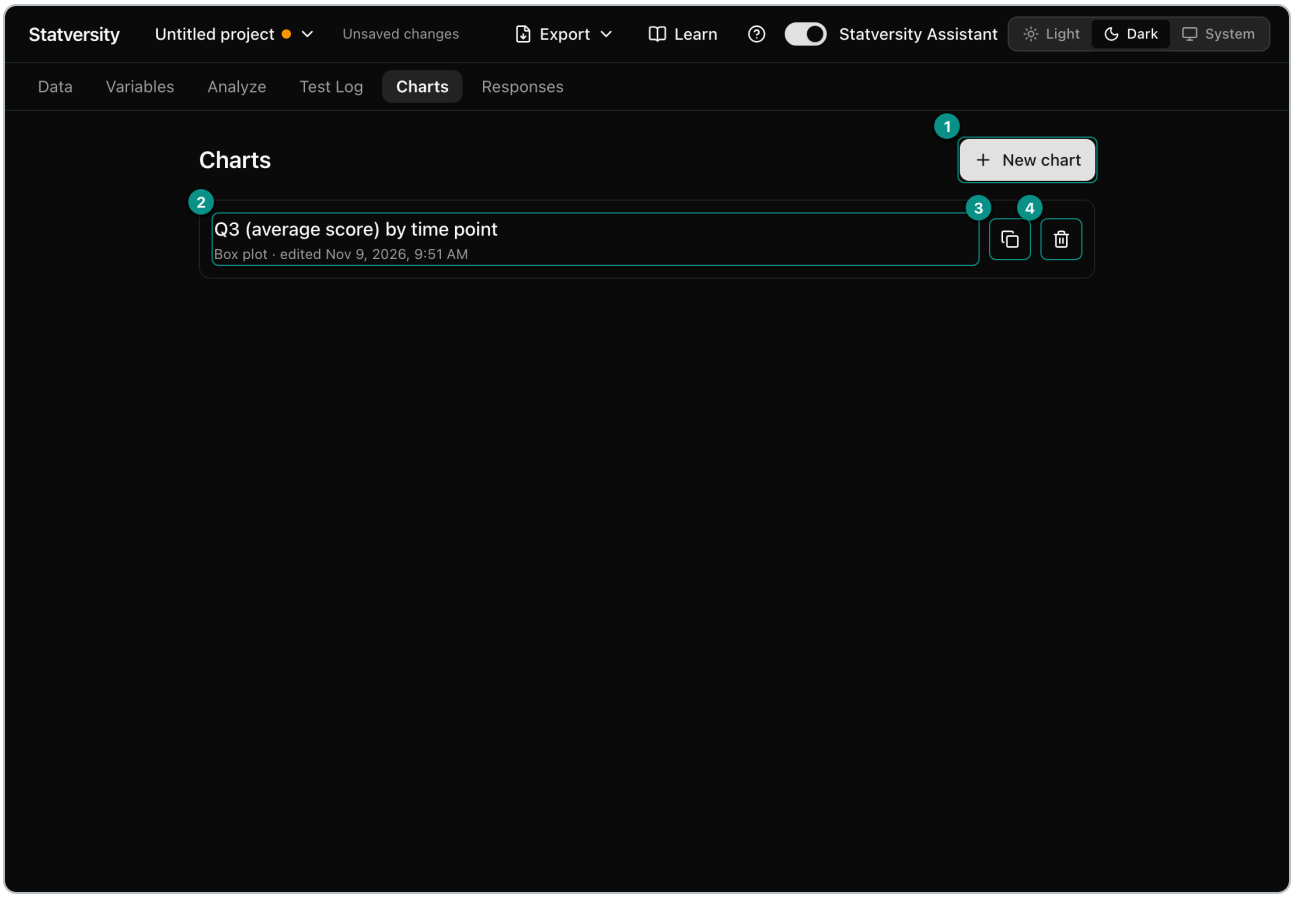


Figure 46. The Charts list.

Number	Control	What it does
1	**New chart**	Opens the Chart builder with an empty chart.
2	A chart row	Shows the chart name, its type, · **APA figure** if it uses that style, and · **edited {date}**. Click to open it in the builder.
3	**Duplicate {name}**	Makes a copy to adjust without touching the original.
4	**Delete {name}**	Removes the chart from the project.

The screenshot shows one saved chart, a box plot named “Q3 (average score) by time point”.

When you have no charts the page says **No charts yet** and explains that you can build a bar chart, box plot, Likert chart and more.

The Chart builder

Click **New chart** or a chart row to open the builder. Everything here updates a live preview.

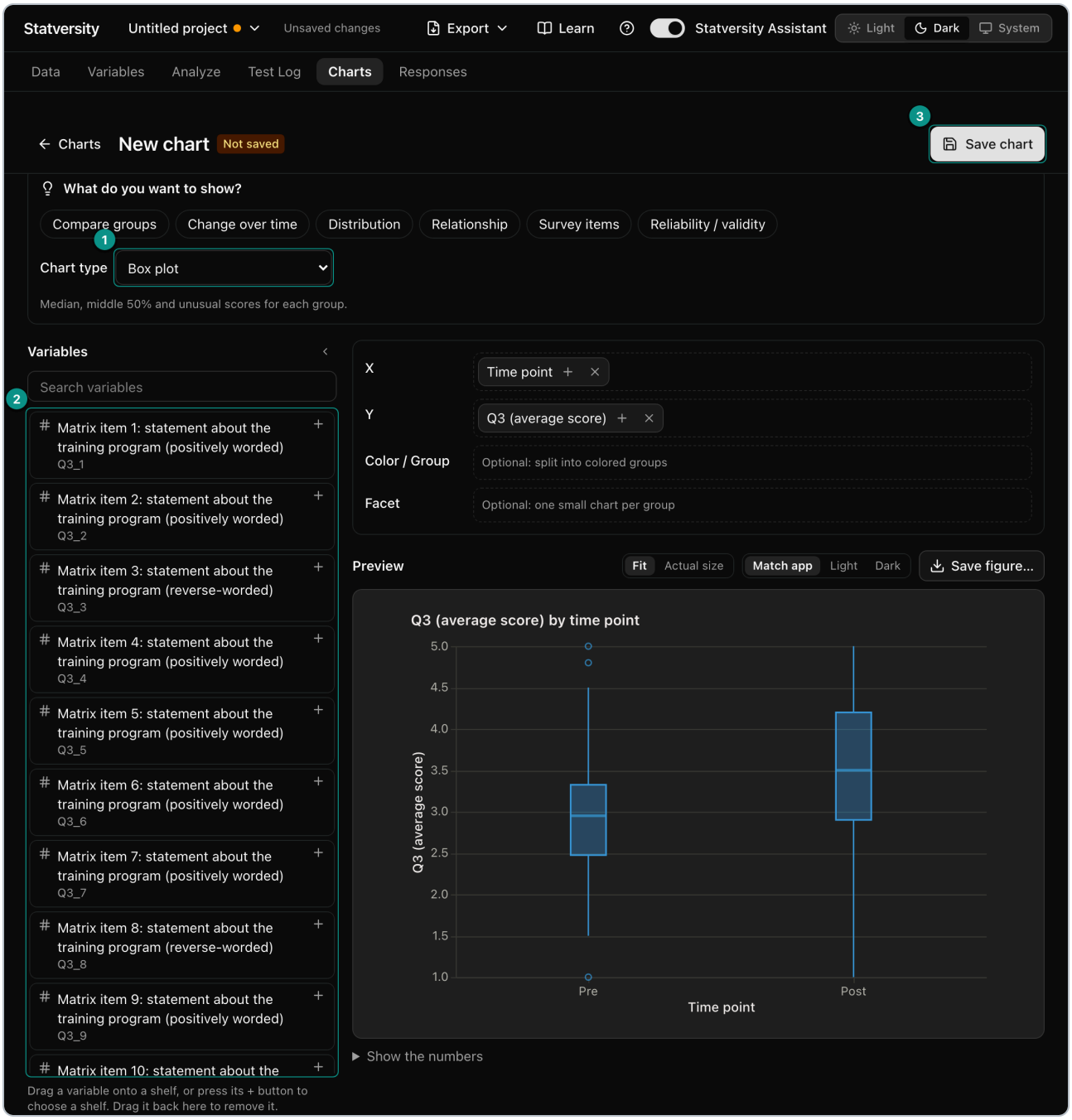


Figure 47. The Chart builder.

Number	Control	What it does
1	**Chart type**	The kind of chart to draw.
2	**Variables** list	The questions you can plot. Click one (the **+** beside it) to put it on a shelf: **X**, **Y**, **Color / Group** or **Facet**. The shelves hold the variables you plot; here **Time point** is on X and Q3 (average score) is on Y.
3	**Save chart**	Adds the chart to the project so reports can use it.

Controls from top to bottom:

- **Charts** (back arrow) returns to the list. If you have unsaved edits it asks **Save this chart?** with **Cancel**, **Don't save** and **Save chart**.
- The chart name sits at the top with a badge, **Not saved** or **Saved in project**. After **Save chart** you see the message **Chart saved to the project.**
- **What do you want to show?** offers six goal buttons: **Compare groups**, **Change over time**, **Distribution**, **Relationship**, **Survey items** and **Reliability / validity**. A goal does not lock you in; it suggests suitable chart types.
- **Chart type** lists the choices: **Bar chart with error bars**, **Grouped bar chart**, **Line chart over time**, **Interaction plot**, **Box plot**, **Violin plot**, **Histogram**, **Density plot**, **Q-Q plot**, **Scatter plot with fit line**, **Correlation heatmap**, **Likert diverging bars**, **Stacked bar chart**, **Percent (100%) bar chart**, **Scree plot**, **CFA path diagram**, **Interaction plot (moderation)**, **Mediation path diagram** and **Comparison with significance bars**. The last five draw from a test you have already run and saved, which you pick under **Analysis to draw from**. The comparison chart shows means (or medians for rank tests) with brackets over the pairs that differ; chapter 8 explains it.
- The **Variables** list has a **Search variables** box. Drag a variable onto a shelf, or use the + button and choose **Add to {shelf}**. Once placed, a variable shows as a pill. Its **Summary of {name}** menu chooses **Mean**, **Median**, **Sum** or **Count**, and **Remove {name} from {shelf}** takes it off.
- The **Variables** panel has a chevron button, **Hide variables**, at its top. Collapsed, it becomes a slim strip with **Show variables**. The **Customize** panel works the same way with **Hide customize** and **Show customize**. Collapsing either gives the preview the freed width; collapsing both gives the largest preview. Both choices are remembered.
- **Preview** shows the chart. To the left of the background buttons is a two-option control, **Fit** | **Actual size**. **Fit** is the default: the chart is drawn at the size set under Customize (**Width** and **Height**, default 480 by 300) and then enlarged or reduced to fill the preview column, so the preview is an exact enlargement of the figure you will save. **Actual size** shows it at its real pixel size, with scroll bars if it does not fit. The choice is remembered. The **Match app**, **Light** and **Dark** buttons change only the preview background, which is useful to see how a chart will look on white paper.

- **Save figure…** opens a menu of **PNG ({n} DPI)** in three sizes, **SVG (vector)** and **PDF**. A save dialog follows, then the message **Saved {path}.**
- **Show the numbers** expands a table of the plotted values.

Customize

The **Customize** panel changes how the chart looks. Not every option applies to every chart type; the panel shows only those that do.

Option	What it does
Style	**APA figure** (black, grey and simple, for papers) or **Statversity style**.
Figure no., **Title** / **Figure title (italic)**, **Subtitle**, **Note below the chart** / **Note (below the figure)**	The text around the figure, in APA layout for the APA style.
Error bars	**95% confidence interval**, **Standard error (SE)**, **Standard deviation (SD)** or **None**.
Significance marks, **Show non-significant pairs**	Comparison figures only. Brackets labelled with **Asterisks**, **Exact p-values** or **Off**; the checkbox also draws pairs that do not differ, labelled **ns**.
Brackets	Two-factor comparison figures only (factorial and mixed ANOVA). Which comparisons get brackets: **Within each {series}: compare {x}** (each group's change across the x axis), **At each {x}: compare {series}** (the groups against each other at each point) or **Both**. The names in braces are your own variables, for example "Within each Group: compare Time point".
Number of bins	How many bars in a histogram.
Fit line, **Correlation**, **Smoothing**	For scatter plots: **Straight line (least squares)** or **Smooth curve (LOWESS)**; **Pearson r** or **Spearman rho (ranks)**.
X axis label, **X from**, **X to**, and the same for Y	Axis titles and ranges.
Palette	**Colorblind-safe (Okabe-Ito)**, **Colorblind-safe (Tol bright)**, **Black and greys** or **Custom colors**, with **Color {n}** pickers.
Font, **Size**	**Default for the style**, **Arial**, **Helvetica**, **Calibri**, **Georgia (serif)** or **Times New Roman (serif)**; text size.
Legend	**Right**, **Bottom**, **Top**, **Left** or **Hidden**.
Show data labels, **Gridlines**	Printed values and background lines.
Width (px), **Height (px)**	Output size. Under them: "Size of the saved figure. The preview is scaled to fit." Saving a figure always uses the Width and Height you set, never the preview's on-screen scale.

IN PLAIN LANGUAGE

Error bars show how uncertain each plotted average is. A 95% confidence interval gives a range that would contain the true average in about 95 of 100 repeated studies. Standard error and standard deviation are other ways to show spread. State which kind you used in the figure note, because readers cannot tell by looking.

TIP

APA publications usually want black-and-grey figures. Choose the **APA figure** style or the **Black and greys** palette for anything going into a thesis or journal.

Edit chart on a result

Every chart on a Results page has an **Edit chart** button. It opens a dialog called **Edit chart** with a live preview and the same Customize options as above, minus the data options (you cannot change which variables are plotted, because the result fixes them). The buttons are **Reset to default**, **Cancel** and **Apply**. **Apply** keeps your look with that result; the result then shows the tag **Edited**, and **Save figure…** and exported reports use your version.

Responses

The **Responses** tab is for open-ended questions: the free-text answers people typed. It is not for numbers. You read the answers, give each a short label called a **tag** (also called a code), and then count the tags. This is the basic method of qualitative coding.

If no data are loaded the page says **Import data first.** If the dataset has no text questions it says **This dataset has no written-answer (text) questions to read.** The screenshot uses the messy Qualtrics practice dataset, which has written answers.

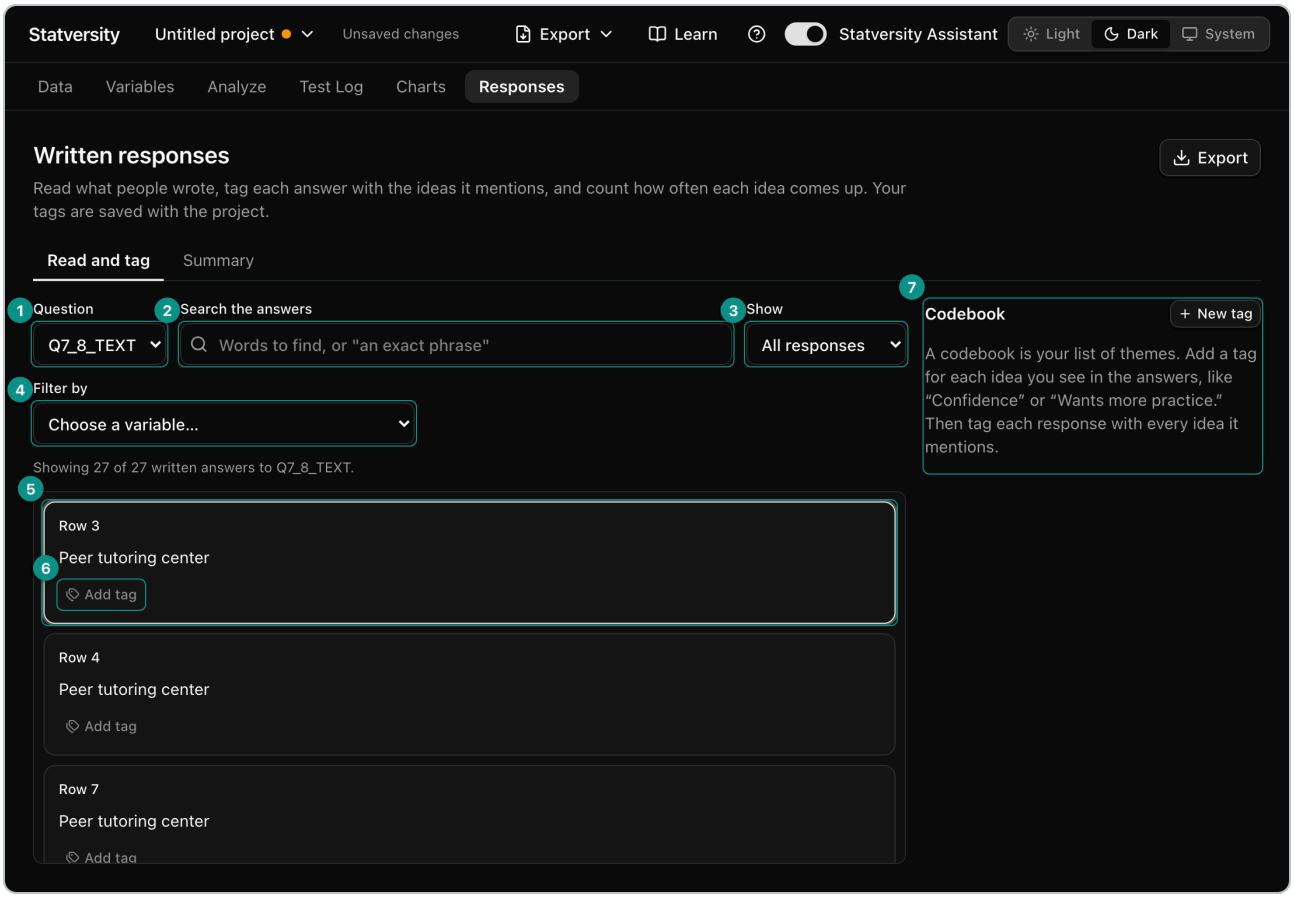


Figure 48. The Responses screen.

Number	Control	What it does
1	**Question**	Picks which written question to read.
2	**Search the answers**	Finds words or an exact phrase (put it in quotes).
3	**Show**	Limits the list to **All responses**, **Not tagged yet**, or one tag.
4	**Filter by**	Limits the list to people who gave certain answers to another question.
5	A response card	One person's answer, with its tags.
6	**Add tag** / **Edit tags**	Opens a menu of tags to apply to that answer.
7	**Codebook**	The panel where you create and edit tags.

The screenshot shows the written answers to Q7_8_TEXT in the messy Qualtrics export (messy_3header.csv), with no tags applied yet.

Reading and tagging

Under the tab **Read and tag**, the filter bar narrows the list and a line such as **Showing {n} of {total} written answers to {variable}.** tells you how many are visible. **Clear filters** resets it. Each card carries tag chips; click the small x (**Remove tag {name}**) to take one off.

You can work without the mouse:

Key	What it does
Down arrow or **j**	Next response.
Up arrow or **k**	Previous response.
Home / **End**	First / last response.
1 to **9**	Toggle the tag in that position on the current response.

The codebook

The **Codebook** panel is your list of tags. Click **New tag**, give a **Tag name**, write a **Definition (when should someone use this tag?)**, pick a **Color**, and click **Add tag** (or **Save tag** when editing). Writing a definition is not optional in spirit: it is what keeps your coding consistent over hundreds of answers, and it is what a second coder needs to agree with you. Use **Edit tag {name}** and **Delete tag {name}** on a row; deleting asks you to confirm with **Delete** or **Keep**.

Summary and turning tags into variables

The **Summary** tab shows counts and percentages of each tag, with a **Compare by** menu (**Everyone together**, or by a grouping question) and a chart with **Show the numbers**. **Make yes/no variables** converts each tag into a 1/0 column (1 if the answer has that tag), and **Go to Analyze** takes you to the Advisor or All tests so you can, for example, compare how often a theme appears in two groups with a chi-square test.

Exporting coded answers

The **Export** menu on this screen saves your work: **Coded responses (Excel, one column per tag)**, **Coded responses (Word, grouped by tag)**, **Codebook (Excel)** and **Codebook (Word)**. A message such as **Saved {file} ({n} responses).** confirms it.

> **CAUTION**
> Statversity helps you organize and count tagged answers. It does not read the answers for you or decide on the tags. The coding is your own analysis, and a thesis committee will expect you to describe how you derived the tags.

Study Planner

The **Study Planner** is for the time before you collect data. It asks the same plain-language questions as the Test Advisor, picks a suitable analysis, works out how many participants you need, and produces a plan you can attach to an ethics application or a proposal. It works without any data loaded. Open it from **Plan a study** on Home, or from a **Sample size for …** test on the All tests menu.

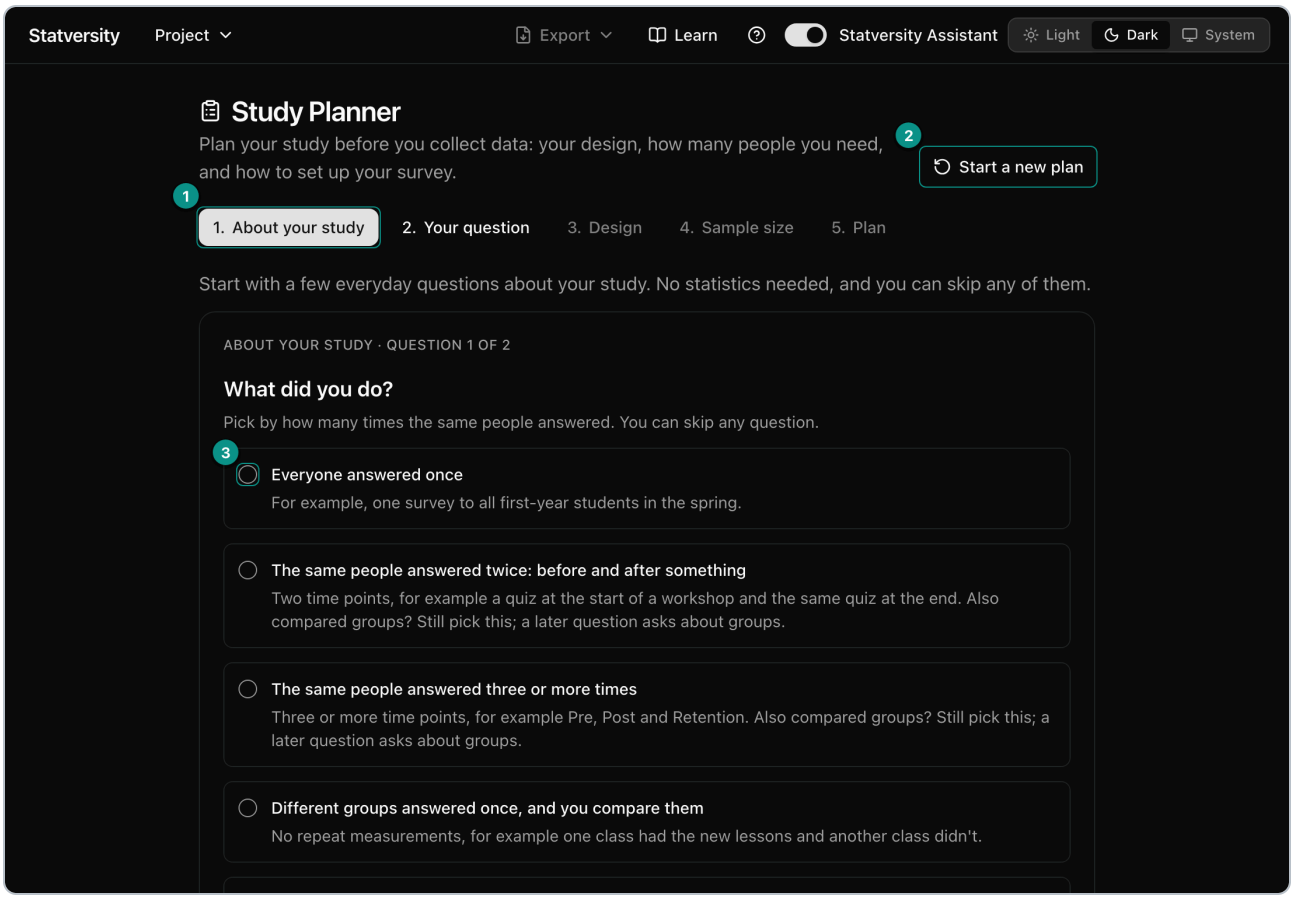


Figure 49. The Study Planner.

Number	Control	What it does
1	Steps (**About your study**, **Your question**, **Design**, **Sample size**, **Plan**)	Show where you are; click to move between steps you have already reached.
2	**Start a new plan**	Clears the planner to begin again.
3	Step content	The questions or results for the current step.

Step by step:

1. **About your study.** The same study questions described in chapter 5 (what you did, whether you compare groups, whether people can be matched across times, what kinds of questions you ask).

Continue with **Next: name your study**.

2. **Your question.** Type a name under **What is your study called?** and write your research question in the box. Continue with **Next: describe your design**.
3. **Design.** A few Advisor-style questions with **Continue** (or **Use this answer**), **Change** and **Back**. At the end Statversity shows **Your planned analysis**. Continue with **Next: how many people?**
4. **Sample size.** Choose how large an effect you want to be able to detect: **Small**, **Medium** or **Large**, or type your own value in **Or type your own {symbol}**. Set **Significance level (alpha)** (**.05 (usual)**, **.01** or **.10**), **Power you want** (**80% (usual)**, **90%** or **95%**) and, depending on the design, **Study design**, **Direction of the test**, **Group size ratio (group 2 ÷ group 1)**, **Number of groups**, **Times each person is measured**, **Expected correlation between time points**, **Number of predictors** and **Expected drop-out (%)**. Click **Calculate sample size**. Statversity reports **You need about {n} people in total**. The panel **What if you can't get that many?** lets you type **People you can realistically get ({unit})** and click **Check** to see what power that gives you. Continue with **Next: see my plan**.
5. **Plan.** A summary document with the sections **Design summary**, **Planned analyses**, **Sample size**, **Data collection recommendations** and **Assumptions to check after you collect data**. **Save plan in project** stores it, **Export as Word (.docx)** writes a file, and **Start analysis project from this plan** seeds a new analysis project.

> **IN PLAIN LANGUAGE**
>
> **Power** is the chance that your study will detect an effect if the effect is really there. 80% is the usual minimum: a study with 50% power is a coin flip. **Alpha** is the false-alarm rate you accept, normally 5%. A **sample size** calculation finds how many people give you the power you want for the effect you expect. Choosing the effect size is the hard part; the honest answer is usually "the smallest difference that would matter in practice".

If a plan is already in progress and you start another from a Learn page or the All tests menu, a dialog **Replace your in-progress plan?** offers **Keep my plan** or **Show the saved plan**.

Export

The **Export** menu in the top bar (or **More actions** in a narrow window, where the items read **Export report…**, **Print report…**, **Export data…**, **Export codebook…** and **Export test log…**) saves your work as files. It is disabled until a project exists.

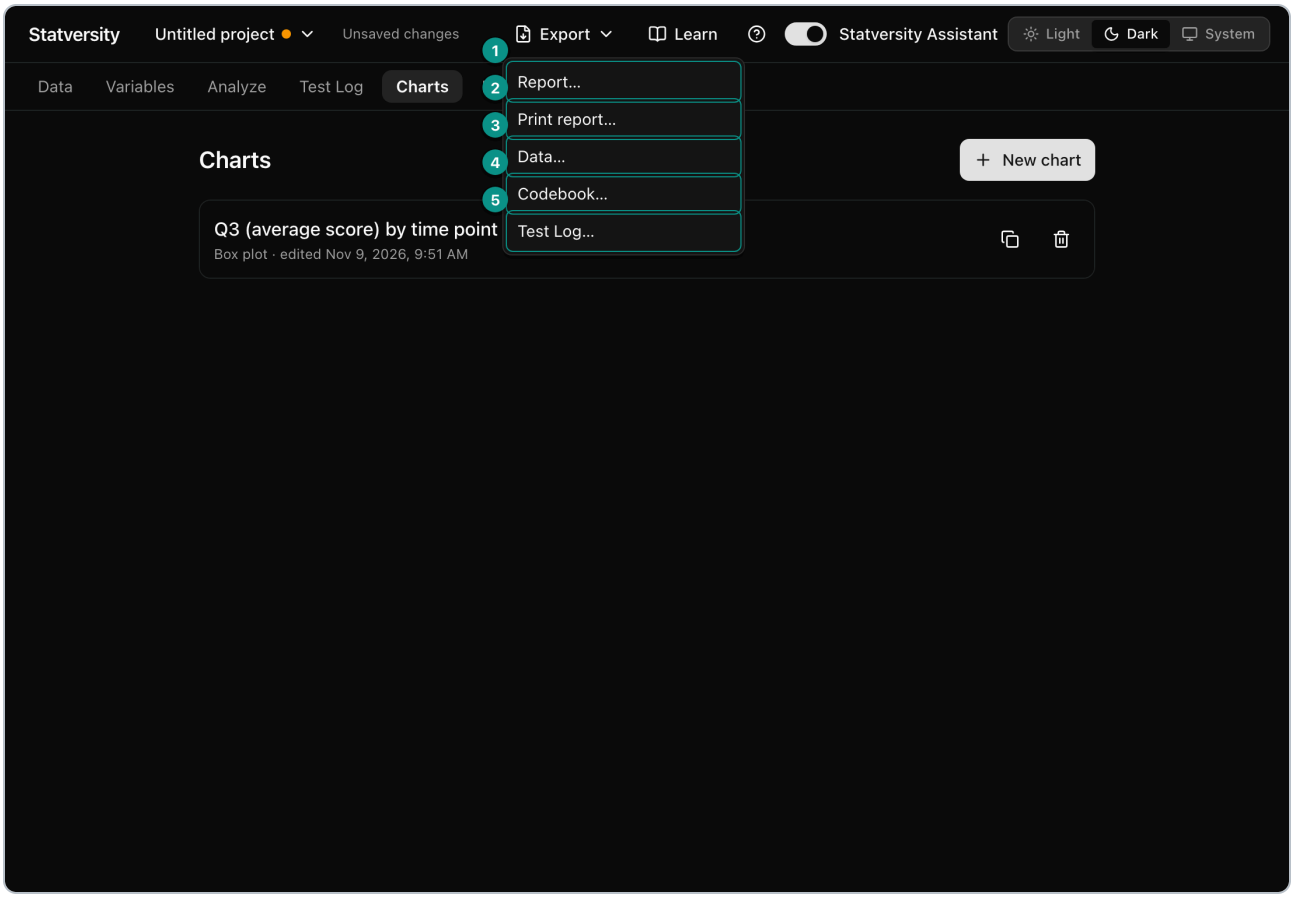


Figure 50. The Export menu.

Number	Control	What it does
1	**Report…**	Builds a Word or PDF report of selected tests.
2	**Print report…**	Opens the report options, then the system print dialog.
3	**Data…**	Saves your data as Excel, CSV, SPSS or Stata.
4	**Codebook…**	Saves a description of each variable as Excel or Word.
5	**Test Log…**	Saves the list of tests as Excel, CSV or Word.

Each item opens a small dialog with a **Format** choice, **Cancel**, and **Export**; after saving you see the saved location and a **Done** button.

Menu item	Formats offered	Notes
Report…	**Word (.docx)**, **PDF**	Contents chosen in the dialog below.
Data…	**Excel (.xlsx)**, **CSV (.csv)**, **SPSS (.sav)**, **Stata (.dta)**	**Rows**: **Included by the filter ({n})** or **All rows ({n})**. Checkboxes: **Include metadata columns**, **Add a variable-label row**, **Leave missing-value codes blank**, **Exclude columns flagged as PII**.
Codebook…	**Excel (.xlsx)**, **Word (.docx)**	Variable names, labels, types, answer choices.
Test Log…	**Excel (.xlsx)**, **CSV**, **Word (.docx)**	One row per logged test.
Save figure… (charts and results)	**PNG ({n} DPI)** in three sizes, **SVG (vector)**, **PDF**	Saves one picture.

> **CAUTION**
>
> **Exclude columns flagged as PII** keeps personal information (names, emails, IP addresses) out of the file. Tick it before you share data with a collaborator or a repository. A file that you export without ticking it will contain whatever columns were in the project.

The Export report dialog

Report… opens the report dialog. Statversity builds the file on your computer; nothing is uploaded.

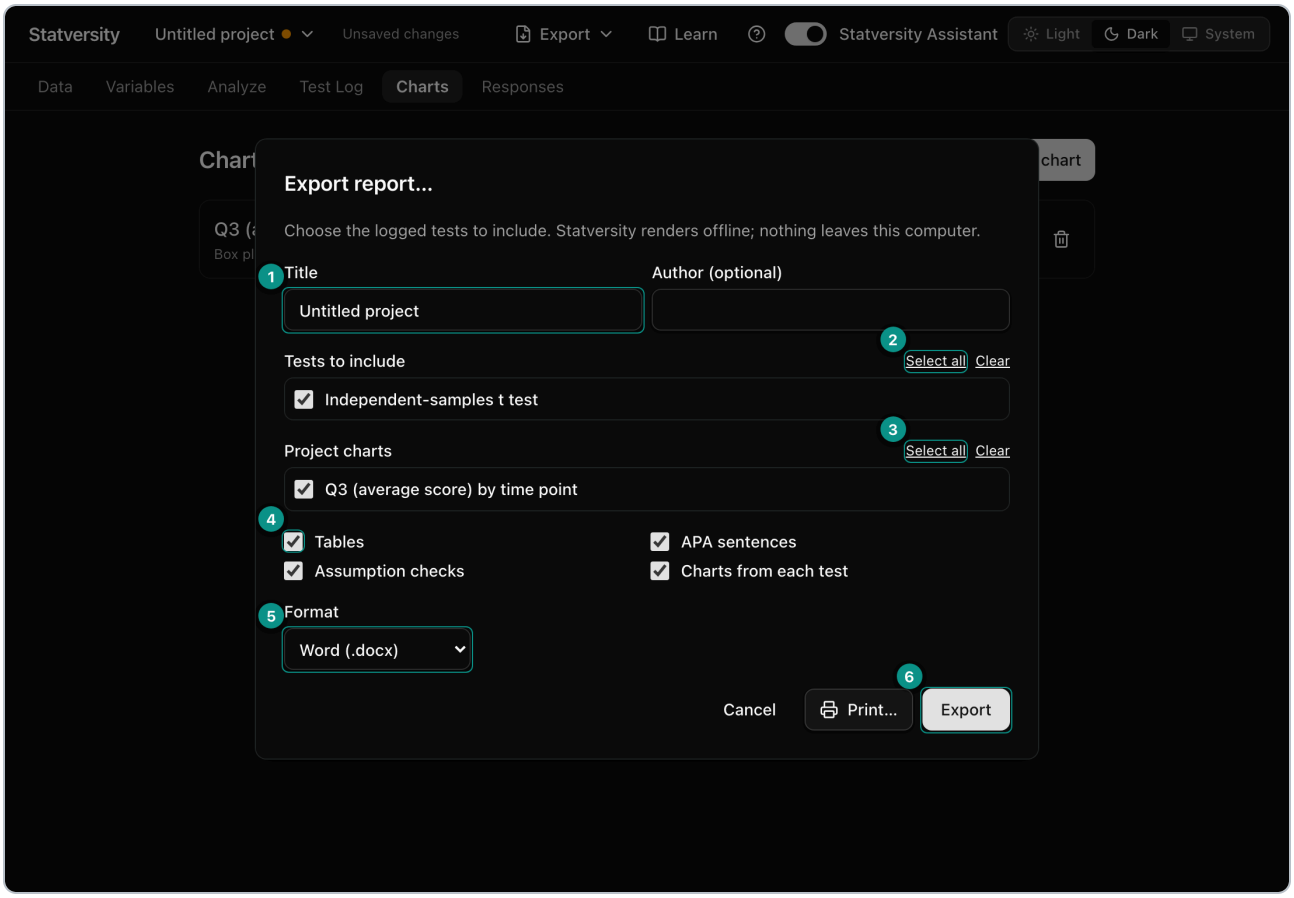


Figure 51. The Export report dialog.

Number	Control	What it does
1	**Title** and **Author (optional)**	Printed at the top of the report.
2	**Tests to include** (with **Select all** and **Clear**)	One checkbox per logged test.
3	**Project charts** (with **Select all** and **Clear**)	Charts you saved in the Chart builder.
4	**Tables**, **APA sentences**, **Assumption checks**, **Charts from each test**	Which parts of each test go in.
5	**Format**	**Word (.docx)** or **PDF**.
6	**Export**	Builds the file. It shows **Rendering charts…** then **Saving…**.

The screenshot shows the report dialog with **Select all** clicked for tests.

When it finishes, the dialog shows **Saved to {path}** with **Copy path** and **Done**. If you have no logged tests the dialog says **Log a test first (Advisor or Results screen) to include it in a report.**

Print

Printing is available in two ways. **Print…** on a Results page prints that one result. **Print report…** (from the **Export** menu or **More actions**) opens the same dialog as **Report…** with the same options, but its button reads **Print…** and opens your computer's print dialog instead of saving a file. From there you can print on paper or choose **Save as PDF**.

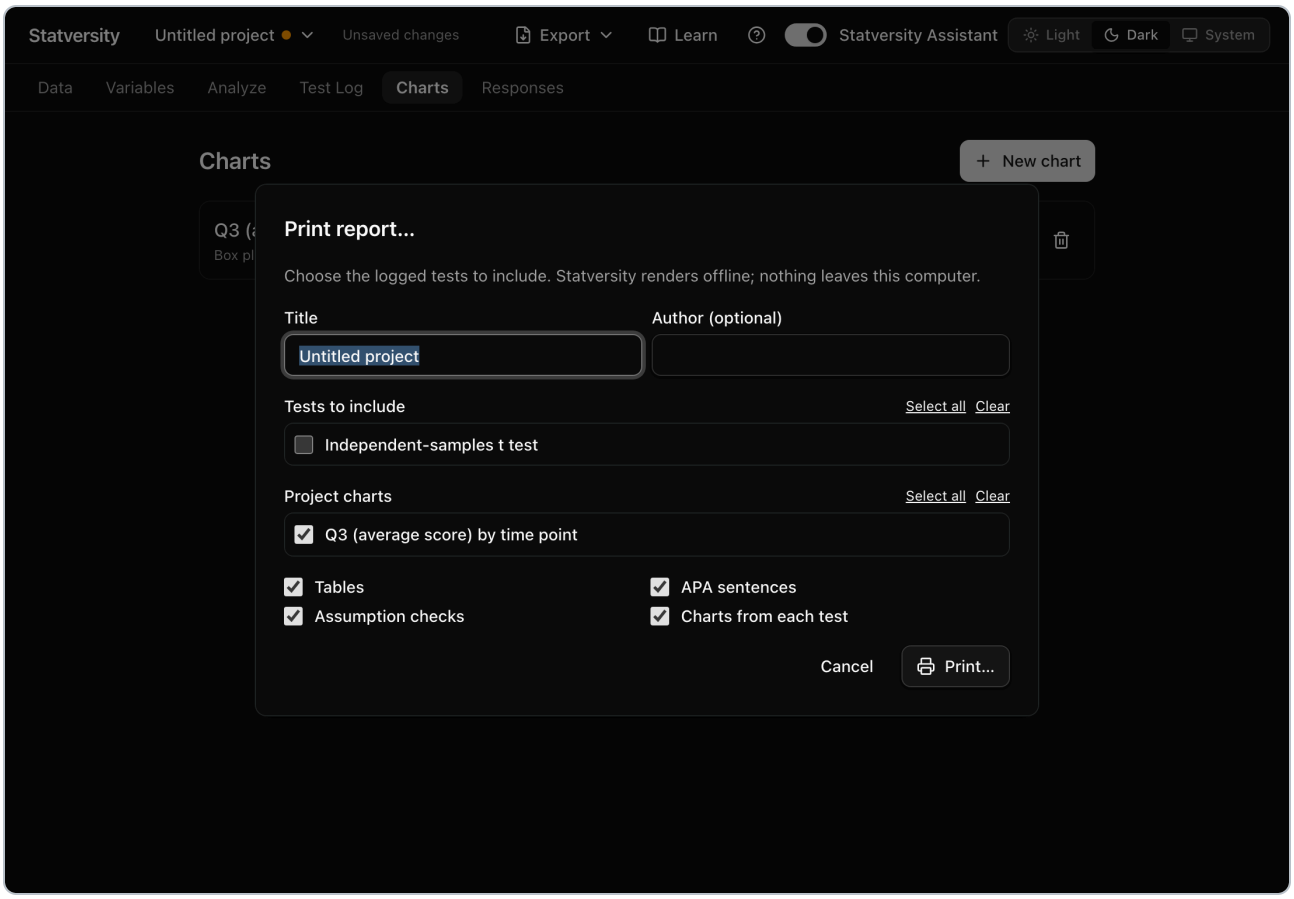


Figure 52. The Print report options dialog.

The screenshot shows the Print report options dialog. The operating system's own print dialog cannot be captured, so it is not shown. No tests are ticked yet, and the dialog has **Print…** where the Export dialog has **Format** and **Export**.

CHAPTER 10

Working without the Assistant

Some people already know which test they want. They have run the same analysis in SPSS or R for years, or their supervisor told them exactly what to run. For them, answering Advisor questions is a detour. Statversity has a mode for this: the **Statversity Assistant** switch, set to off.

> **NOTE**
> You may see this called “expert mode” in older notes or conversations. **There is no control called Expert mode.** Turning the **Statversity Assistant** switch off is what puts Statversity in that state, and this chapter uses “Assistant off” throughout.

The switch

The **Statversity Assistant** switch sits in the top bar, to the right of the **Learn** and tour buttons. It is on by default. Hover over it and the tooltip explains: on, Statversity asks about your study and suggests tests; off, you go straight to your variables and pick tests yourself. In a narrow window it moves into the **More actions** menu as a **Statversity Assistant** item with a tick when on. The same switch also appears on the **Choose files** step of the import wizard with a line of explanation.

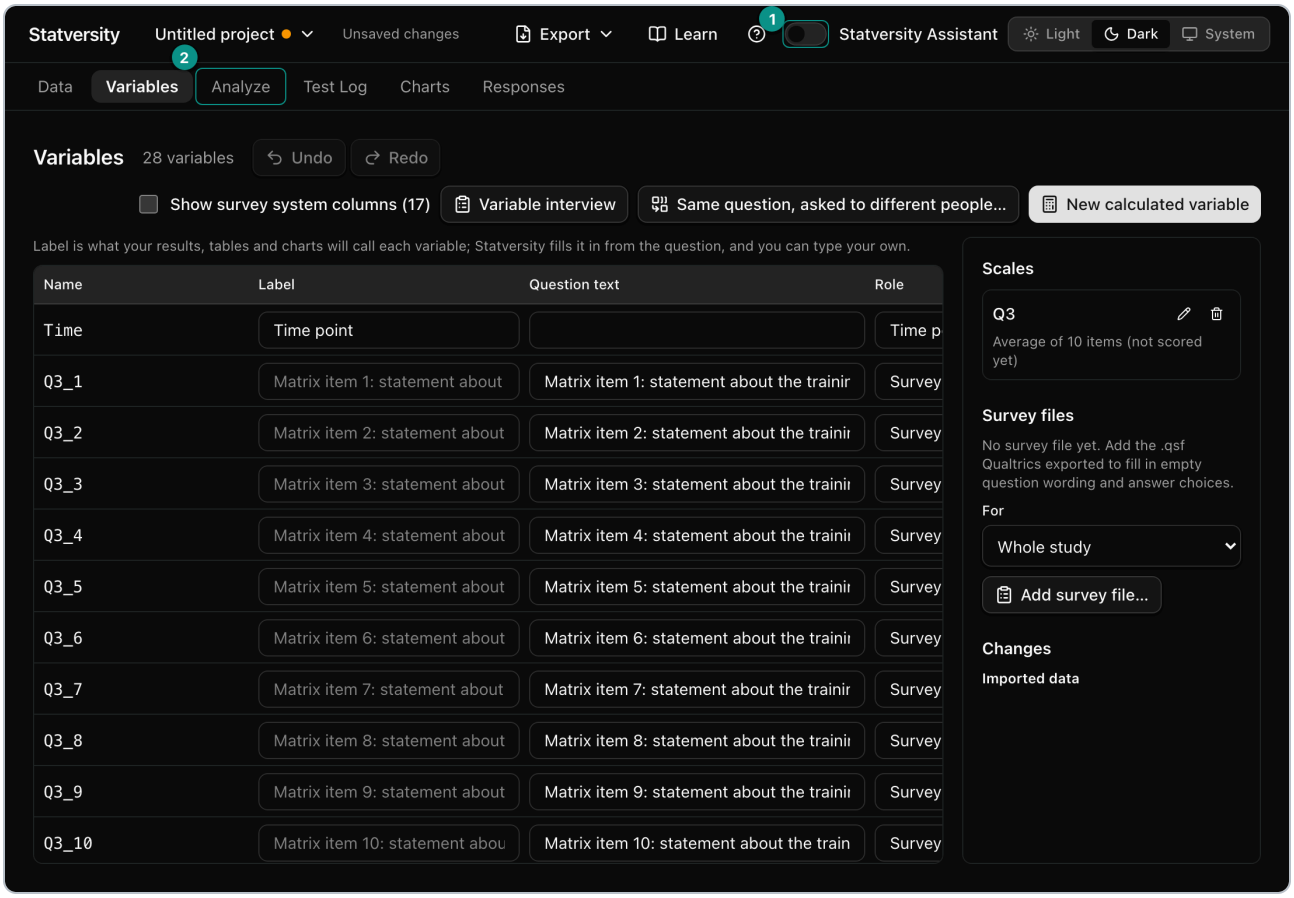


Figure 53. The top bar with the Statversity Assistant switch off.

Number	Control	What it does
1	**Statversity Assistant** (switch)	Turns the Assistant on or off. Here it is off.
2	**Analyze** tab	With the Assistant off, opens the All tests menu instead of the Test Advisor.

The screenshot shows the Variables screen of the one-group pretest and posttest practice data, imported with the Assistant off, so the Variable Interview did not open by itself.

The setting is remembered by the app, so it is still off the next time you open Statversity on the same computer. It applies to the whole app, not to one project. You can flip it at any moment. If you are on the Advisor when you switch it off, the screen swaps to the All tests menu, and the other way round.

What changes

Switching the Assistant off changes four things. Everything else stays the same.

Area	Assistant on (default)	Assistant off
Import wizard	Includes the **About your study** step. At the end the button reads **Set up variables** and starts the Variable Interview.	No **About your study** step. At the end the button reads **Go to variables** and you land on the Variables screen.
Analyze tab	The **Test Advisor** asks questions and recommends a test.	The **All tests** menu, a searchable list of every test. Choosing a test goes straight to **Choose your variables**, and the main button reads **Run**. The assumption walk is skipped.
Results	No backup-test suggestion.	A **Run {backup} instead** suggestion appears when assumptions look doubtful.
Variables screen	The **Variable interview** button is available.	The same.

Two points need emphasis:

- **The Variable Interview is not removed.** It no longer opens by itself after import, but the **Variable interview** button on the Variables screen still works. Types, roles, scales and answer keys still matter, because tests need to know what each column is.
- **The checks are not gone.** Without the assumption walk you do not step through them before running, but the result has an **Assumption checks** section, and if the assumptions look doubtful Statversity shows the backup suggestion.

The All tests menu

All tests is a single page listing every test Statversity can run. You reach it from the **Analyze** tab (Assistant off), from **Choose a test myself** on the Advisor, from **All tests** on the Analyze landing, or from **Run this test** on a Learn page.



Figure 54. The All tests menu.

Number	Control	What it does
1	**Search tests**	Filters the list as you type. It also knows common names, such as “chi square”, “Cronbach” and “Mann-Whitney U”. With no match you see **No matching tests**.
2	**Descriptives**	The family of summary tests.
3	**Compare Means**	Tests of averages (t tests and their rank-based forms).
4	**ANOVA**	Comparisons of three or more means, repeated measures and related methods.
5	**Categorical Data**	Tests on counts and proportions.
6	**Correlation**	Tests of how strongly variables go together.
7	**Regression**	Models that estimate one variable from others.
8	**Reliability & Agreement**	How consistent a scale or a set of raters is.
9	**Factor Analysis**	Finding the structure behind a set of questions.
10	**Power Analysis**	Sample size calculations.

The page also has a **Plan my analyses** card at the top, above the search box (the screenshot is a tall capture, so the last family is cut off at the bottom). Badge 1 is the search box and badges 2 to 10 are the family headings, top to bottom.

Tests that use ranks rather than raw scores carry the label **(nonparametric)**. A test that is not built into this version of Statversity is shown as **Not available in this version**. The names below are Statversity’s default labels; the engine’s own label wins if it differs, so a button may read slightly differently.

Descriptives

- **Descriptive statistics**
- **Gain score (pre/post)**
- **Normalized gain (Hake’s g)**

Compare Means

- **One-sample t test**
- **Independent-samples t test**
- **Paired-samples t test**
- **Mann-Whitney U test**
- **Wilcoxon signed-rank test**
- **Wilcoxon signed-rank test (one sample)**
- **Sign test**

ANOVA

- **One-way ANOVA**
- **Welch's ANOVA**
- **Repeated-measures ANOVA**
- **Factorial ANOVA**
- **Mixed ANOVA**
- **Aligned rank transform ANOVA**
- **ANCOVA**
- **Quade's test**
- **MANOVA**
- **MANCOVA**
- **Kruskal-Wallis test**
- **Friedman test**

Categorical Data

- **Chi-square test of independence**
- **Chi-square goodness-of-fit test**
- **Fisher's exact test**
- **McNemar's test**
- **Cochran's Q test**

Correlation

- **Pearson correlation**
- **Spearman correlation**
- **Kendall's tau-b**
- **Point-biserial correlation**
- **Partial correlation**
- **Correlation matrix**

Regression

- **Linear regression**
- **Hierarchical regression**
- **Logistic regression**
- **Ordinal regression**
- **Moderation analysis**
- **Mediation analysis**

Reliability & Agreement

- **Cronbach's alpha**
- **McDonald's omega**
- **Kuder-Richardson 20 (KR-20)**
- **Split-half reliability**
- **Item analysis**
- **Intraclass correlation (ICC)**
- **Cohen's kappa**
- **Fleiss' kappa**
- **Kendall's W**

Factor Analysis

- **Exploratory factor analysis (EFA)**
- **Confirmatory factor analysis (CFA)**

Power Analysis

- **Sample size for a t test**
- **Sample size for ANOVA**
- **Sample size for a correlation**
- **Sample size for a chi-square test**
- **Sample size for a regression**

The **Power Analysis** tests are different: they do not use your data, so choosing one opens the Study Planner at the sample-size step (chapter 9). The **Plan my analyses** card at the bottom of the page offers the plan builder from chapter 7.

If you cannot find a test, try the **Search tests** box with the name you know it by. The empty result suggests alternatives such as t-test, chi square or Cronbach.

Going straight to Choose your variables

Click any statistical test and you land on **Choose your variables**, the same screen described in chapter 7. A back button at the top reads **All tests**.

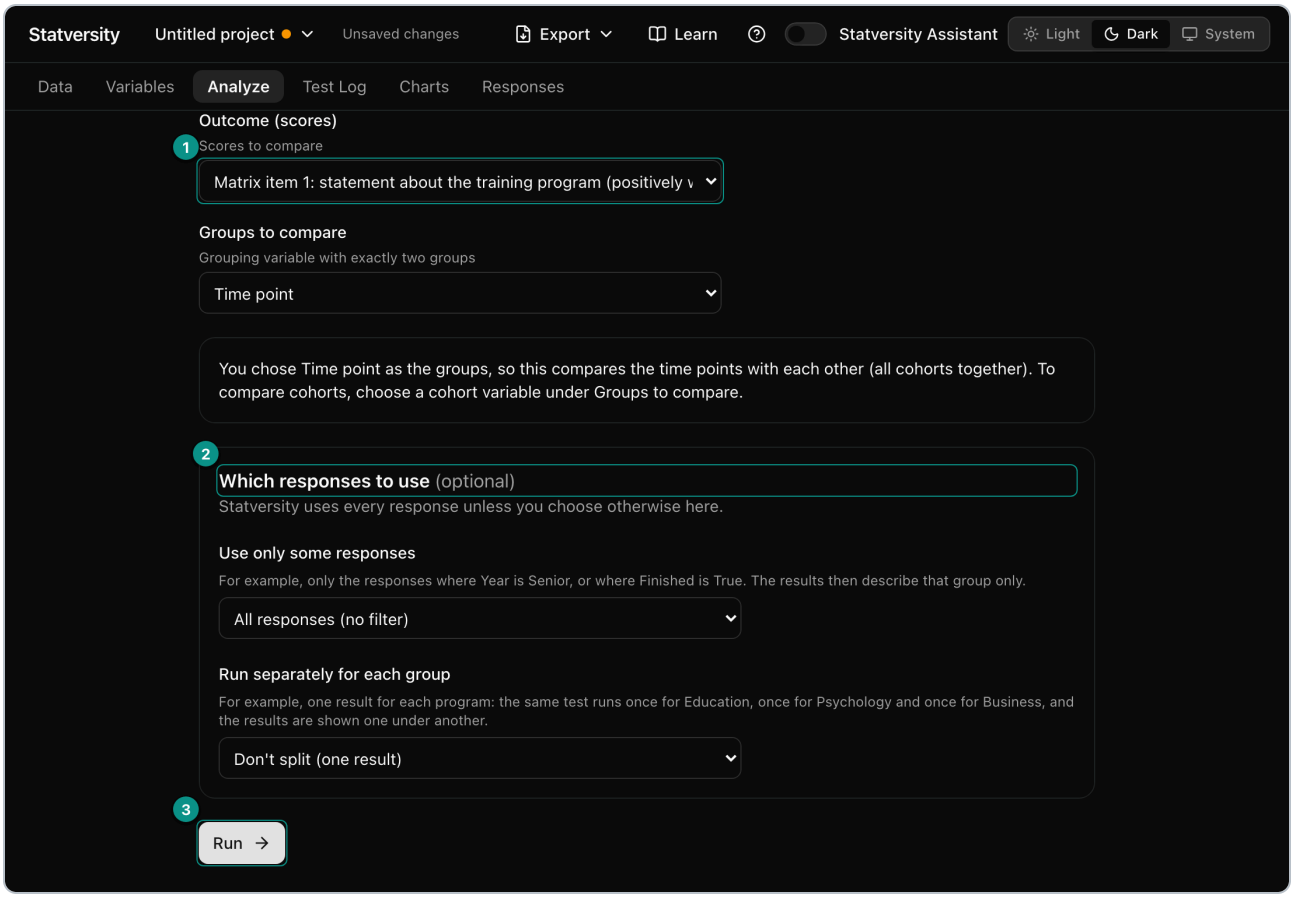


Figure 55. Choose your variables, reached straight from All tests.

Number	Control	What it does
1	Role slots (here **Outcome (scores)**)	Assign your columns to the parts of the test.
2	**Which responses to use (optional)**	Limits rows, or runs the test once per group.
3	**Run**	Runs the test immediately.

The screenshot is the independent-samples t test picked from All tests, with Q3_1 as the outcome and **Time point** as the groups. Unlike the guided route in chapter 7, there is no **Analysis** drop-down at the top; to change the test, go back with the **All tests** button.

The difference from the guided route is the last button. In the guided route it reads **Check the assumptions** and leads to the checks and a decision. Here it reads **Run** and the result opens at once.

> **CAUTION**
>
> Running directly means you skip the walk through the assumptions. If you pick a test whose assumptions your data do not meet, Statversity will still run it. Read the **Assumption checks** section of the result, and watch for the backup suggestion described next.

Run {backup} instead

On a result, when assumptions look doubtful, Statversity shows a notice with a button **Run {backup} instead**, where {backup} is the name of the better-suited test. This notice is a feature of the Assistant-off route: with the Assistant on you decide between the recommended test and its alternative on the decision screen before running, so there is nothing to suggest afterward.

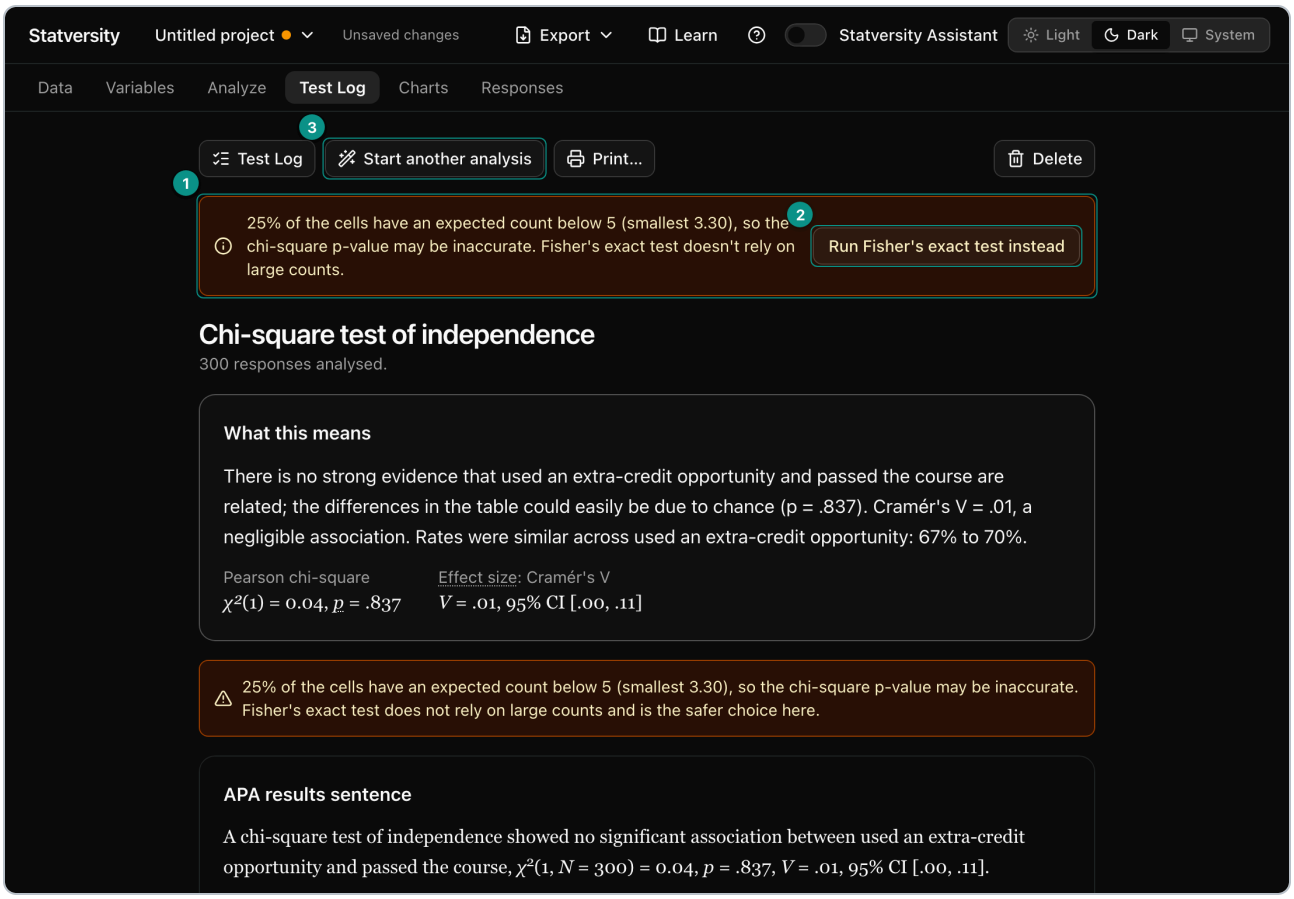


Figure 56. A result showing the Run backup instead suggestion.

Number	Control	What it does
1	The notice	Says which assumption looked doubtful.
2	**Run {backup} instead**	Runs the backup test on the same variables and opens its result.
3	**Start another analysis**	Opens All tests, since the Assistant is off.

The screenshot is a chi-square test of independence on the categorical_outcomes practice data (Q4 by Q3), run with the Assistant off. A quarter of the cells have small expected counts, so the button reads **Run Fisher's exact test instead**.

Click **Run {backup} instead** and a second result appears in the Test Log. Your first result is kept, so you can compare the two. Chapter 8 has advice on what to report when they agree and when they differ.

When to switch the Assistant off

Switch it off when:

- You already know which test you need, for example you are repeating a published analysis or following a protocol.
- You are running many similar tests and the questions have become repetition.
- You have used the Advisor enough to recognize its recommendations and want the shortcut.
- You want a test the Advisor would not have chosen, such as a particular follow-up.

Leave it on, or switch it back on, when:

- You are new to statistics. The questions and the assumption walk are the teaching part of Statversity. Skipping them is how a beginner ends up with a result for the wrong test.
- You are not sure whether your data are linked, how many groups there are, or which kind of outcome you have. The Advisor will ask in plain language.
- You are checking a test you have not used before. Leave it on for the first run, then switch it off for the repeats.

> **TIP**
> A good compromise is to leave the Assistant on for your first analysis of each kind and read every check. Once you can predict what each screen will say, switching it off saves time without losing the safety net, because the doubtful-assumption notice still appears on results.

> **TRY IT YOURSELF**
> Switch the **Statversity Assistant** off. Click **Analyze**, search for “paired”, open **Paired-samples t test**, and fill in the pre and post columns from the practice data. Click **Run**, and compare the result with the one you got from the Advisor route in chapter 3. They should be identical. Then switch the Assistant back on.

SECTION 3

Walkthroughs

CHAPTER 11

Did the training help?

This chapter takes you through one complete analysis, from opening the data to reading the finished result. Follow along on your own screen. Every number in the chapter is the number Statversity showed when this walkthrough was captured, so yours should match.

The question

Did people's scores on a ten-question survey go up after a training program?

This is a **before-and-after study** (also called a pre/post design). A group of people answered a survey before a training program, took the training, and then answered the same survey again. If the scores moved up, that is evidence the training did something. It is not proof, because other things could have changed between the two surveys, but it is the first thing you would want to know.

There is one catch, and it shapes the whole chapter. The two sets of answers were saved in two separate files, and nothing in either file says which row belongs to which person. Statversity cannot line up one person's "before" with the same person's "after." That changes which statistical test is honest to use, and you will see Statversity handle it for you.

The data

The practice dataset is called **Did the training help?** It is made up (nobody real is in it), but it behaves like real classroom data. It has two files:

- `pre.csv`, with 60 survey responses collected before the training.
- `post.csv`, with 60 survey responses collected after the training.

Each file has ten rating questions named `Q3_1` to `Q3_10`. Each is a statement about the training program, answered on a 1 to 5 scale from strongly disagree to strongly agree. Together the ten questions measure one idea, how positively people feel about the training. Two of the ten statements are worded in the opposite direction from the rest, a detail you will meet in the Scales step.

The files are Qualtrics exports (Qualtrics is a common survey tool), which is why you will see a few survey-system columns along the way. Chapter 5 explains each import screen in detail. Here we will move through them quickly and stop where a decision matters.

Open the practice data

1. Start Statversity. The Home screen shows **Welcome to Statversity**.

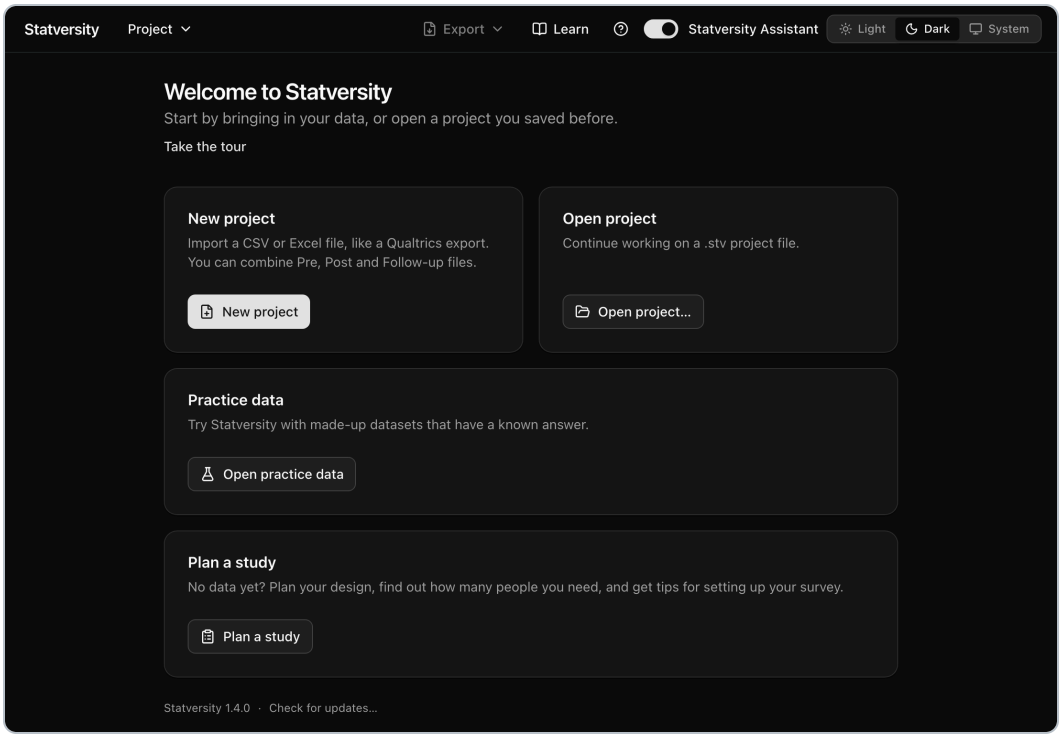


Figure 57. The Home screen, Welcome to Statversity, with the Open practice data button.

2. Click **Open practice data**. The **Practice data** screen lists the practice datasets.

3. Find the card titled **Did the training help?** It says “One group took the same 10-item survey before and after a training program” and “2 files.” Click **Import this dataset** on that card.

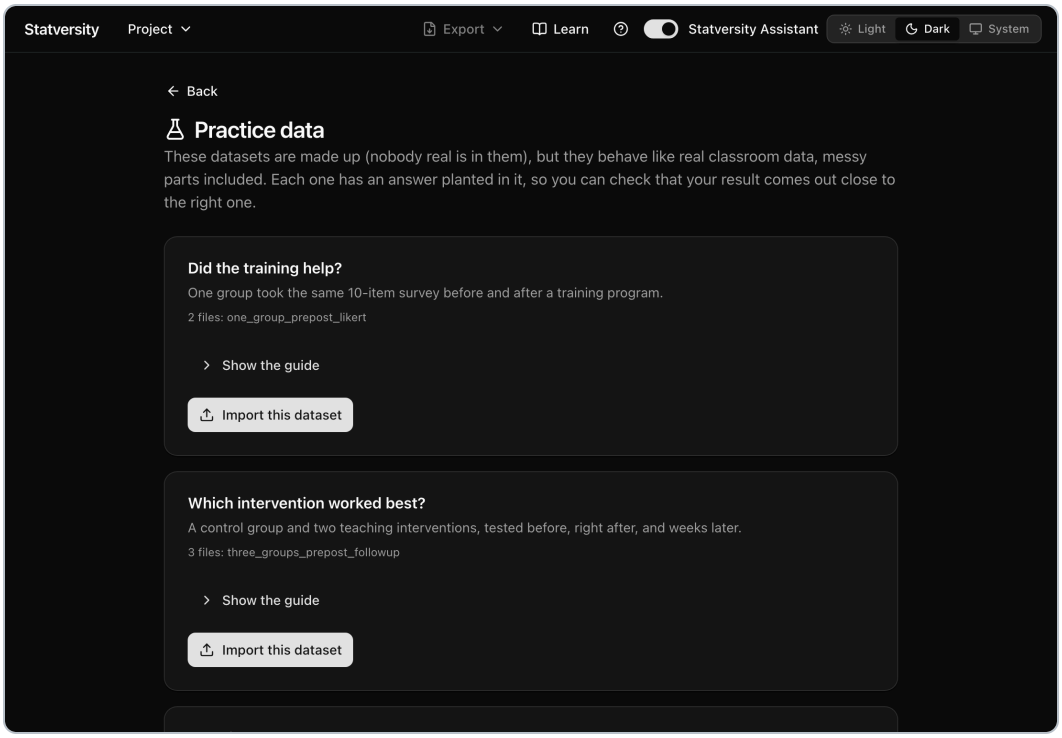


Figure 58. The Practice data screen with the Did the training help? card at the top.

Steps: bringing the data in

The import runs as a seven-step wizard. The list down the left side shows where you are.

4. The screen shows **Choose files** with `pre.csv` and `post.csv` already listed. Statversity loaded the two practice files for you. Click **Read files**.

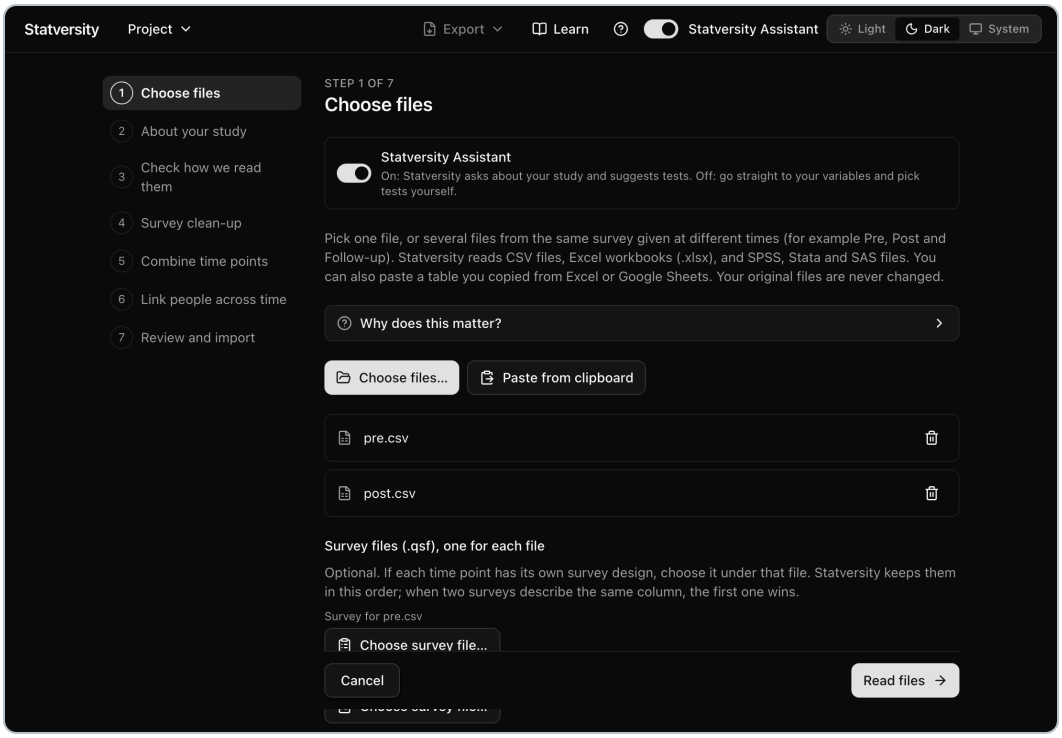


Figure 59. Step 1 of 7, Choose files, showing pre.csv and post.csv.

5. The screen shows **About your study**. Statversity asks what you did, such as “The same people answered twice: before and after something.” These answers only sharpen Statversity’s first guesses, and you can skip them. Click **Skip these questions**.

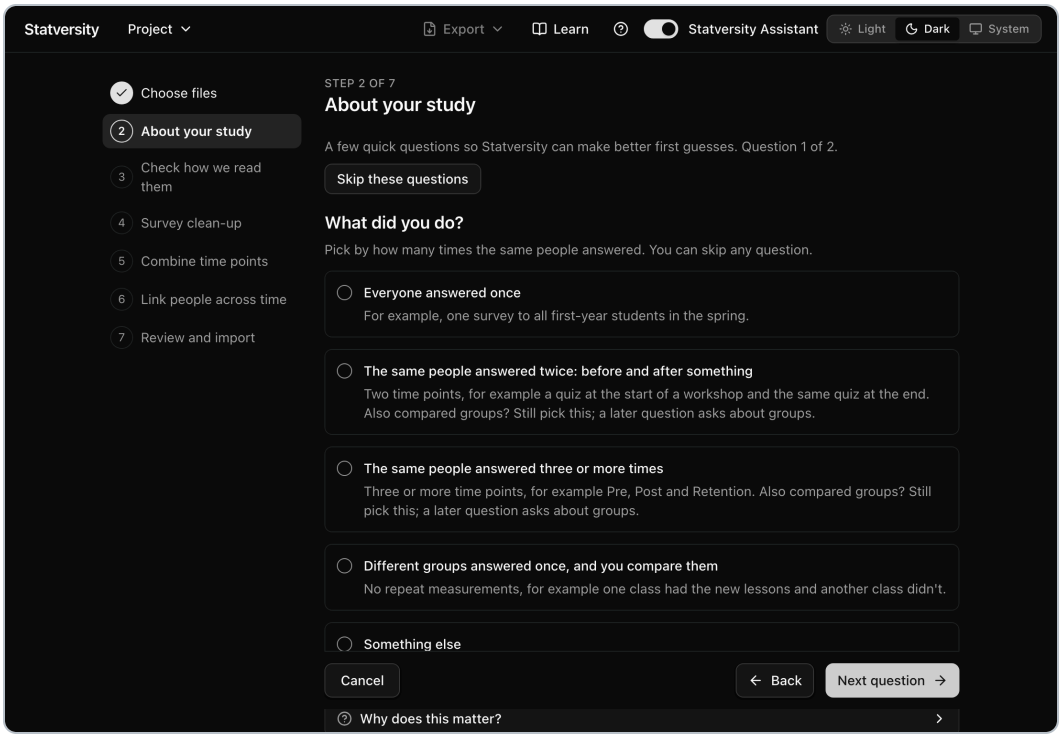


Figure 60. Step 2 of 7, About your study, with the Skip these questions button.

> **NOTE**
>
> Skipping is fine here because Statversity will ask you the real questions later, at the point where the answer changes what happens. If you do answer this one, choose “The same people answered twice,” which is how the study was run, even though Statversity cannot match the people.

6. The screen shows **Check how we read them**. Statversity reports how it read each file: 60 rows and 27 columns in `pre.csv`, recognized as a Qualtrics export. It also points out that `Q3_1` through `Q3_10` look like one matrix question and suggests treating them as a scale named Q3. Nothing needs changing. Click **Continue**.

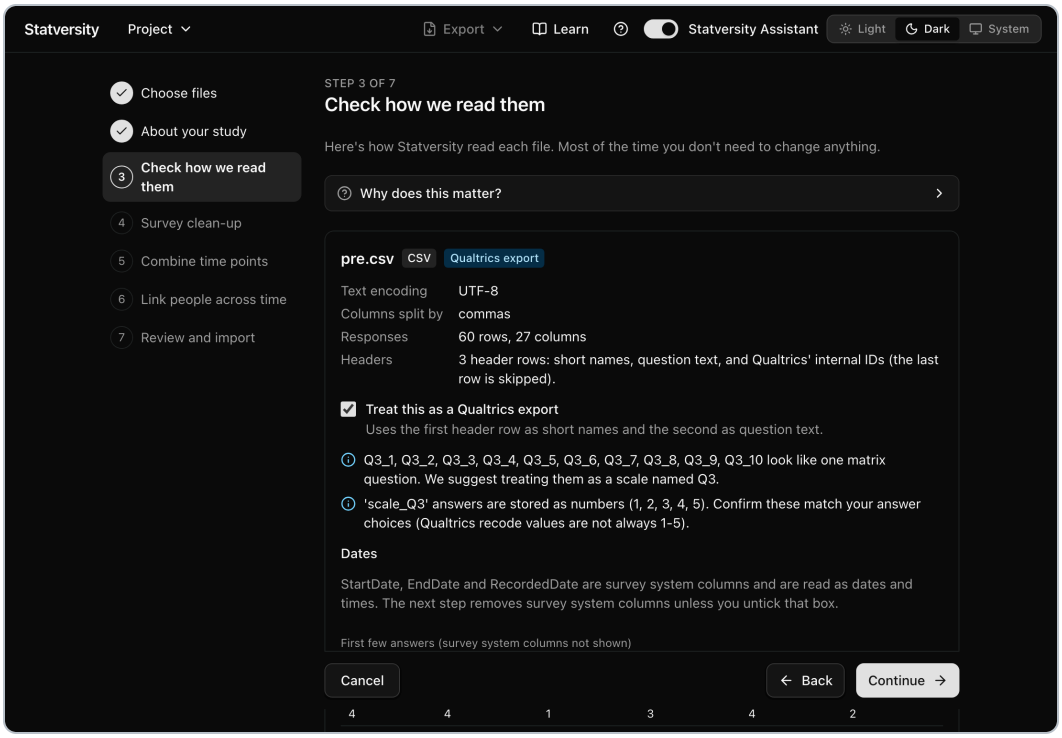


Figure 61. Step 3 of 7, Check how we read them, showing how pre.csv was read.

7. The screen shows **Survey clean-up**. Statversity picked safe defaults: hide the survey-system columns, and leave out any “Survey Preview,” “Survey Test,” or “Spam” responses. The counts beside each option show 0 rows to remove in both files, so there is nothing to clean here. Click **Continue**.

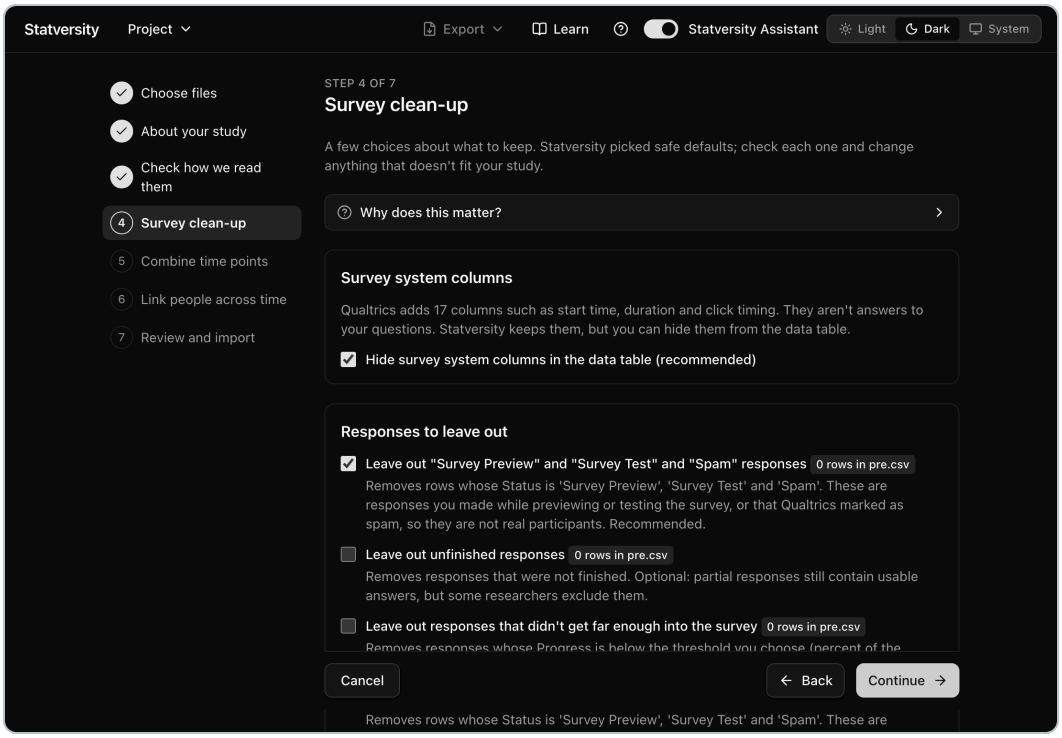


Figure 62. Step 4 of 7, Survey clean-up, with the recommended boxes ticked.

8. The screen shows **Combine time points**. Statversity will stack the two files into one table and add a new column named **Time** that says which file each row came from. Check that the labels read **Pre** for `pre.csv` and **Post** for `post.csv`. They already do, and 27 questions matched by their wording across both files, so there is no need to change anything. Click **Continue**.

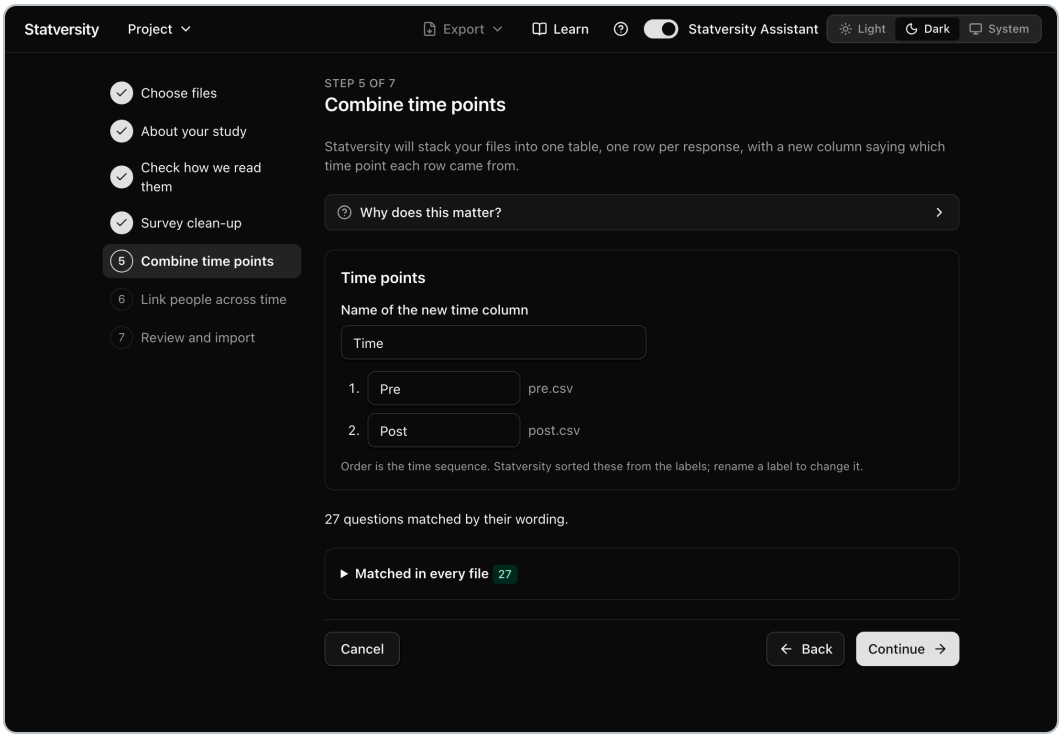


Figure 63. Step 5 of 7, Combine time points, with the labels Pre and Post.

The decision that matters: linking people across time

9. The screen shows **Link people across time**. This is the most important choice in the import, so read the options before you click.

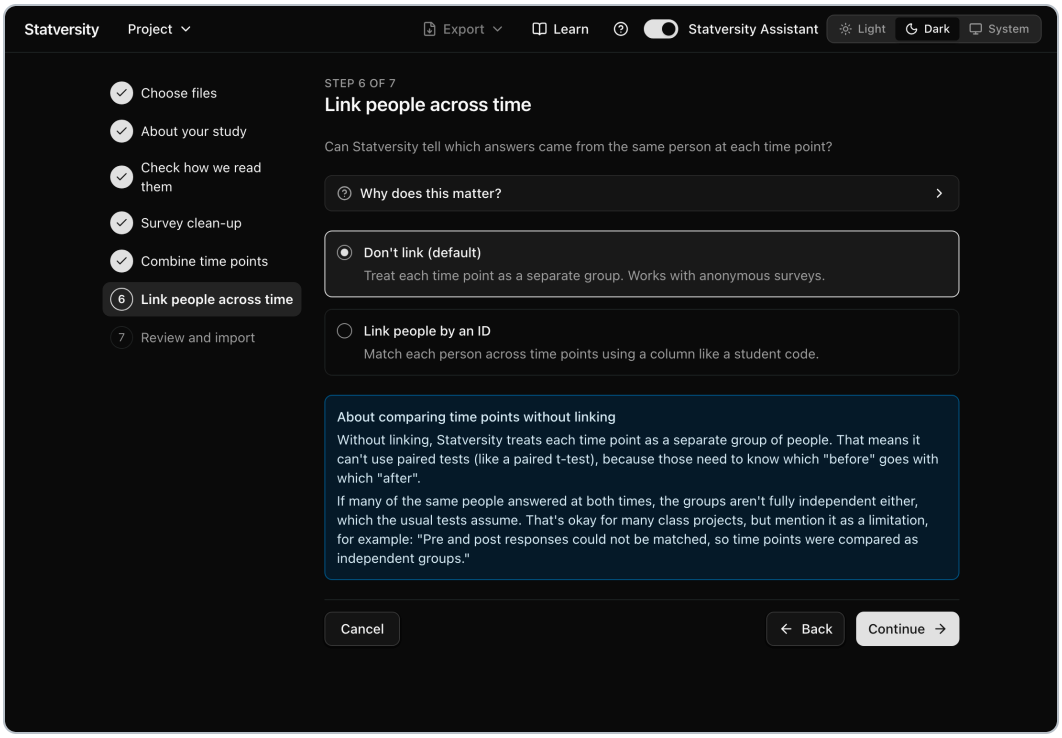


Figure 64. Step 6 of 7, Link people across time, with Don't link (default) picked.

The two options are:

- **Don’t link (default).** “Treat each time point as a separate group. Works with anonymous surveys.”
- **Link people by an ID.** “Match each person across time points using a column like a student code.”

> IN PLAIN LANGUAGE
>
> **Why can’t Statversity just assume the first row of the pre file is the same person as the first row of the post file?** Because row order tells you nothing. Surveys are returned in whatever order people finished them, so row 7 in the “before” file and row 7 in the “after” file are almost certainly different people. To compare a person with themselves you need something that identifies them, like a student ID or a chosen code, in both files. These two files have no such column, and the people answered anonymously.

Leave **Don’t link (default)** picked. The blue box on screen explains the consequence: without linking, Statversity treats each time point as a separate group of people, so it cannot use paired tests (like a paired t test), “because those need to know which ‘before’ goes with which ‘after’.” It also warns that if many of the same people answered at both times, the groups are not fully independent, and suggests mentioning that as a limitation, for example: “Pre and post responses could not be matched, so time points were compared as independent groups.”

Hold on to that sentence. It is what you will carry into your own write-up. Chapter 14 shows what changes when the files do share an ID.

10. Click **Continue**. The screen shows **Review and import**. It summarizes what Statversity will do: two files, 120 rows read, 27 columns kept, survey-system columns hidden, time points Pre to Post (column “Time”), and the last line, **Linking: not linked (separate groups)**. Check that this matches what you meant, then click **Import**.

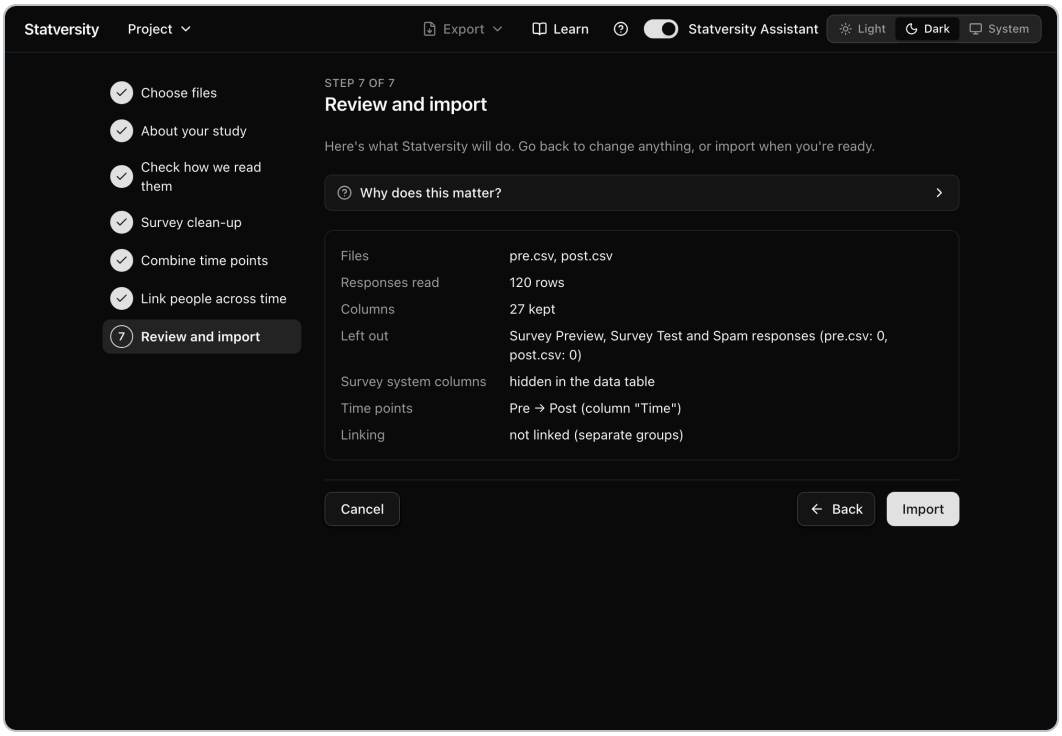


Figure 65. Step 7 of 7, Review and import, ending with Linking: not linked (separate groups).

Steps: telling Statversity what each question is

After the import the screen shows **Welcome**, the start of the Variable Interview. Chapter 6 covers it in full. A **variable** is just one column of your data, and the interview is a short conversation in which Statversity asks what each column means. Statversity has already guessed, so most of the time you check its guess and press **Continue**.

11. On **Welcome**, click **Start**. It notes that there is 1 question to check, because the ten Q3 items are grouped as one matrix question.

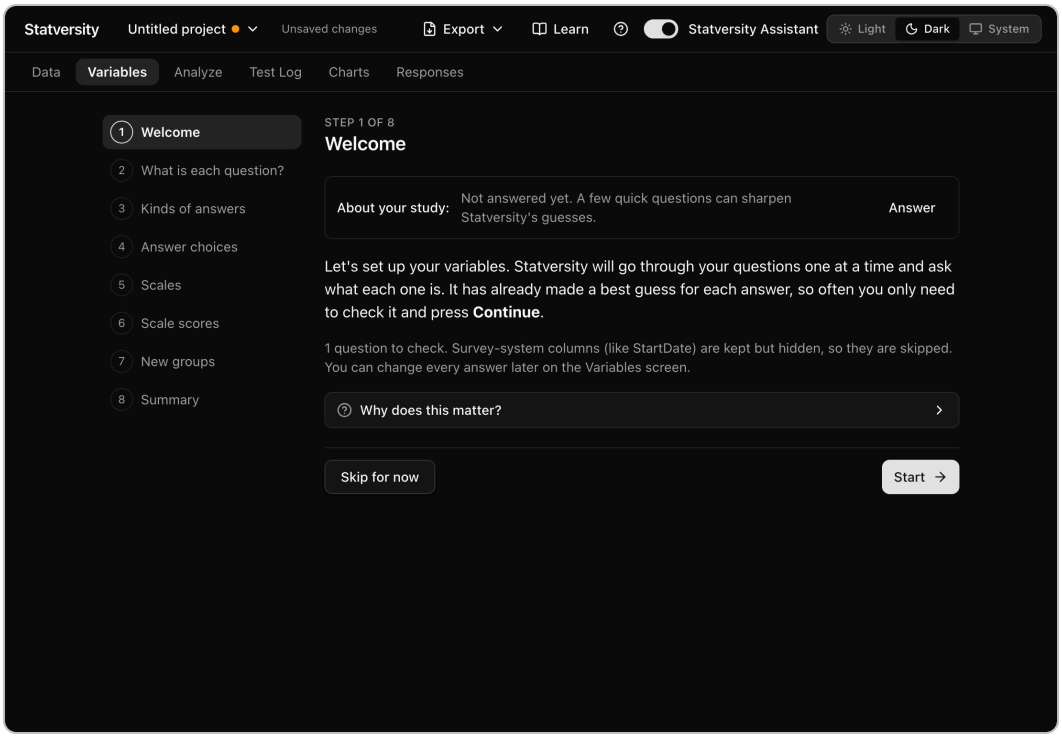


Figure 66. The Welcome step of the Variable Interview.

12. The screen shows **What is each question?** Statversity shows the question, “Matrix item 1: statement about the training program (positively worded),” a few sample answers (4, 3, 2, 5, 1), and a row of choices. It has picked **Survey (Likert) item**, which is right: an agree/disagree rating question. Click **Continue**.

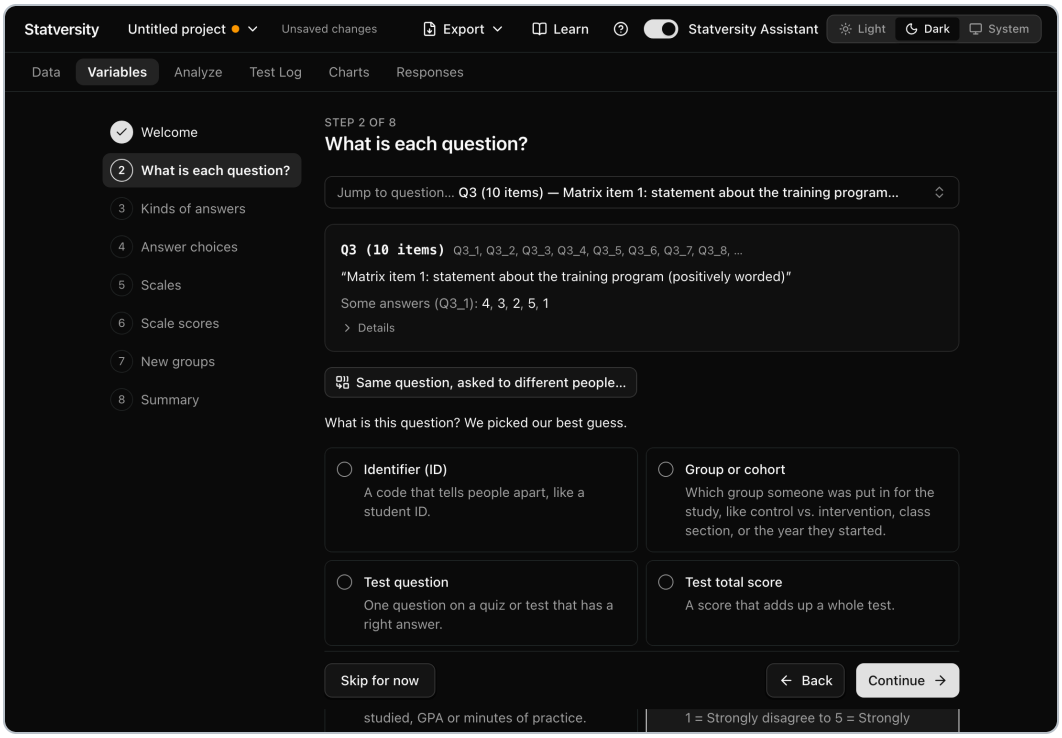


Figure 67. What is each question?, with Survey (Likert) item selected.

13. The screen shows **Kinds of answers**. Statversity picked **Ordered categories (ordinal)**. Read the note in the blue box, because it explains something that surprises many new users: a single agree/disagree question is ordered categories, since "Agree" is more than "Neutral" but the gaps may not be equal. Once several questions are combined into a scale score, that score is usually treated as a number. Click **Continue**.

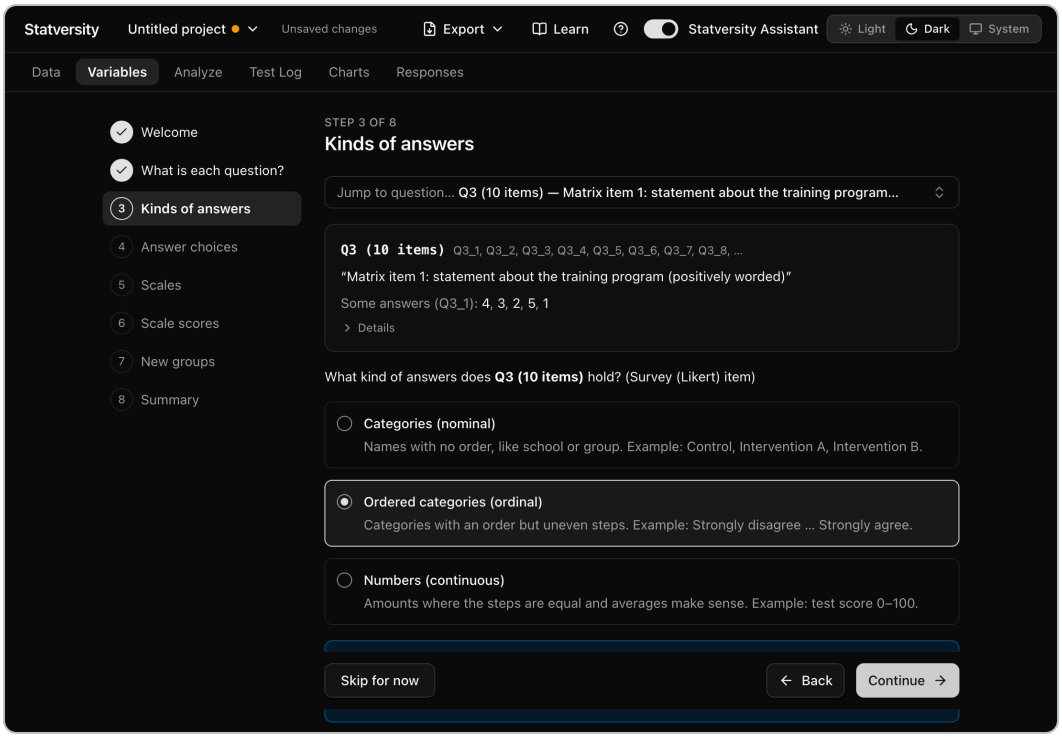


Figure 68. Kinds of answers, with Ordered categories (ordinal) selected and the explanatory note.

> **IN PLAIN LANGUAGE**
>
> **Ordinal versus numbers.** The difference between 1 and 2 on a rating question might not equal the difference between 4 and 5. Averaging a single ordinal question is therefore a little shaky. Averaging ten of them into one score is much more defensible, and that average is what we will compare.

14. The screen shows **Answer choices**. Your file stores the answers as numbers, and Statversity offers to attach labels such as 1 = Strongly disagree. The note says "Numbers still work for every test, so you can skip this." Click **Continue** and keep the plain numbers.

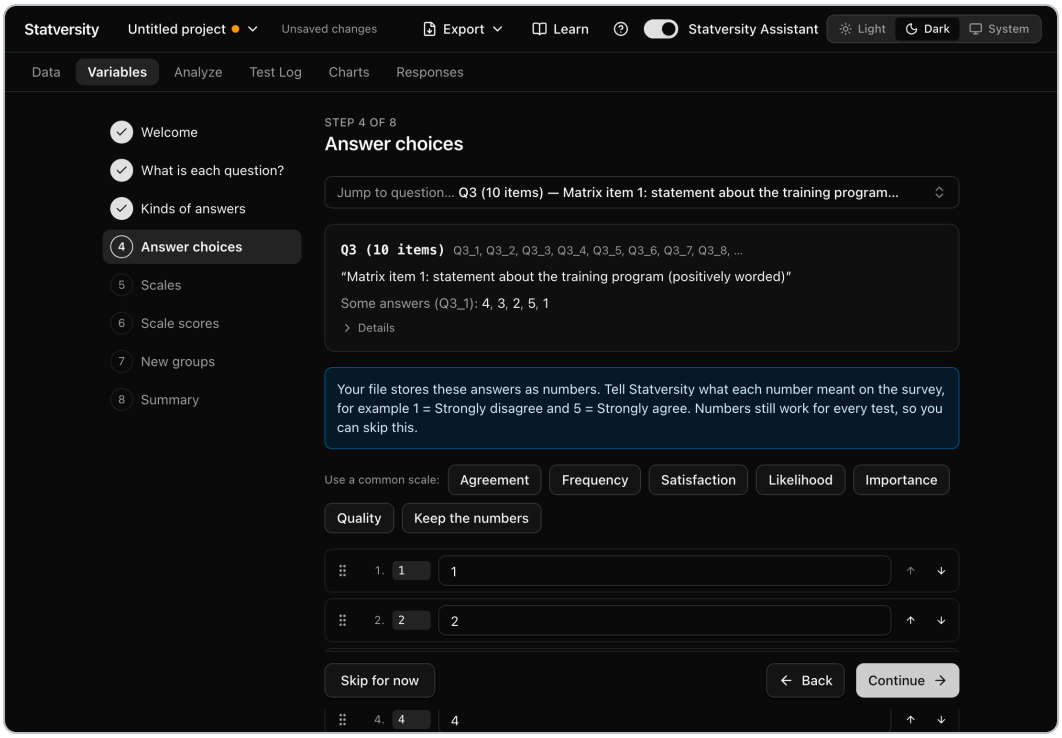


Figure 69. Answer choices, where the numbers 1 to 5 can be given labels but do not have to be.

15. The screen shows **Scales**. A **scale** is a set of questions that together measure one idea. Statversity averages each set into one score per person, and that score is usually what you analyze. Statversity has already grouped `Q3_1` to `Q3_10` into one 10-item scale.

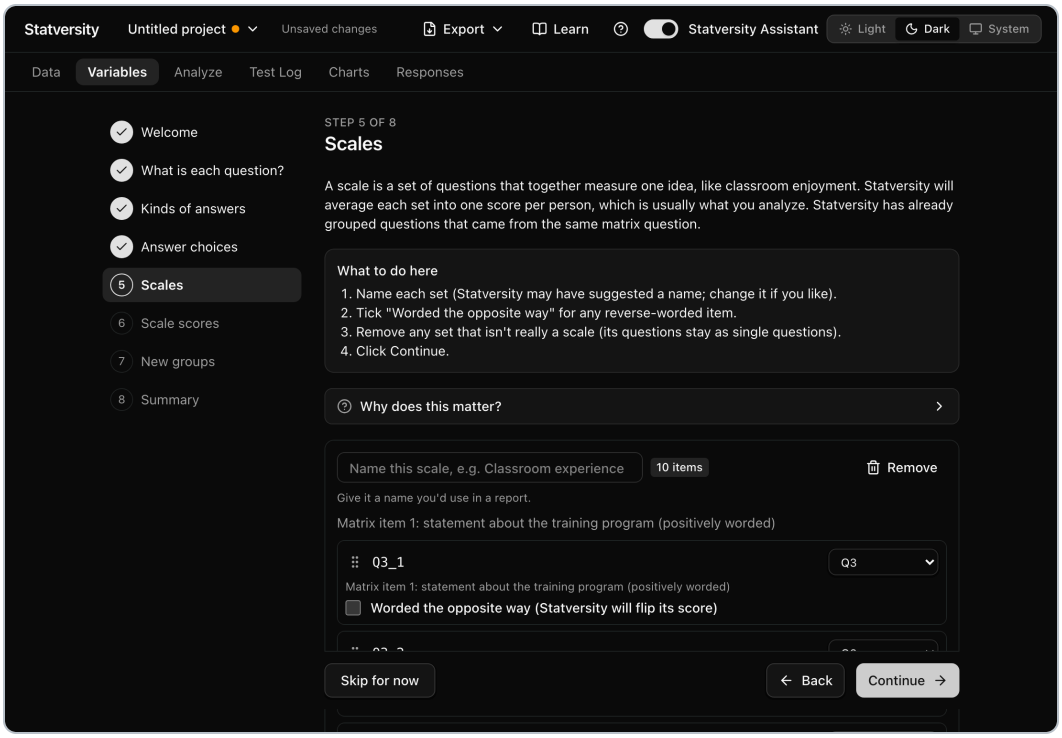


Figure 70. The Scales step, with the Q3 scale of 10 items.

16. Scroll down the list of items. Two of them have **Worded the opposite way (Statversity will flip its score)** already ticked: **Q3_3** and **Q3_8**. Under Q3_3 Statversity says “Ticked because the question text says it’s reverse-worded.” No other item is ticked. Leave them as they are. Click **Continue** when you are ready.

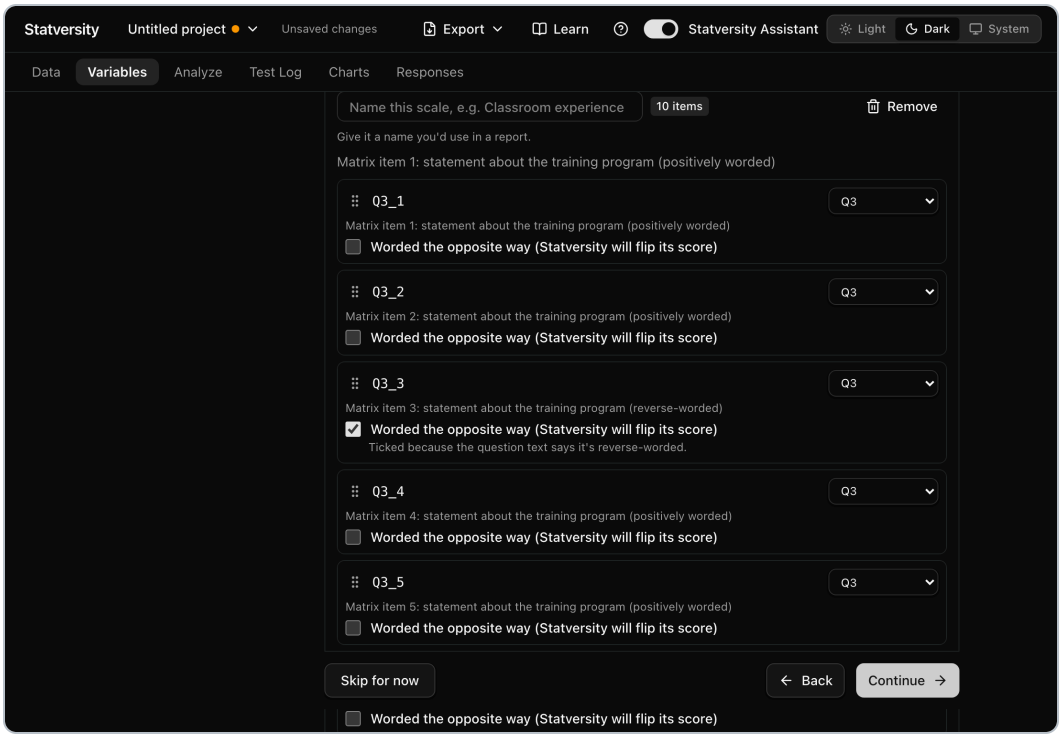


Figure 71. The Scales step scrolled down, with Q3_3 ticked as worded the opposite way.

> **IN PLAIN LANGUAGE**
>
> **Why reverse-score?** Suppose most statements are positive (“The training was useful”) and a few are negative (“The training was a waste of time”). Agreeing strongly with the negative one means 5, but it signals a worse opinion. If Statversity averaged the raw numbers, those items would pull the average the wrong way. Ticking **Worded the opposite way** tells Statversity to flip that item’s 1-to-5 numbers so that a high number always means a more positive opinion. Here Statversity read the words “reverse-worded” in the question text and ticked the right two for you. Always glance at this list, though, because Statversity cannot read intent when the wording is subtle.

17. The screen shows **Scale scores**. Statversity chose **Average of the answered questions (recommended)**, which keeps the score on the same 1-to-5 range as the questions. It will only give a score to people who answered at least 5 of the 10 questions, so someone who skipped most of the survey gets a blank score instead of a misleading one. The defaults suit most surveys. Click **Continue**.

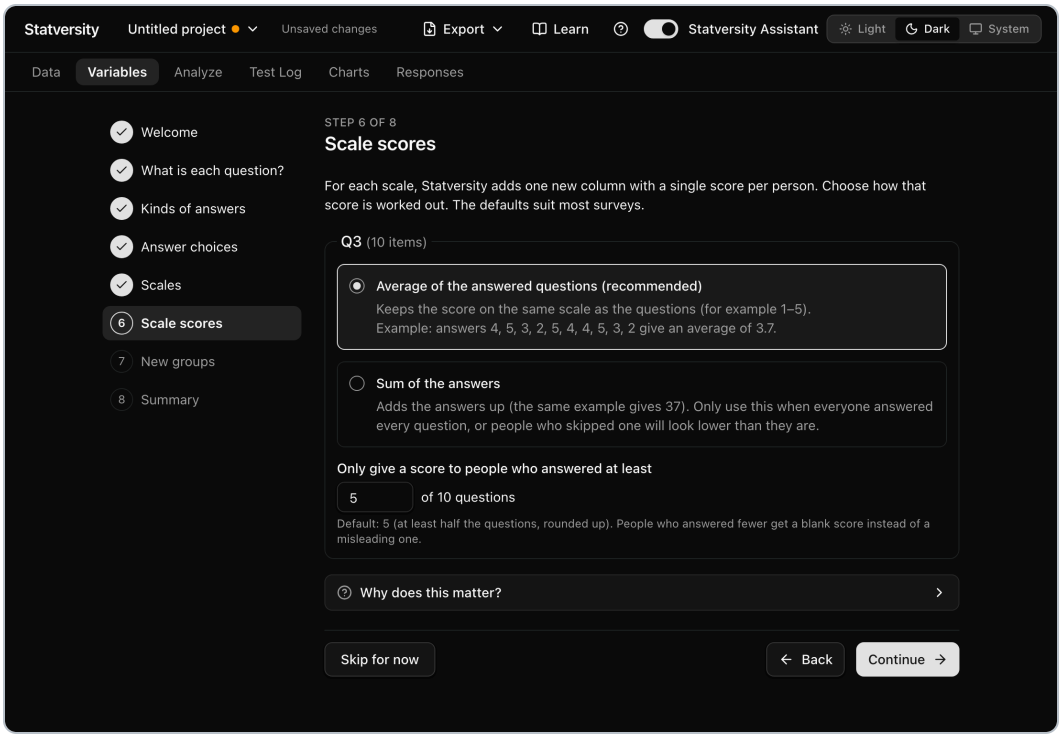


Figure 72. Scale scores, with the average recommended and a minimum of 5 of 10 questions.

18. The screen shows **New groups**, an optional step for building new grouping variables. We do not need one. Click **Continue**.

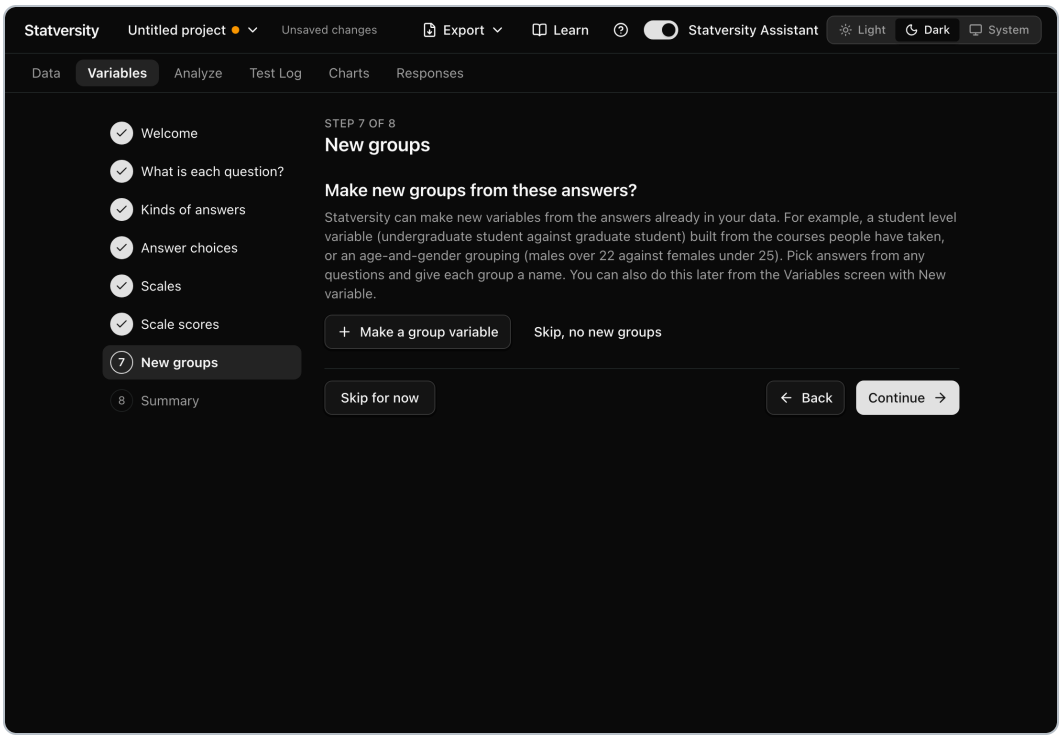


Figure 73. The New groups step.

19. The screen shows **Summary**. It lists what you told Statversity. Read the **Scales** line: “Q3: average of Q3_1, Q3_2, Q3_3, Q3_4, Q3_5, Q3_6, Q3_7, Q3_8, Q3_9, Q3_10 (reverse-scored: Q3_3, Q3_8); needs 5 answered.” The footer says 10 variables to update and 1 scale score to compute. Click **Finish**.

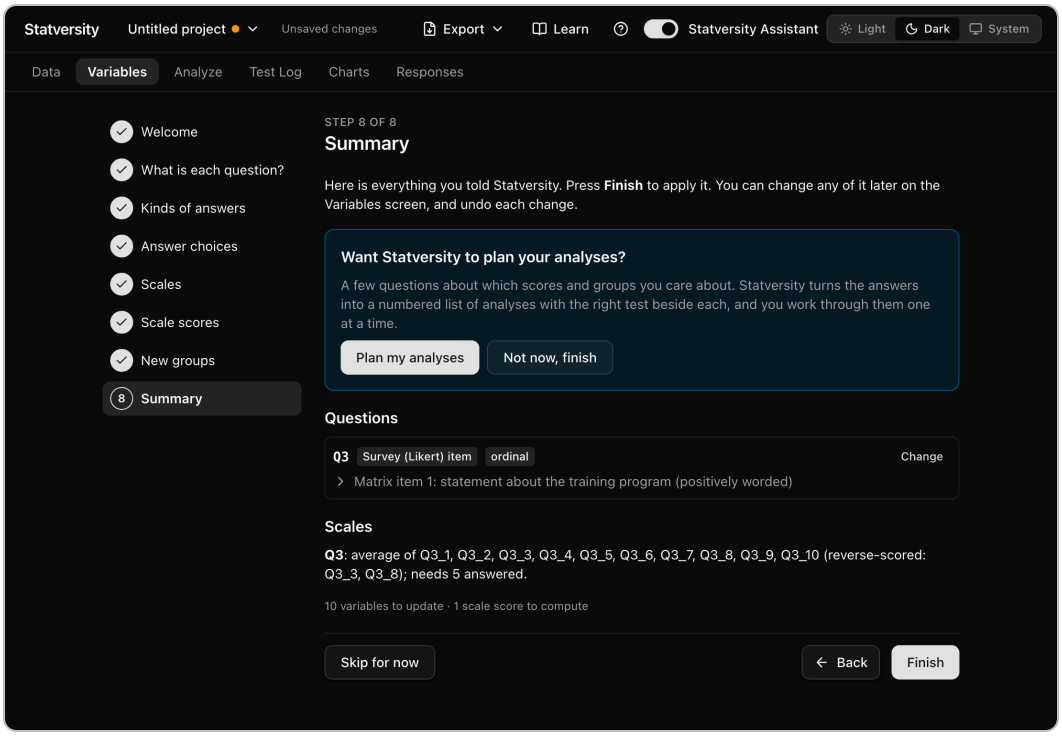


Figure 74. The Summary step with the Q3 scale described in one line.

20. The screen shows **Variables**, a table of every column. Notice the new column at the bottom, **Q3 (average score)**, with the role **Scale score** and the level **continuous**. Statversity made it by averaging the ten items for each person. Notice also the **Time** column at the top, with the role **Time point**. Those are the two columns we will use.

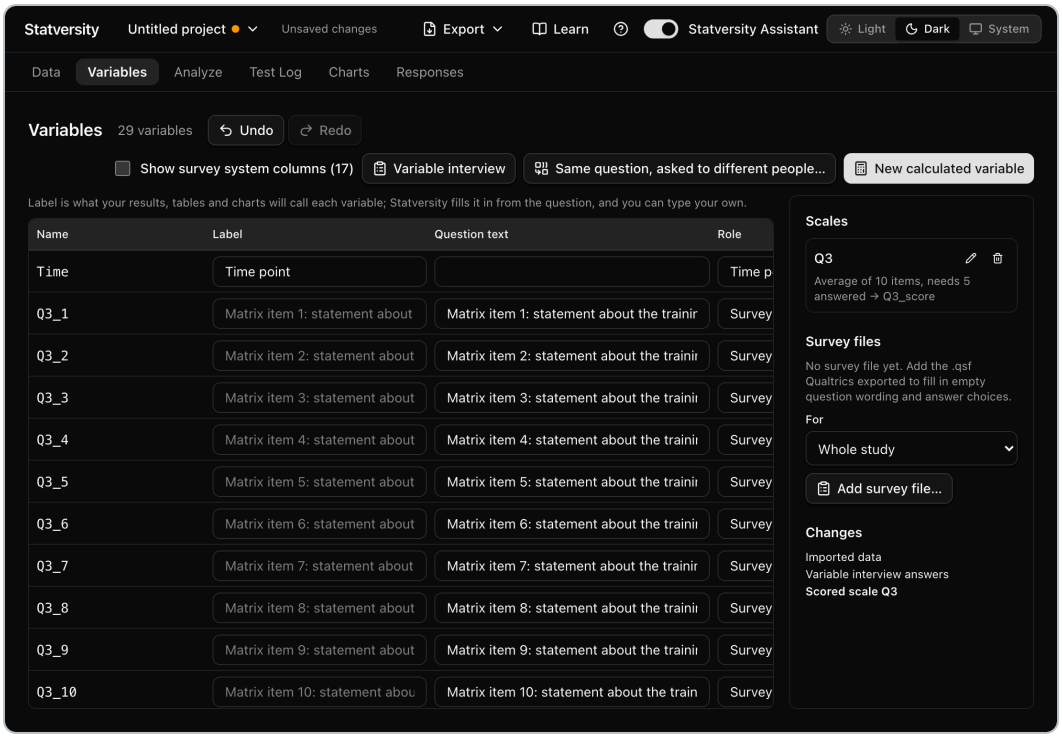


Figure 75. The Variables screen, with the Time point column at the top and Q3 (average score) at the bottom.

Steps: asking Statversity which test to use

21. Click **Analyze** in the top row of tabs. The **Test Advisor** opens with the question “What do you want to know?” Chapter 7 describes the Advisor in detail. It asks a few plain-language questions and then recommends a test.

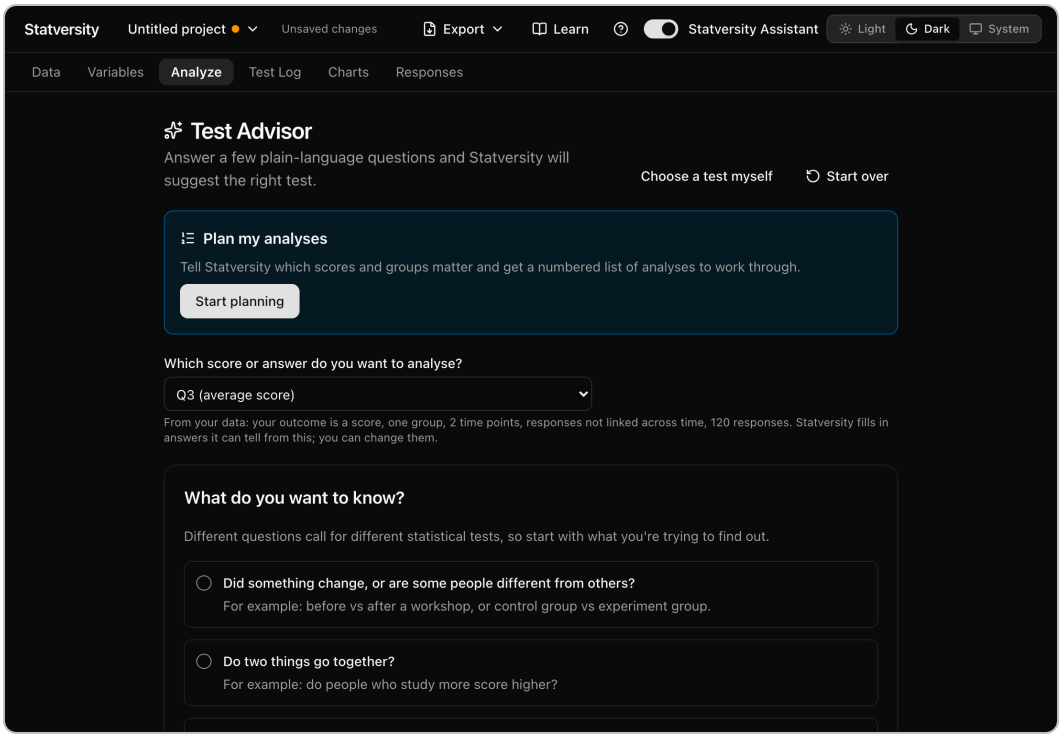


Figure 76. The Test Advisor, ready to ask What do you want to know?

22. Under **Which score or answer do you want to analyse?**, choose **Q3 (average score)**. Below it Statversity summarizes what it figured out from your data: “your outcome is a score, one group, 2 time points, responses not linked across time, 120 responses.”

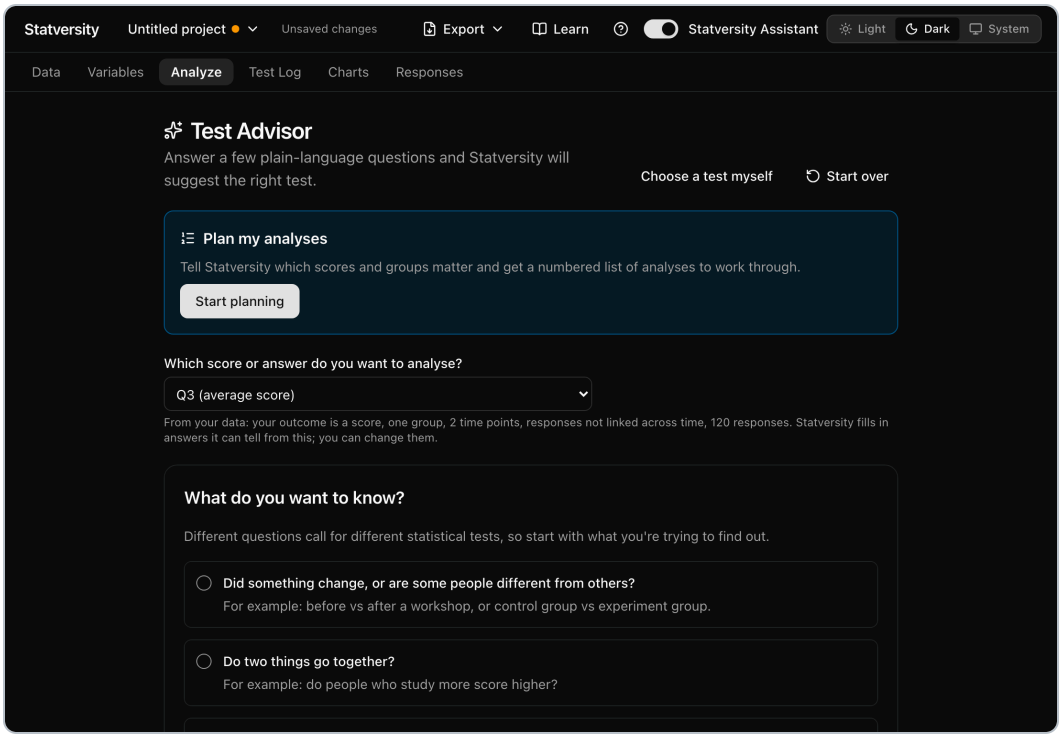


Figure 77. The outcome picker set to Q3 (average score), with Statversity's summary of the data.

23. Under **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**.

> **IN PLAIN LANGUAGE**
>
> **What is this question really asking?** It is the "big picture" question: are you looking for a difference, a relationship, a prediction, or whether a set of survey items hangs together? Ours is a difference question. We want to know if scores changed between two occasions, so this is the right answer. (If instead you wanted to know whether hours studied goes with scores, you would pick "Do two things go together?")

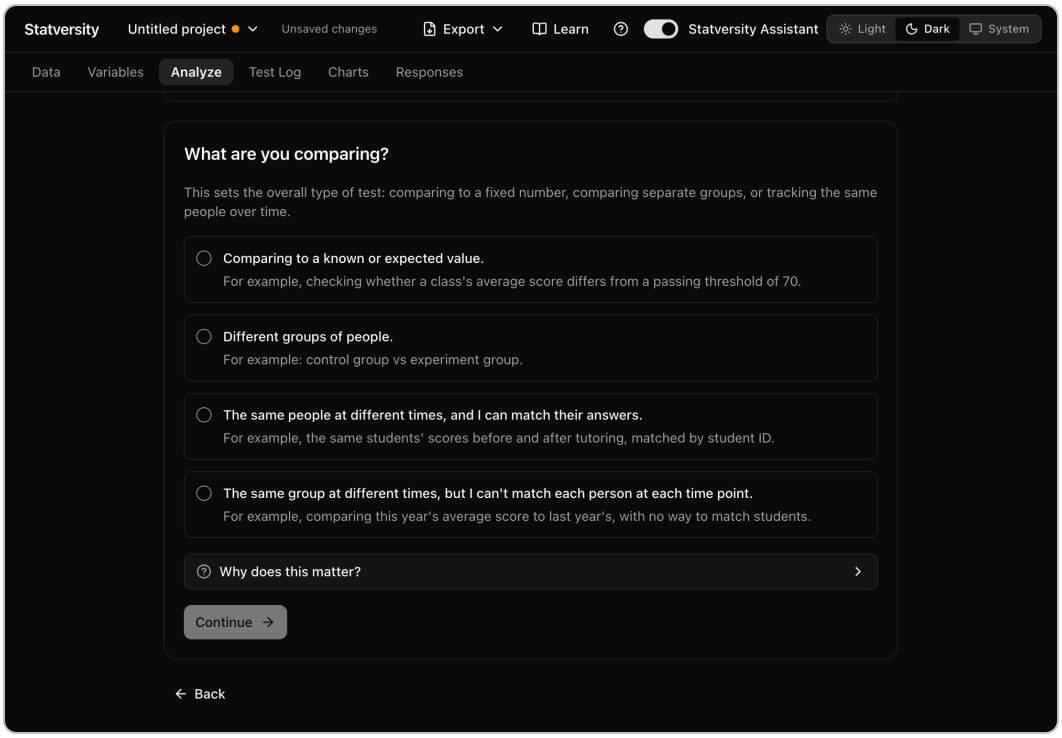


Figure 78. Under What are you comparing?, showing the four options.

24. The next question is **What are you comparing?** Choose **The same group at different times, but I can't match each person at each time point.** and click **Continue**.

> **IN PLAIN LANGUAGE**
>
> **This is the answer that matches the import choice.** There are two "same group at different times" answers: one where you can match each person and one where you cannot. Because we chose **Don't link** during import, we cannot match, so we pick the second one. Statversity had even noticed this on its own, since the summary line said "responses not linked across time." Picking the matched version here would be a mistake, because the data have no matching to use.

25. The Advisor shows **Statversity recommends: Independent-samples t test**. An orange box explains: "Your time points are compared as if they were separate groups of people, because the answers can't be linked to the same students. That is the honest choice, but it is less powerful than a paired test and can't tell you how individual students changed. If you can link responses (for example with a student ID), a paired test is better."

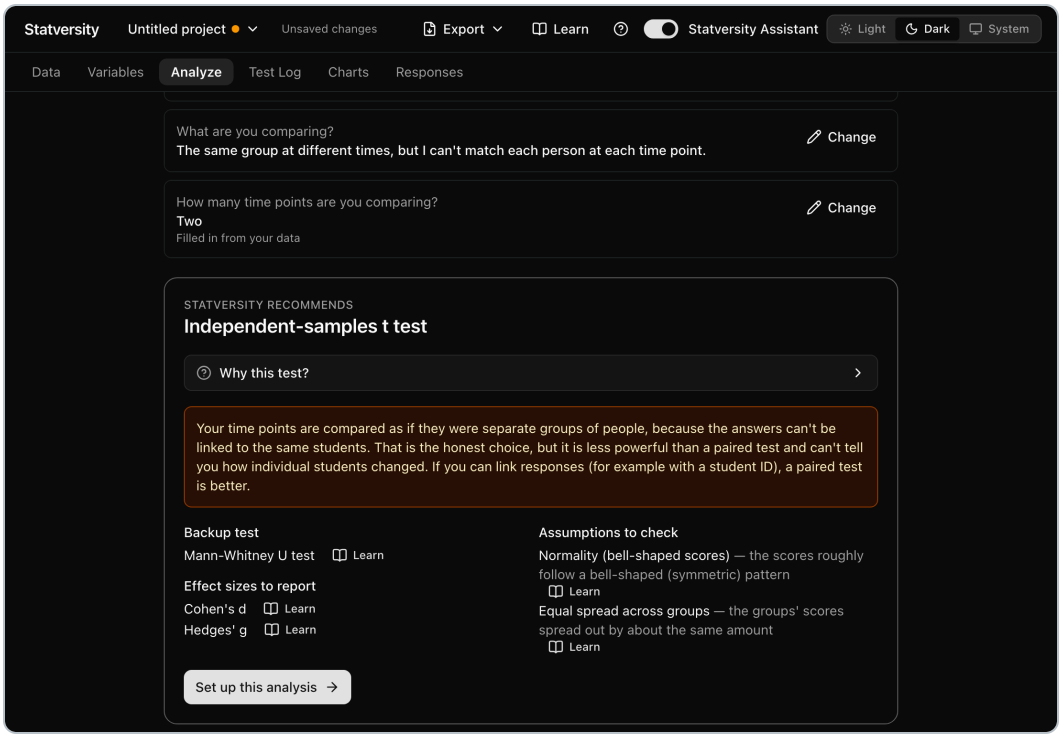


Figure 79. The recommendation, Independent-samples t test, with the orange warning about unlinked responses.

Statversity also lists a **backup test** (the Mann-Whitney U test), the effect sizes it will report (Cohen's d and Hedges' g), and the **assumptions to check**: normality (bell-shaped scores) and equal spread across groups.

> IN PLAIN LANGUAGE
>
> **Independent-samples t test, in one paragraph.** A t test compares the averages of two groups and asks whether the gap between them is bigger than you would expect from luck alone. "Independent" means the two groups are made of different people, so nobody's score in one group is connected to anybody's score in the other. Statversity is treating the Pre group and the Post group as if they were two separate classes. That is a stand-in for the truth (they are probably mostly the same people), and the orange box is Statversity being honest about that.

26. Click **Set up this analysis**. The screen shows **Choose your variables**.

Steps: choosing the variables and checking assumptions

27. On **Choose your variables**, check three things. **Analysis** shows **Independent-samples t test (recommended)**. **Outcome (scores)** shows **Q3 (average score)**. **Groups to compare** shows **Time point**. Statversity filled all three from your variable roles, so there is nothing to change.

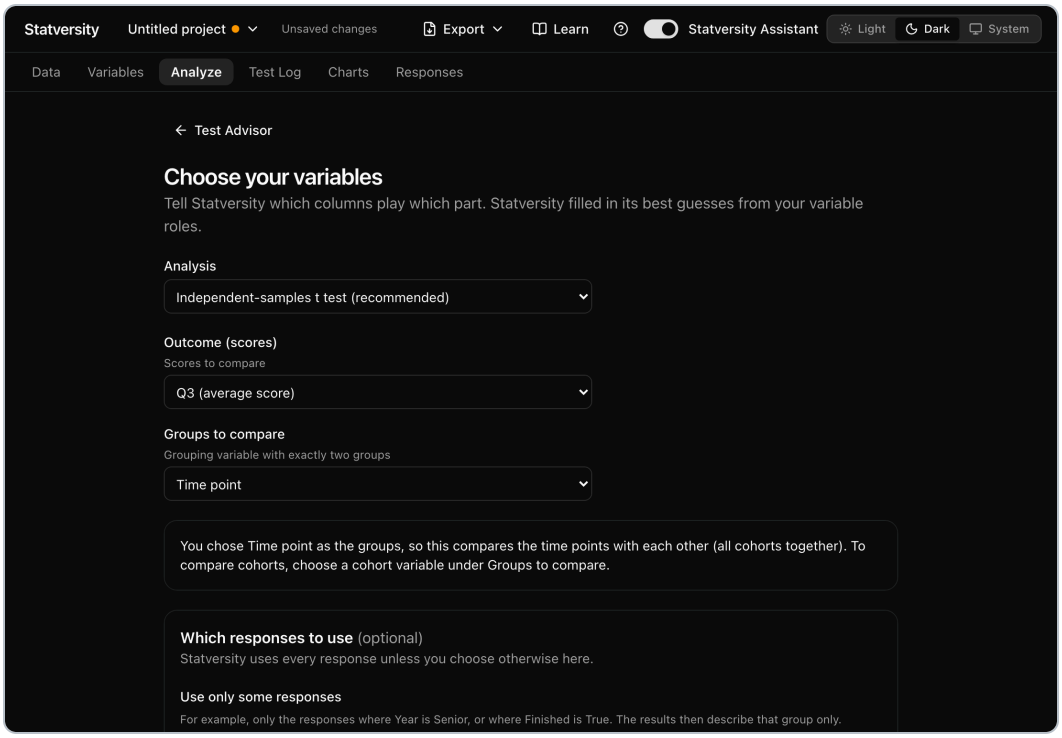


Figure 80. Choose your variables, with Q3 (average score) as the outcome and Time point as the groups.

A note on the screen says "You chose Time point as the groups, so this compares the time points with each other (all cohorts together)." The optional **Which responses to use** box lets you filter or split, and we leave both alone ("All responses (no filter)" and "Don't split (one result)").

28. Click **Check the assumptions**. A t test works properly only if your data behave in certain ways. Statversity checks three things, one screen each. If you want the general idea first, chapter 23 explains assumptions in plain language.

Check 1 of 3: normality for Pre

The first check is **Normality: Pre**. "Normality" asks whether the scores form a roughly bell-shaped pile, with most people in the middle and fewer at the extremes. Statversity used the Shapiro-Wilk test on the 60 Pre scores and found $W = 0.99$, $p = .764$. The result is **Looks fine**.

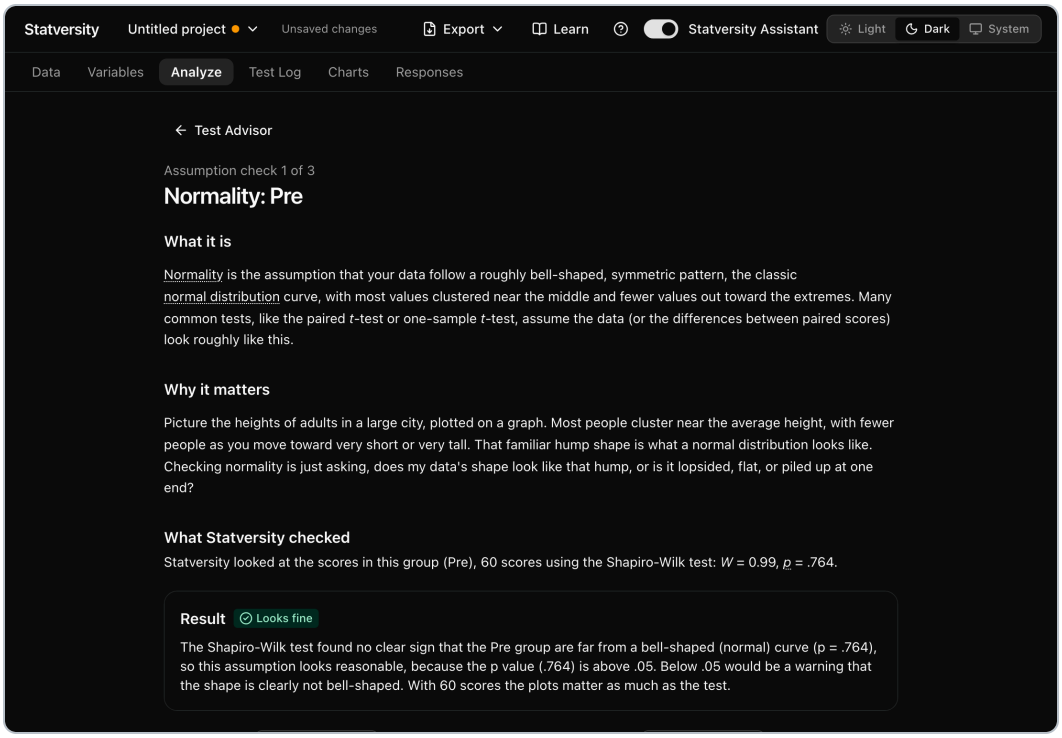


Figure 81. Normality for the Pre group, Looks fine with p = .764, with its Q-Q plot and histogram starting below.

> IN PLAIN LANGUAGE
>
> **For assumption checks, you want a big p-value.** In the main test a small p-value is the interesting finding. In an assumption check it is the opposite. The check's "boring" starting idea is "the data are fine," so a p-value above .05 (here .764) means there is no sign of a problem. Statversity says exactly that in the Result box. The pictures matter too: the Q-Q plot should hug the dashed line, and the histogram should look like a hump, which is what you see.

29. Click **Next check**.

Check 2 of 3: normality for Post

The second check is **Normality: Post**. The 60 Post scores give W = 0.97, p = .158, also **Looks fine**. A p-value of .158 is smaller than the Pre group's, but it is still well above .05.

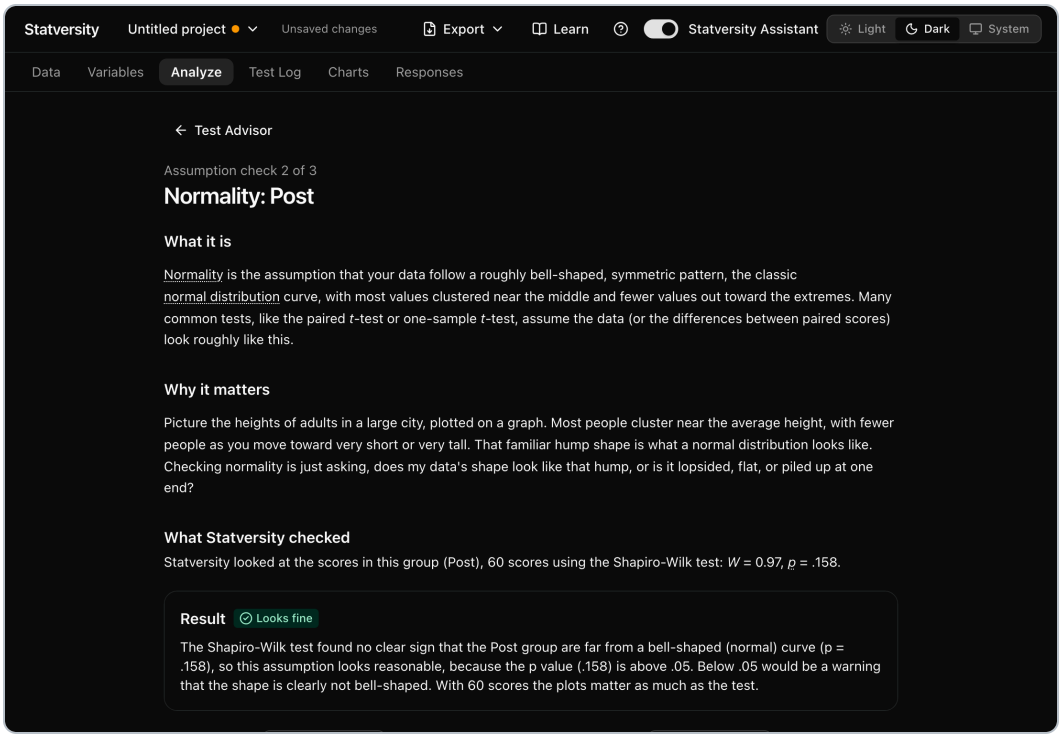


Figure 82. Normality for the Post group, Looks fine with p = .158.

30. Click **Next check**.

Check 3 of 3: equal spread

The third check is **Equal spread (homogeneity of variance): Pre vs Post**. This asks whether the two groups are scattered by about the same amount around their own averages. Statversity used Levene's test (the Brown-Forsythe version) on all 120 scores and found $F(1, 118) = 0.47$, $p = .493$. Again **Looks fine**: "The groups' scores are spread out by similar amounts."

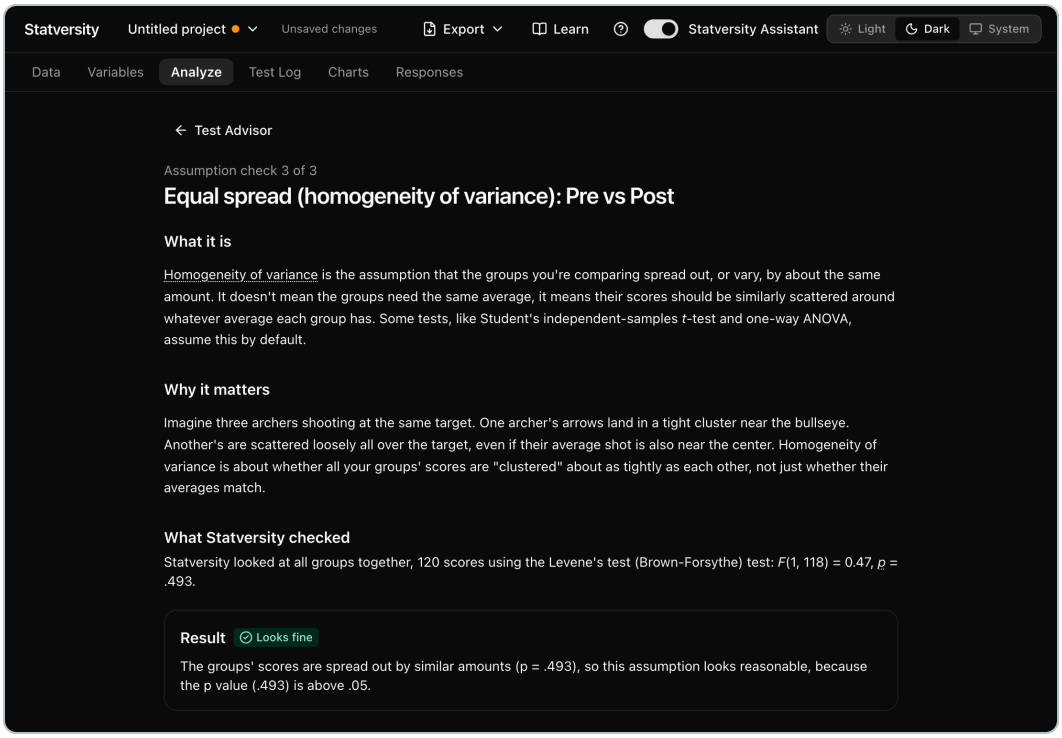


Figure 83. Equal spread for Pre and Post, Looks fine with p = .493.

31. Click **Decide which test to use**.

Steps: the decision and the result

The screen shows **Which test should you use?** and gathers the three checks in one place, each with a green **Looks fine** tag. Below them Statversity says: “Statversity suggests: Independent-samples t test. Every check looks reasonable, so the recommended test is a good fit.” You can pick **Use the recommended test: Independent-samples t test** or the **nonparametric alternative, the Mann-Whitney U test**, which ranks the scores instead of averaging them and does not need bell-shaped data. A tick at the bottom confirms “This matches Statversity’s suggestion.”

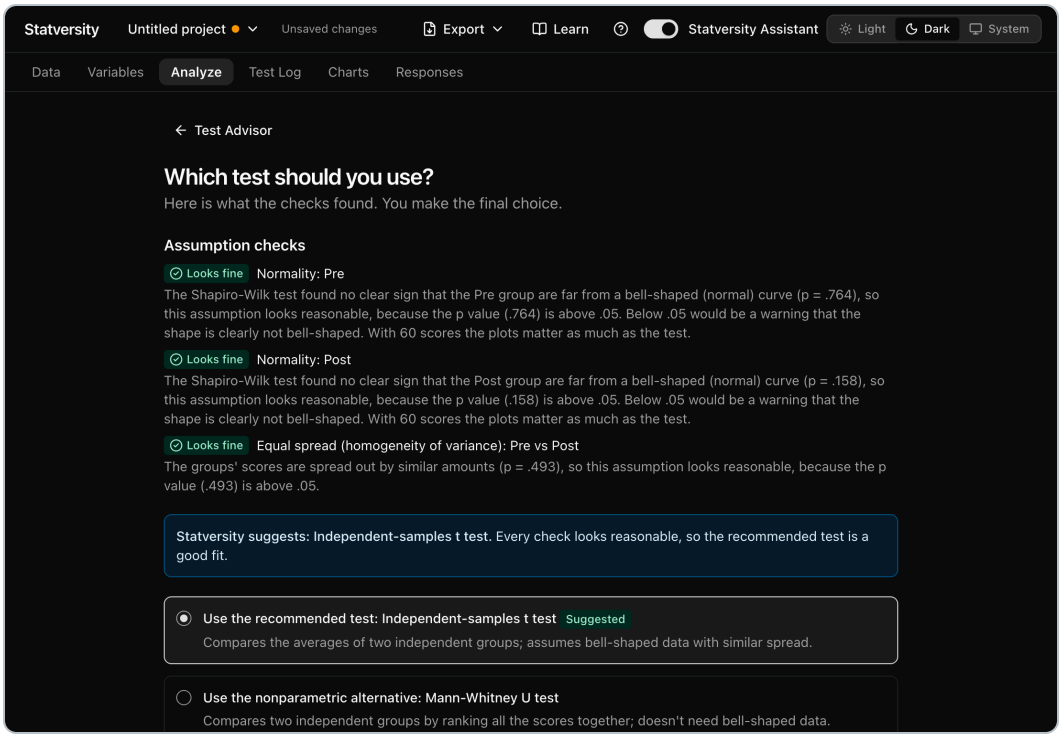


Figure 84. Which test should you use?, with all three checks Looks fine and the recommended test selected.

> **TIP**
>
> You make the final choice. Statversity recommends, but it will let you run the other test if you have a reason, for example a plan you wrote down before collecting data. In this walkthrough every check passed, so the recommended test needs no second guessing.

32. Keep the recommended test and click **Run and see the results**.

The Results screen opens for the **Independent-samples t test**. The line under the heading says “These 120 responses come from the same survey given at 2 time points (Pre: 60; Post: 60), not from 120 different people. People could not be matched across time points, so each time point is compared as a separate sample.” That is Statversity being careful: the pre and post surveys were stacked without a student ID (chapter 14 shows how to link them), so the 120 is a count of answer sheets, 60 from each survey, and not 120 different students. The test itself compares the 60 Pre answers with the 60 Post answers.

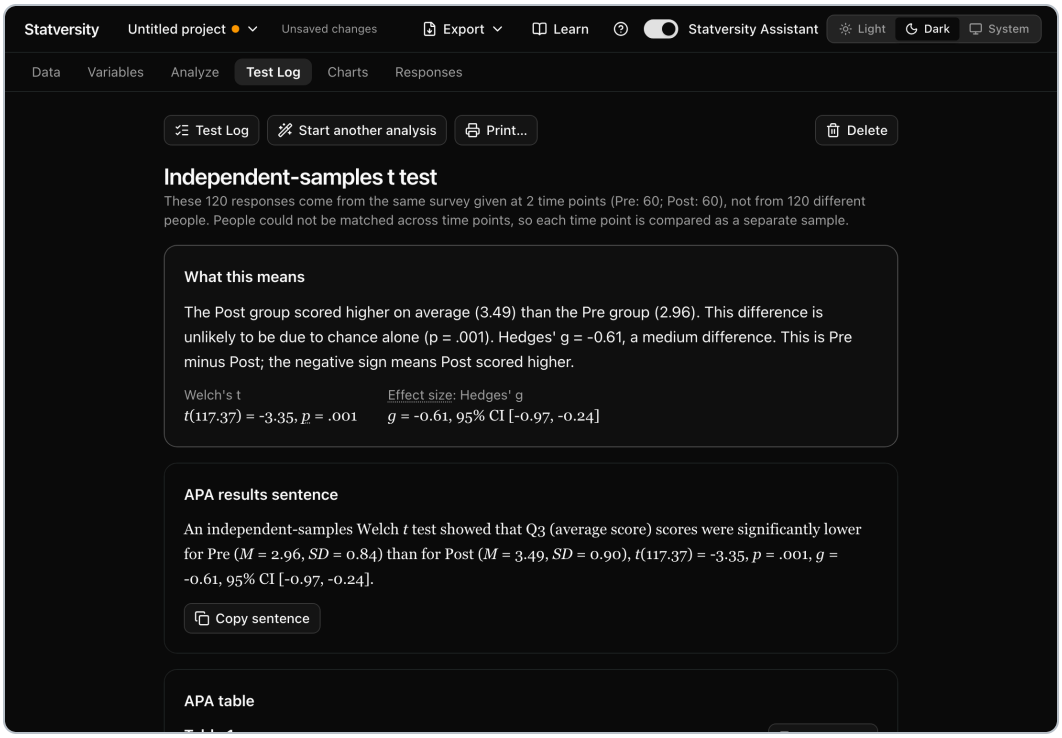


Figure 85. The top of the results, showing the What this means box and the APA sentence.

33. Read **What this means** at the top of the result. This is Statversity's plain-language summary.

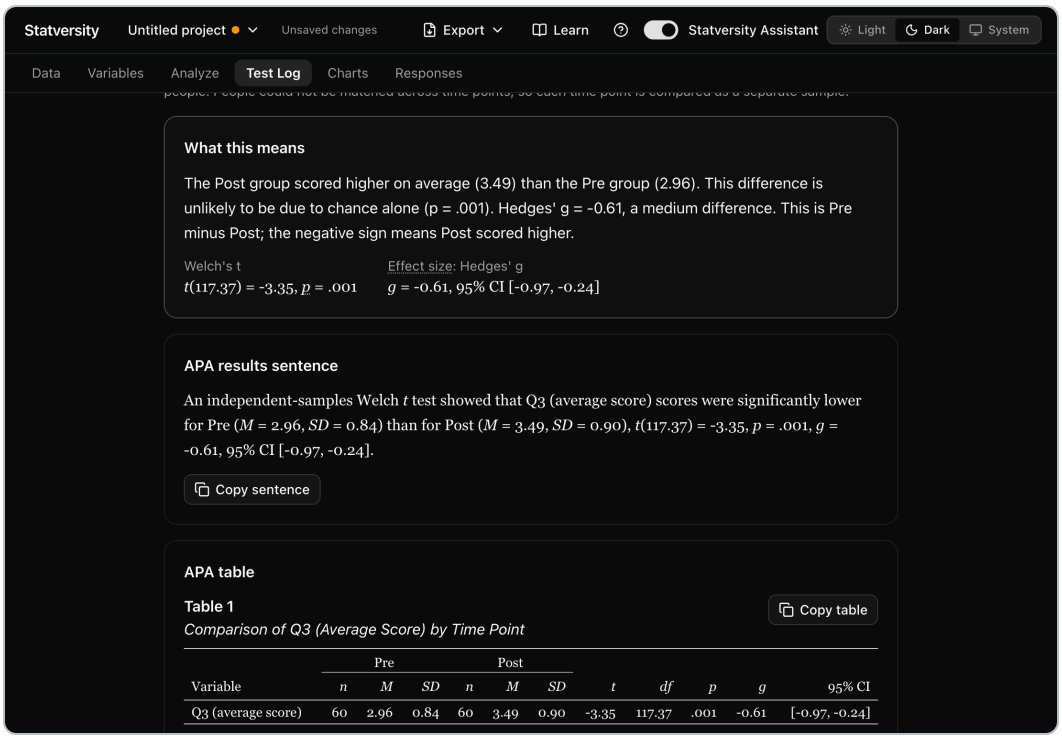


Figure 86. The What this means box, with the Welch t statistic and the Hedges' g effect size.

34. Scroll down to **APA results sentence** and **APA table**. Chapter 8 explains every block on this screen.

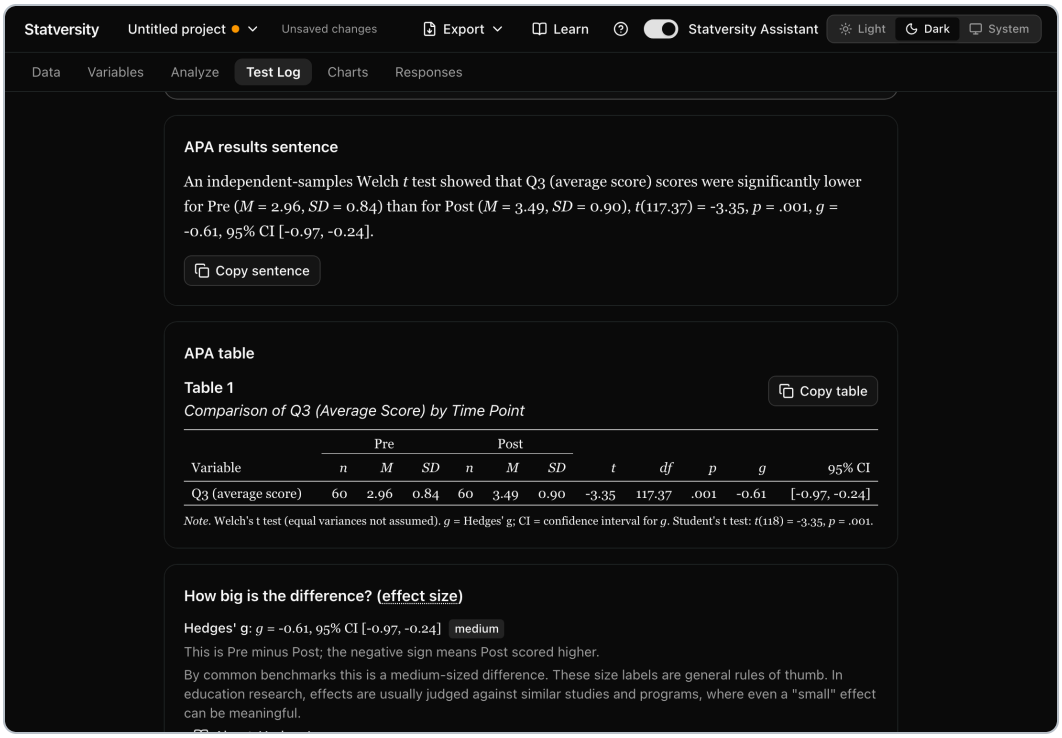


Figure 87. The APA results sentence and the APA table.

35. Keep scrolling to the charts. The first chart on the result is a Q-Q plot for the Pre group, the same normality picture you saw earlier, now saved with the result. Each chart has **Save figure…** and **Edit chart** buttons.



Figure 88. The normality charts that travel with the result, with Show the numbers links.

36. Click **Test Log**. Statversity keeps a record of every analysis you run. The entry shows **Independent-samples t test**, the variable **Q3_score**, the first line of the APA sentence, the date, "n = 120 responses over 2 time points", and p = .001. You can come back and reopen it any time, and it is saved with your project.

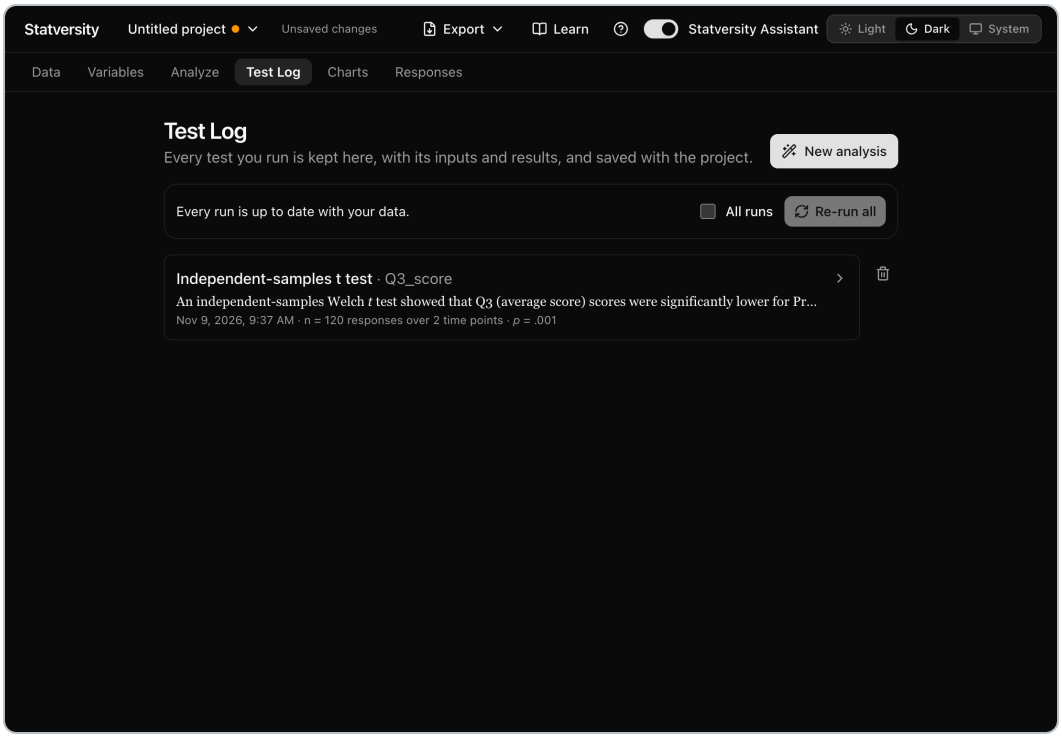


Figure 89. The Test Log with the finished analysis listed.

What Statversity found

	Pre	Post
People (n)	60	60
Mean (M)	2.96	3.49
Standard deviation (SD)	0.84	0.90
Median	2.95	3.50

Result	Value
Test	Independent-samples t test (Welch’s version)
Statistic	t(117.37) = -3.35
p-value	.001
Effect size	Hedges’ g = -0.61, 95% CI [-0.97, -0.24]
Label	medium

Cohen’s d, which Statversity also reports, comes out at -0.61 with 95% CI [-0.98, -0.24].

A note on the version of the test. The result says “Welch’s t,” and the APA sentence says “Welch t test.” Welch’s t test is the version of the independent-samples t test that does not require the two groups to be equally spread out. That is why the degrees of freedom are a fractional 117.37 rather than a whole number like 118. Our equal-spread check passed, so the choice makes almost no difference here. The note under the APA table gives the classic Student’s version for comparison, t(118) = -3.35, p = .001. Both tell the same story.

> **IN PLAIN LANGUAGE**
>
> **Reading the minus sign.** Statversity always subtracts in the order Pre minus Post. Post scored higher, so Pre minus Post is negative. A negative g does not mean the training backfired. Statversity spells this out right next to the number: “This is Pre minus Post; the negative sign means Post scored higher.”

What this means in plain English

The Post group’s average rating of the training (3.49 out of 5) was higher than the Pre group’s (2.96). That gap is bigger than you would expect by luck alone: p = .001 means that if the training did nothing, you would see a gap this large (or larger) only about 1 time in 1,000. The gap is also not tiny. Hedges’ g = -0.61 says the two groups are about six-tenths of a standard deviation apart, which Statversity labels a medium difference. The 95% confidence interval runs from -0.97 to -0.24, so even the most cautious end of the range is a small difference in the same direction.

Because the two files could not be matched, we cannot say that any particular person improved. We can say that, as a group, the Post responses sit higher than the Pre responses. For a reminder of how to read p-values, effect sizes, and confidence intervals, see chapter 23.

What you would write in a paper

Statversity generates this sentence for you, and you can copy it with the **Copy sentence** button:

> An independent-samples Welch *t* test showed that Q3 (average score) scores were significantly lower for Pre (*M* = 2.96, *SD* = 0.84) than for Post (*M* = 3.49, *SD* = 0.90), *t*(117.37) = -3.35, *p* = .001, *g* = -0.61, 95% CI [-0.97, -0.24].

To adapt it, replace “Q3 (average score)” with the name of your scale as readers will know it (for example, “satisfaction with the training”), and add a sentence naming the limitation Statversity suggested: “Pre and post responses could not be matched, so time points were compared as independent groups.” The **Copy table** button gives you the APA table, ready to paste.

What if

What if the data had been linked by an ID? Then you would choose **Link people by an ID** at the import step, and the Advisor would recommend a **paired-samples t test**. That test compares each person with themselves, which removes the differences between people and usually gives a more sensitive result. Chapter 14 walks through exactly that, including what happens when the IDs do not match perfectly.

What if the assumption check had failed? Suppose the Post normality check had given a p-value below .05, with a Q-Q plot bending far from the line. Statversity would not show **Looks fine** for that check, and its recommendation would move toward the **Mann-Whitney U test**, which works on ranks. If instead only the equal-spread check had failed, the Welch version, which you already used, is designed for that situation. Chapter 23 explains these repairs.

What if the effect had been small? Say g had been -0.15 with a p-value of .40. The plain-language summary would say the difference could easily be due to chance, and the effect-size label would read negligible. You would report it exactly as you do here, just with a different conclusion. A non-significant result in a small sample does not show that the training did nothing. It shows that this study could not tell.

What if I want the answer for each question rather than the average? You could skip the scale step and run the test on one item at a time, but ten separate tests invite false alarms. Averaging the ten into a scale score first, as you did here, is the standard approach.

> **TRY IT YOURSELF**
>
> **Try it yourself.** Run the same analysis again, but this time choose the **nonparametric alternative: Mann-Whitney U test** on the **Which test should you use?** screen. Compare its result with the t test: is the direction the same? Is the p-value close? Then write one sentence explaining why the two tests agree, and check the Test Log for the second entry.

CHAPTER 12

Which intervention worked best?

The question

“We tried two new teaching interventions, A and B, and also kept a Control group that got neither. Which one gave the best test scores?”

This is a **comparison between three separate groups of people**. Nobody is in more than one group, and you want to know whether the groups’ average results differ by more than luck would explain. Researchers sometimes call this a between-groups design, and the classic tool for it is a one-way ANOVA (analysis of variance), which compares three or more groups at once.

This chapter has a twist that makes it a good teaching example. Statversity started out recommending a one-way ANOVA. Then it checked whether your data were in a shape that test can handle, found one group that was not, and switched to a more cautious test. You will see exactly how and why that happens, and you do not have to understand any of the arithmetic to follow it.

> **NOTE**
> Screens that are the same in every walkthrough, such as Home, the import steps and the Variable Interview, are explained once in the reference chapters: Getting data in (chapter 5), Variables (chapter 6), Analyze (chapter 7) and Results (chapter 8). This chapter tells you which button to press and why the answer is right for this study, and points you to those chapters for the controls themselves.

The data

The practice dataset is a classroom-style study with three files: `pre.csv`, `post.csv` and `followup.csv`. The same 20-question multiple-choice test was given three times: before the interventions (Pre), right after them (Post), and a few weeks later (Follow-up). Every person also answered one question, **Which group are you in?**, with three possible answers: Control, Intervention A or Intervention B.

Two details matter.

First, the test is **not pre-scored**. The files contain the raw answer each person picked (A, B, C or D) for each of the 20 questions. A separate answer key tells Statversity which letter is correct for each question, and Statversity does the marking for you.

Second, the three files hold **different numbers of people**. The screenshot you will see on the Review and import screen shows 352 responses in total. Pre has 135, Post has 119 and Follow-up has 98. Fewer people turned up each time, which is normal in real studies when some participants drop out. Statversity does not know whether the same person appears in more than one file, and in this walkthrough we treat each time point as its own set of people.

To open it, go to Home, click **Open practice data**, and find the card titled **Which intervention worked best?**.

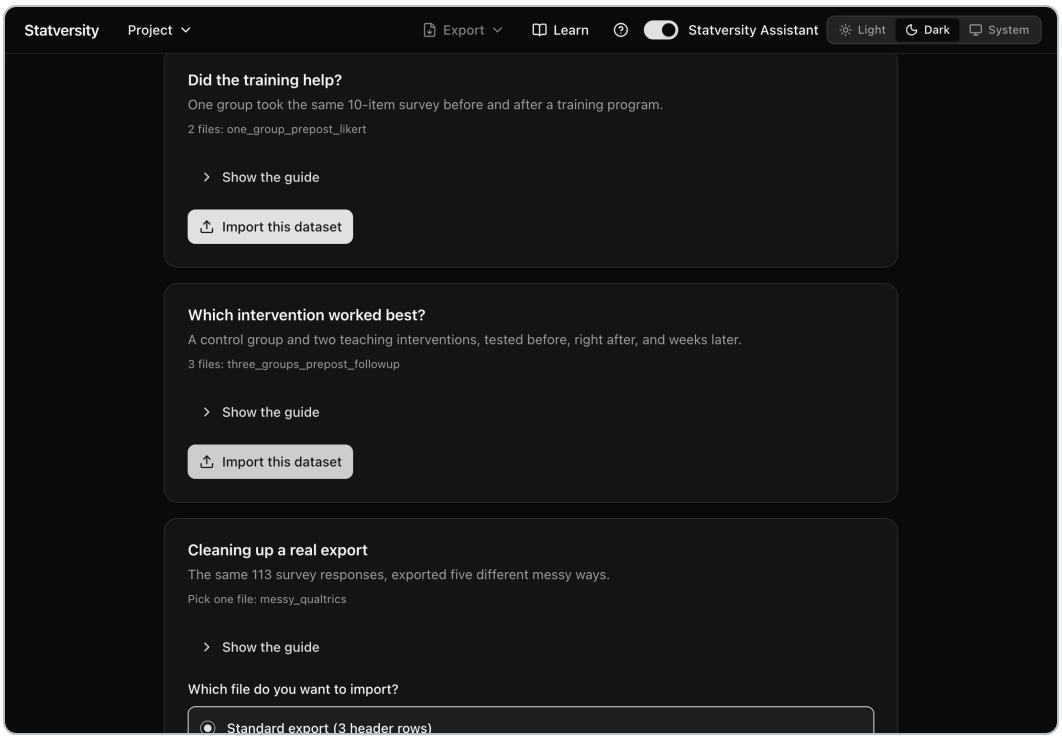


Figure 90. The Practice data screen with the card titled Which intervention worked best? in view, showing its Import this dataset button.

Steps

Part 1: Import the three files

1. On Home, click **Open practice data**. The **Practice data** screen lists the practice datasets.
2. On the card titled **Which intervention worked best?**, click **Import this dataset**. The **Choose files** screen opens with the three files already filled in.

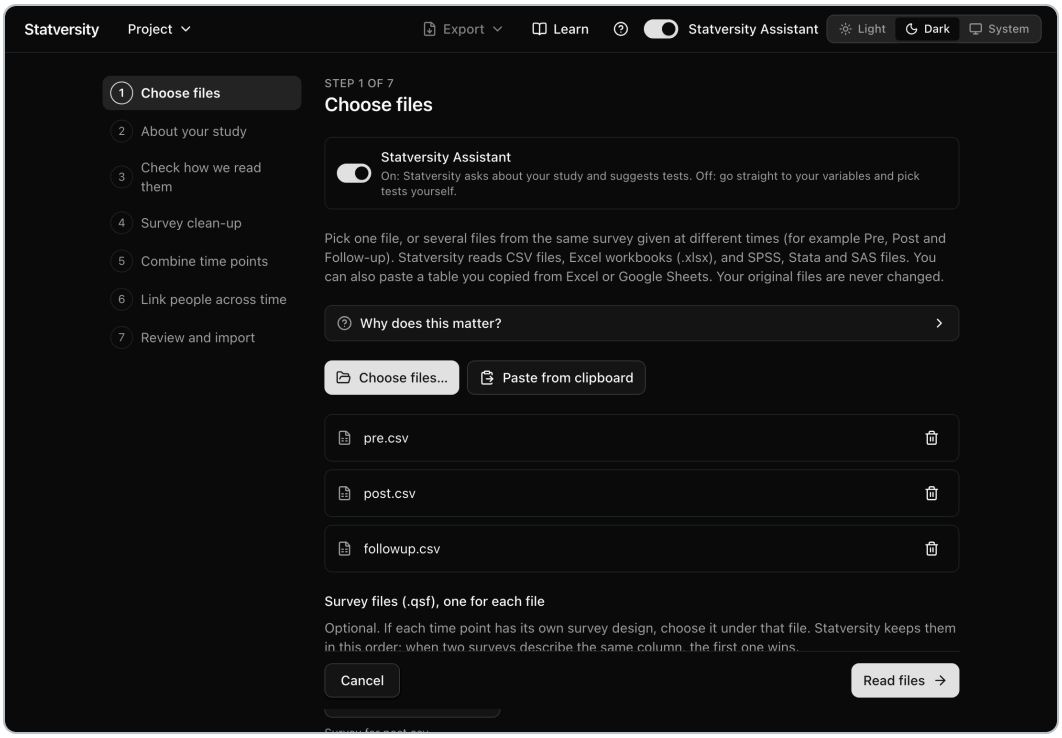


Figure 91. Choose files, with pre.csv, post.csv and followup.csv listed and a Read files button at the bottom.

3. Click **Read files**. The **About your study** screen appears. It asks a couple of quick questions so Statversity can make better first guesses.
4. Click **Skip these questions**. You can always answer them later, and this walkthrough does not need them. The **Check how we read them** screen appears.

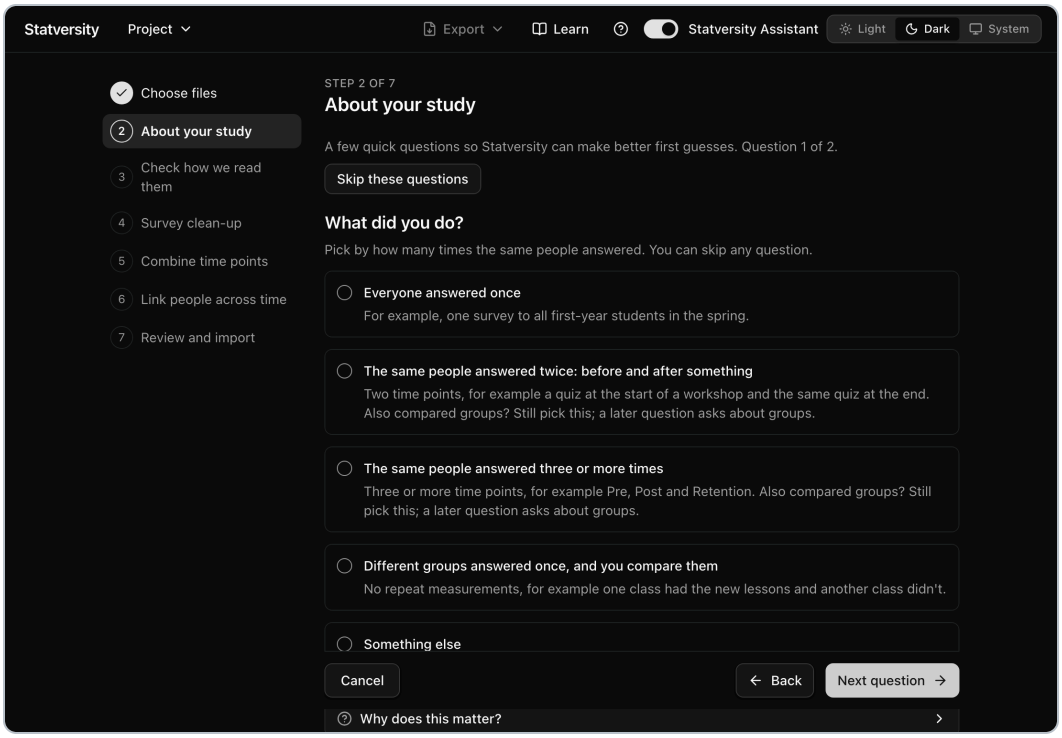


Figure 92. About your study, offering choices such as "The same people answered three or more times". We skipped it.

5. Look at the screenshot below. For each file Statversity reports how it was read: 135 rows and 39 columns for `pre.csv`, a comma-separated layout, and a note that it recognizes a Qualtrics export. The table of first few answers shows the group column (Q2) and the first answers to the test questions, which are letters. Nothing needs changing. Click **Continue**. The **Survey clean-up** screen appears.

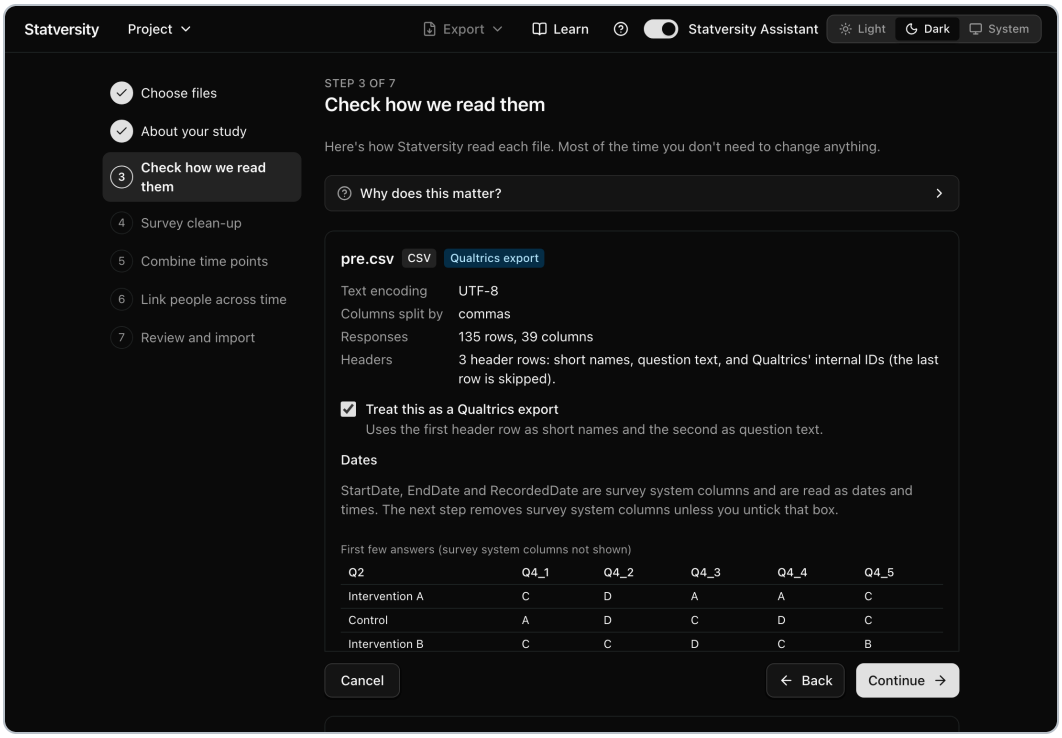


Figure 93. Check how we read them. The first few rows show the group name and a letter answer for each question.

6. On **Survey clean-up**, Statversity has picked safe defaults: hide the survey system columns, and leave out “Survey Preview”, “Survey Test” and “Spam” responses. The labels beside each choice show that zero rows would be removed from each file, so there is nothing to worry about here. Click **Continue**. The **Combine time points** screen appears.

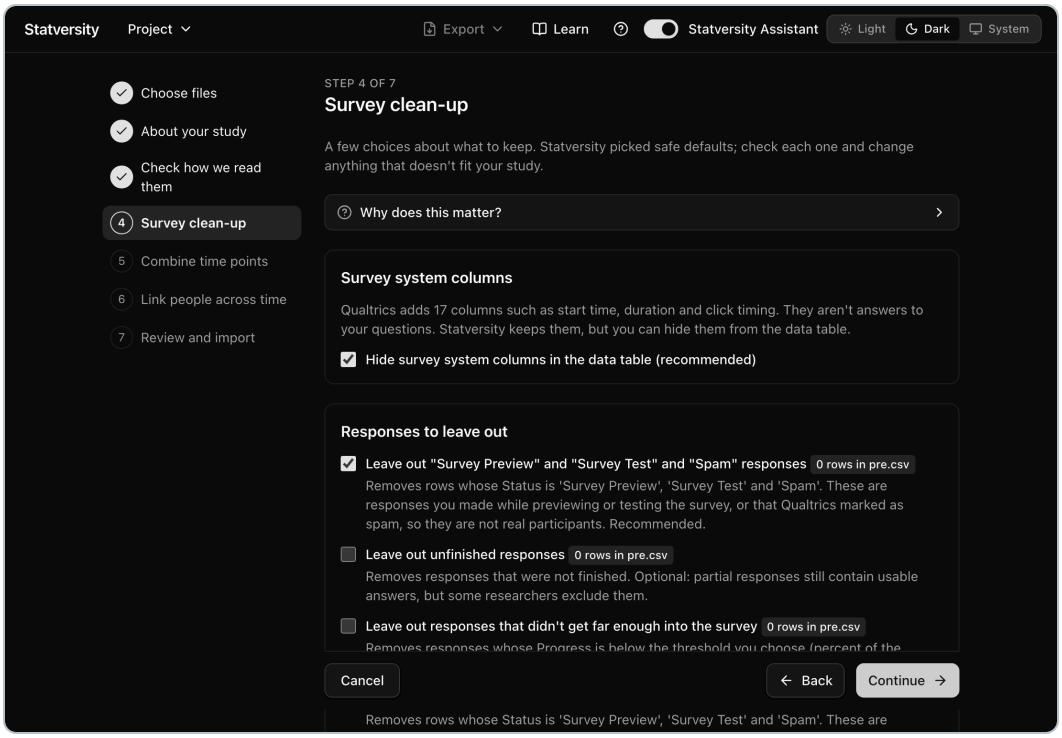


Figure 94. Survey clean-up with the recommended defaults ticked.

7. This is the screen that matters most for a three-file study. Statversity will stack the three files into one table and add a new column called **Time** that says which file each row came from. It has already ordered them Pre, Post, Follow-up from the file names, and reports that all 39 questions were matched by their wording across the files. Click **Continue**. The **Link people across time** screen appears.

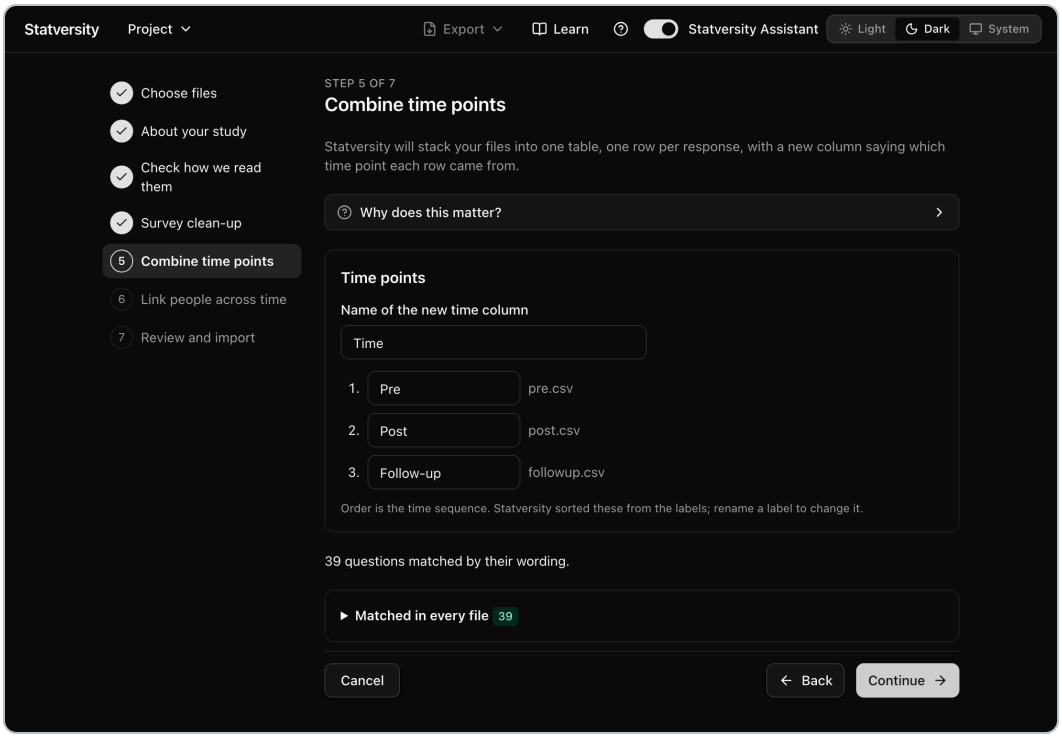


Figure 95. Combine time points, naming the new column Time and putting the files in order Pre, Post, Follow-up.

8. Statversity asks whether it can tell which answers came from the same person at each time point. The default, **Don’t link (default)**, treats each time point as a separate group of people, which is what we want because this dataset has no ID column. The blue note under the options explains the trade-off: without linking, Statversity cannot run paired tests, and you should mention that in your limitations. Click **Continue**. **Review and import** appears.

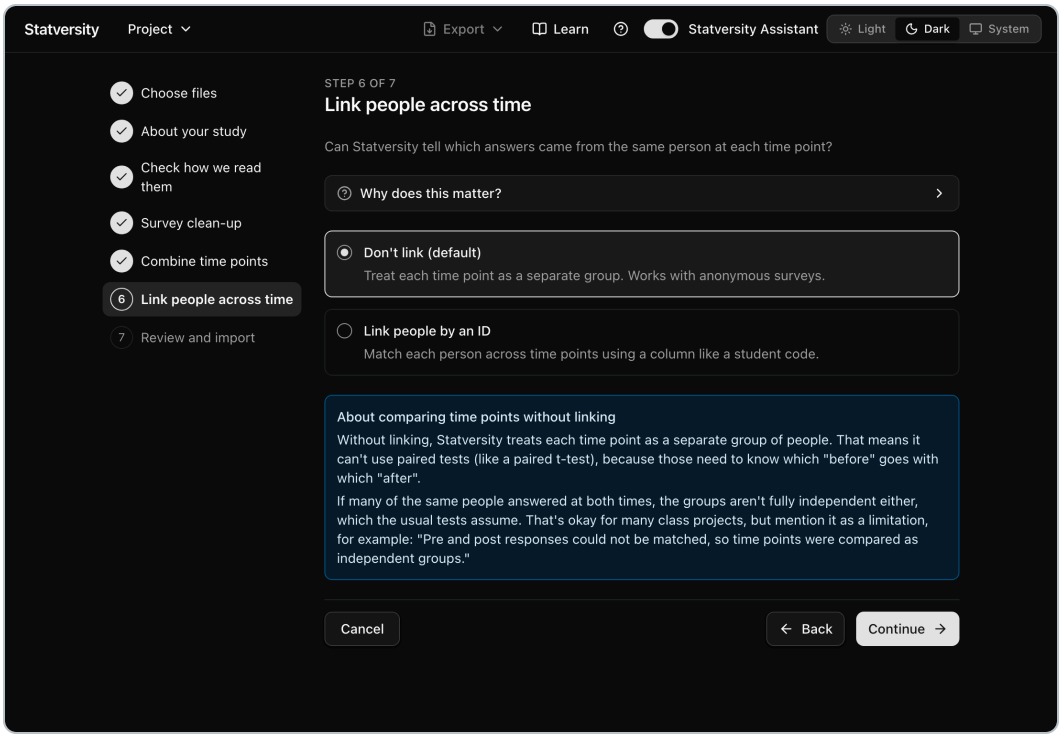


Figure 96. Link people across time, with Don't link (default) selected and an explanatory note.

9. Check the summary. It should read: Files pre.csv, post.csv, followup.csv; Responses read 352 rows; Columns 39 kept; Time points Pre, Post, Follow-up (column “Time”); Linking not linked (separate groups). Click **Import**.

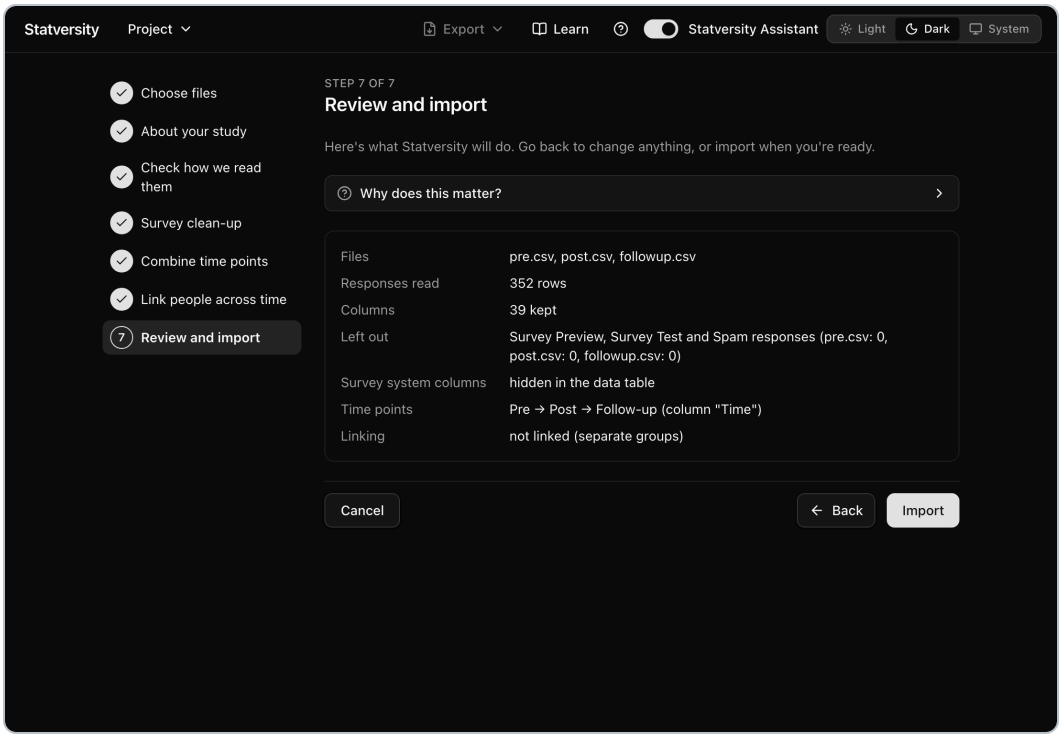


Figure 97. Review and import, listing 352 rows, 39 columns and the time points.

Part 2: The Variable Interview

The Variable Interview is a short series of screens in which Statversity asks what each column is. It has already made a best guess for each, so most screens only need a glance and **Continue**. Chapter 6 explains every control. Here is what each screen means for this study.

10. The **Welcome** screen says there are 3 questions to check. Click **Start**.

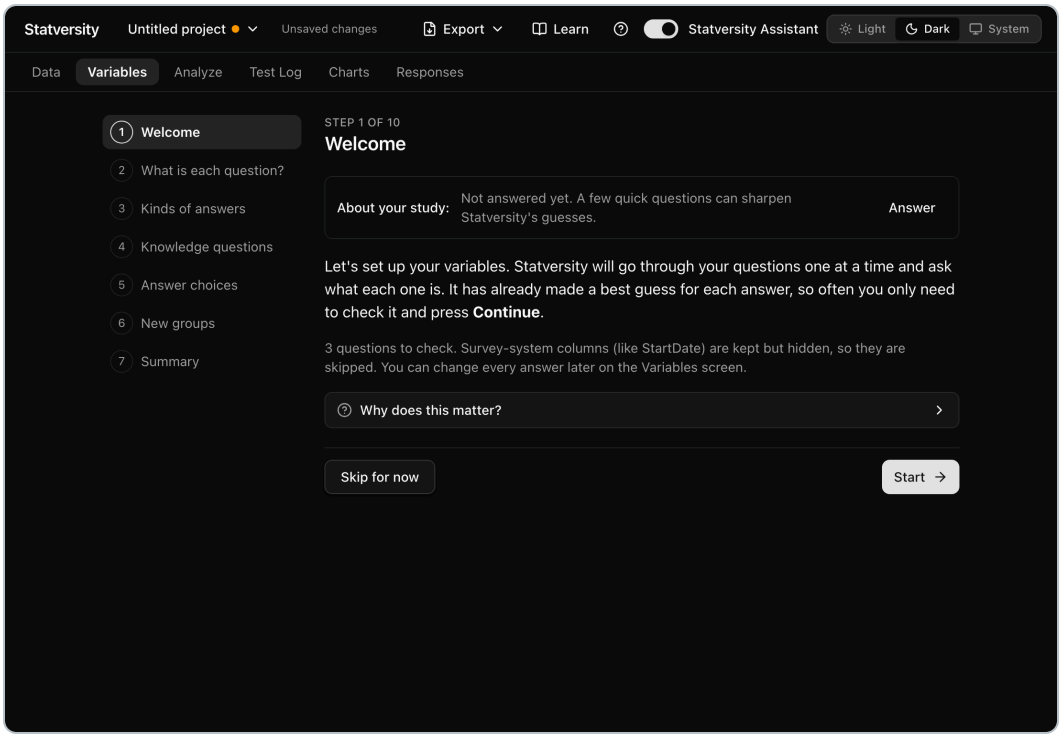


Figure 98. The Welcome step of the Variable Interview.

11. The first **What is each question?** screen shows **Q2**, “Which group are you in?”, with the answers Intervention A, Control and Intervention B. Statversity has chosen **Group or cohort**, meaning “which group someone was put in for the study”. That is right. Click **Continue**.

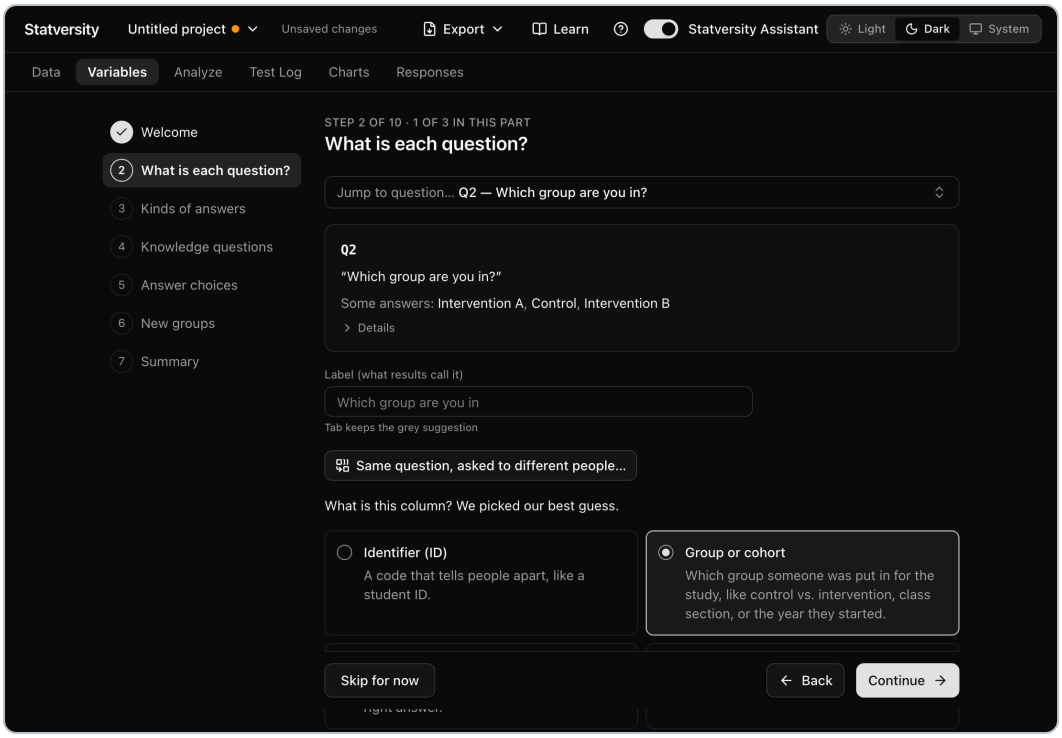


Figure 99. What is each question? for Q2. Group or cohort is selected.

> **IN PLAIN LANGUAGE**
>
> **Why does Statversity ask what a column is?** A spreadsheet only holds letters and numbers. It cannot tell on its own whether a column of numbers is a student ID, a score or a group code. Statversity needs to know, because the right test depends on it. Calling Q2 a group is what lets Statversity offer "compare these three groups" later.

12. The next screen shows **Q4 (20 parts)**, "Knowledge item 1: which statement is correct?". Statversity grouped the 20 test questions together and chose **Test question**, which means "one question on a quiz or test that has a right answer". Correct. Click **Continue**.

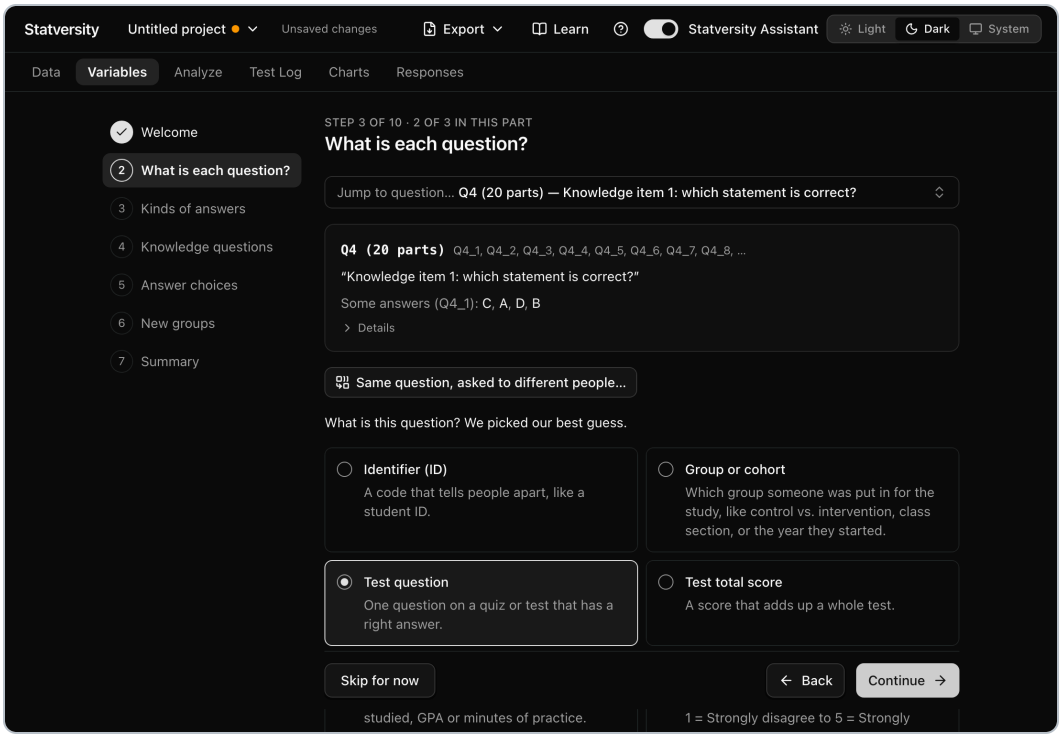


Figure 100. What is each question? for the 20 test questions, labeled Test question.

13. The third screen shows **SC0**, "Total score (auto-scored)", which Qualtrics added and which holds numbers such as 9, 4, 10 and 7. Statversity chose **Test total score**. Click **Continue**. We will not use it; Statversity will make its own total from the answer key.

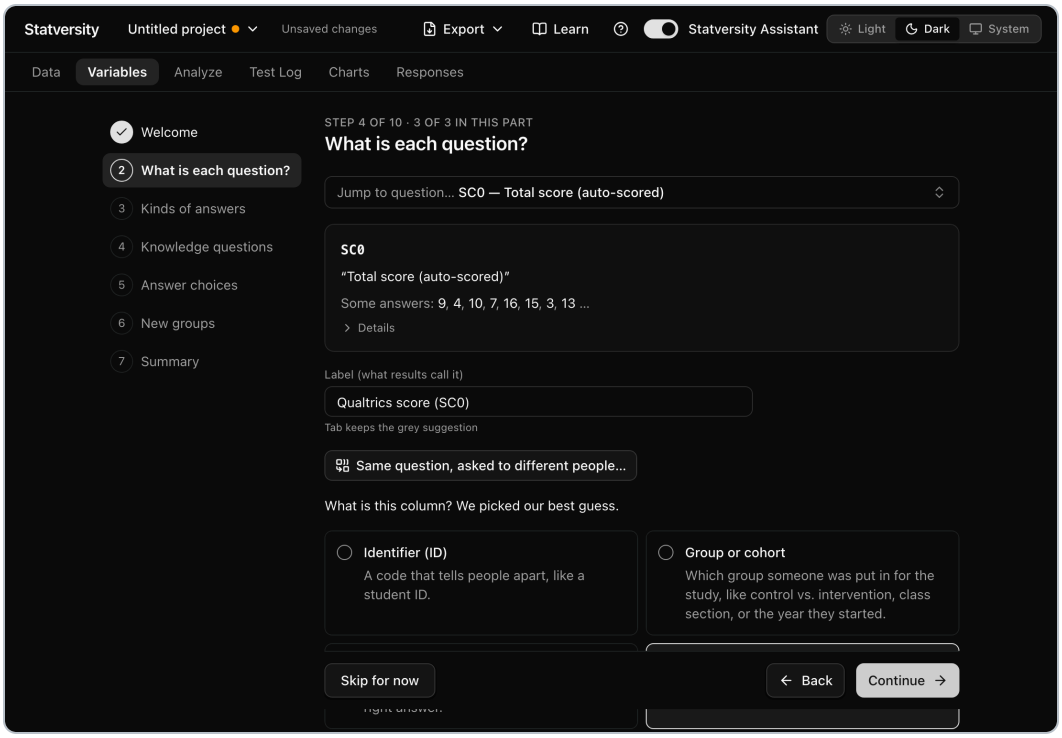


Figure 101. What is each question? for SC0, labeled Test total score.

14. Next is **Kinds of answers**. For **Q2** Statversity chose **Categories (nominal)**: "names with no order, like school or group". Control, Intervention A and Intervention B are names, not amounts, so this is right. Click **Continue**.

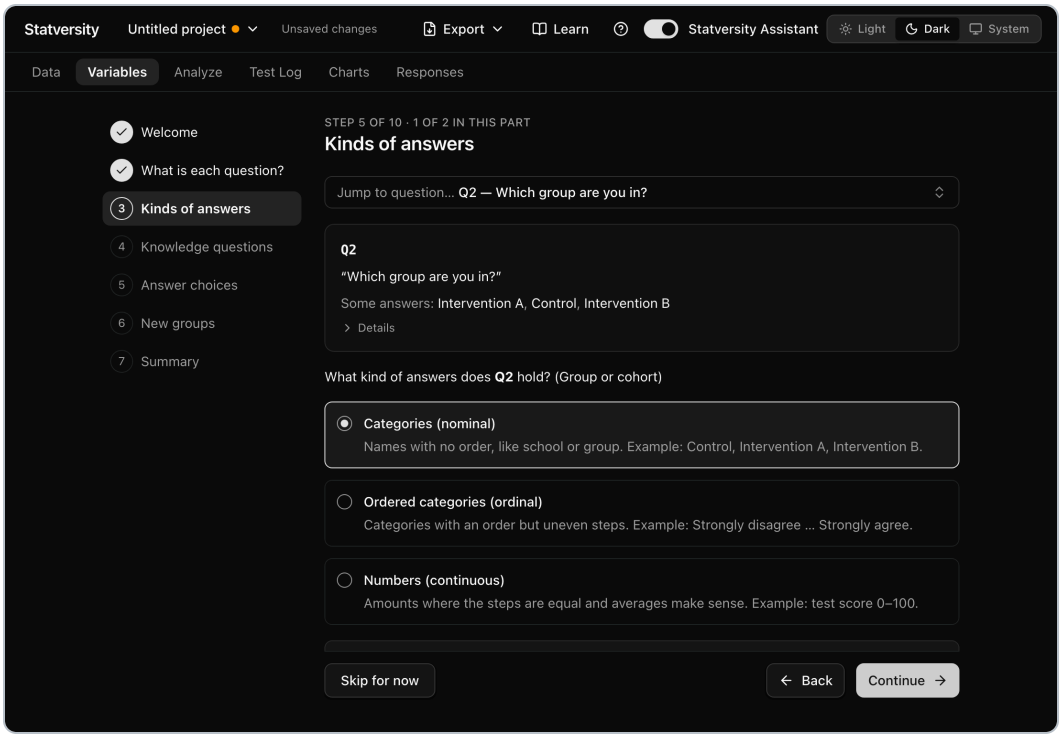


Figure 102. Kinds of answers for Q2. Categories (nominal) is selected.

15. For **SC0** Statversity chose **Numbers (continuous)**: amounts where the steps are equal and averages make sense. Click **Continue**.

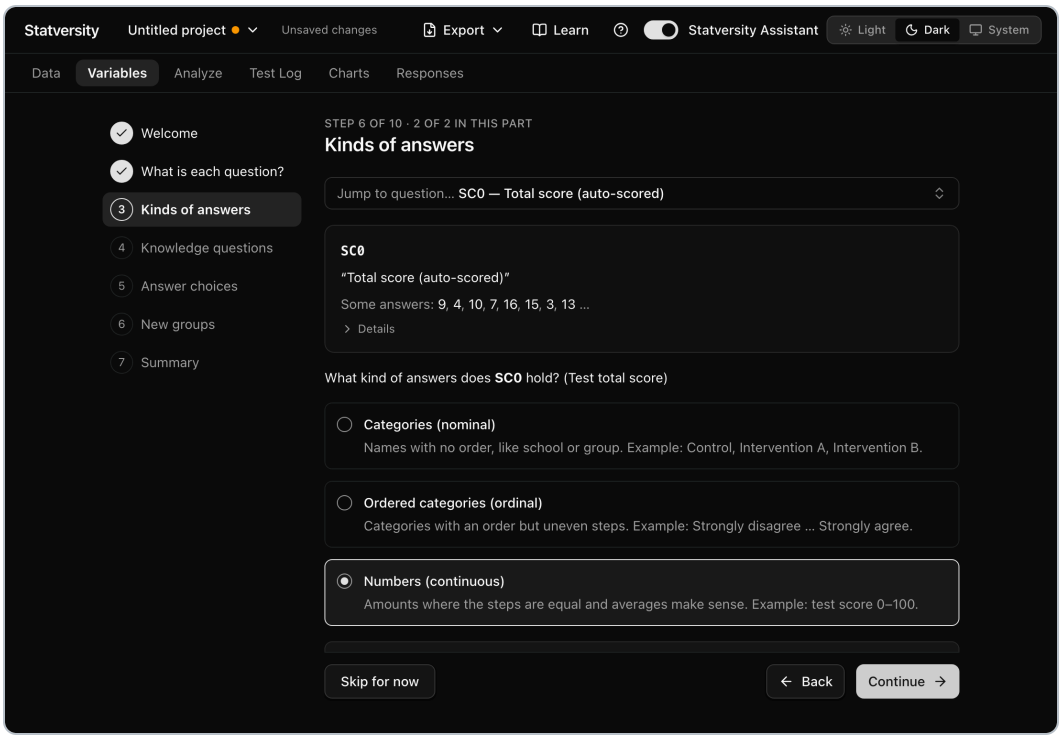


Figure 103. Kinds of answers for SC0. Numbers (continuous) is selected.

16. **Knowledge questions** is the screen that scores the test. Statversity says: “If any of them are knowledge questions with one right answer, tick them and pick the correct answer.” Each of the 20 questions is already ticked **Knowledge question**, with the count of people who chose each letter beside it, but the correct answer is **Not set** for each. Click **Use the practice answer key**. (For your own study you would use **Load answer key…** to choose your file, or **Download answer key template** to get a blank one to fill in.)

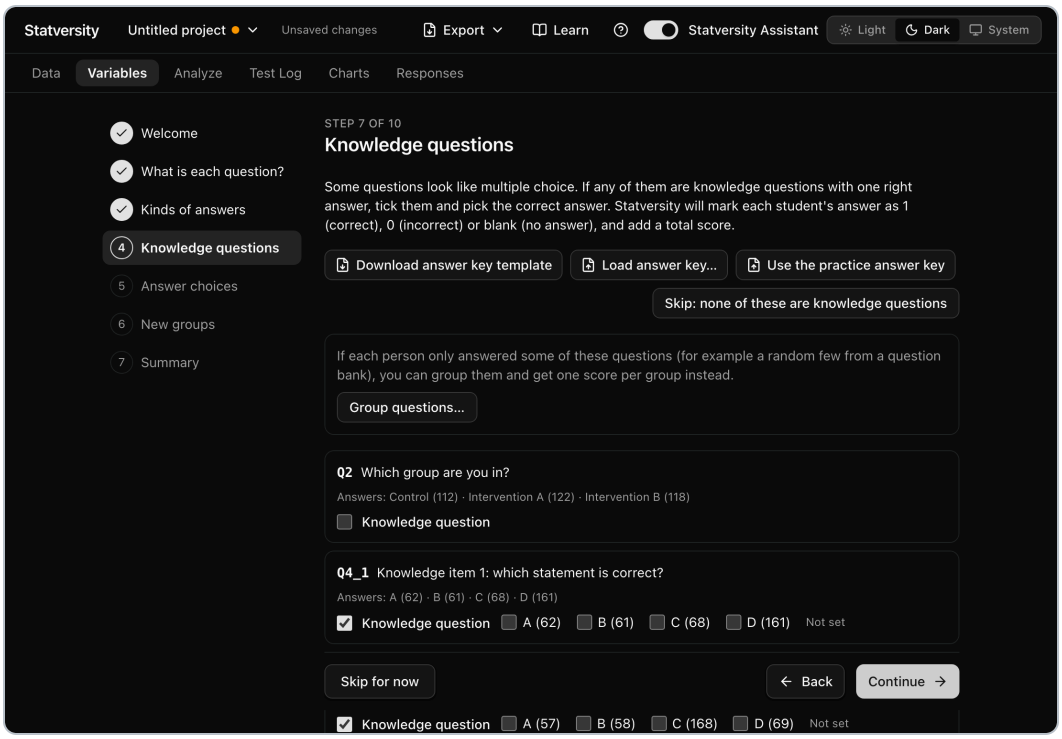


Figure 104. Knowledge questions before the answer key is loaded. Each question says Not set.

17. Statversity fills in the correct answer for each of the 20 questions from the key, and you can still change any of them. Click **Continue**.

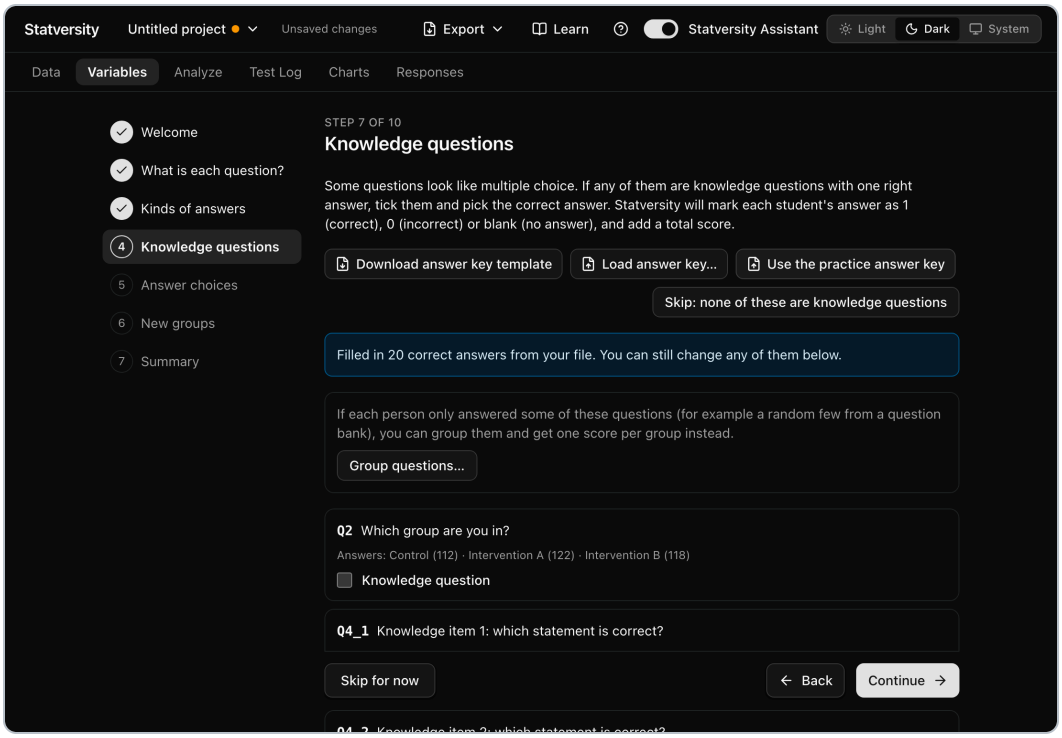


Figure 105. Knowledge questions after loading the practice answer key. The correct letters are now ticked.

> **IN PLAIN LANGUAGE**
>
> **What this step does.** Statversity will create a new 0/1 column for each test question (1 for a correct answer, 0 for an incorrect one, blank for no answer) and add them up into one total per person. This total is the score we compare across groups. Without the answer key you would only have letters, and you cannot average letters.

18. **Answer choices** shows Control, Intervention A and Intervention B, in that order, with the answer in the data beside each. The blue note is general advice for answers stored as numbers; our groups are already words, so there is nothing to change. Click **Continue**.

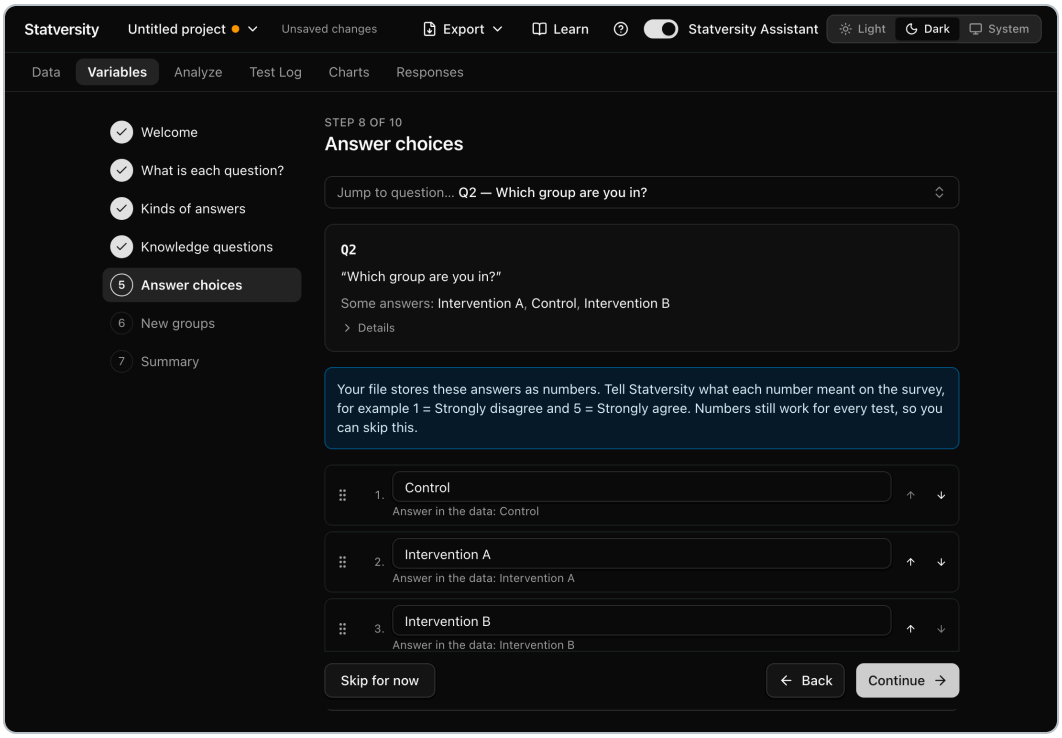


Figure 106. Answer choices for Q2, with the three group names in order.

19. **New groups** offers to build new variables from existing answers. We do not need any, so just click **Continue**.

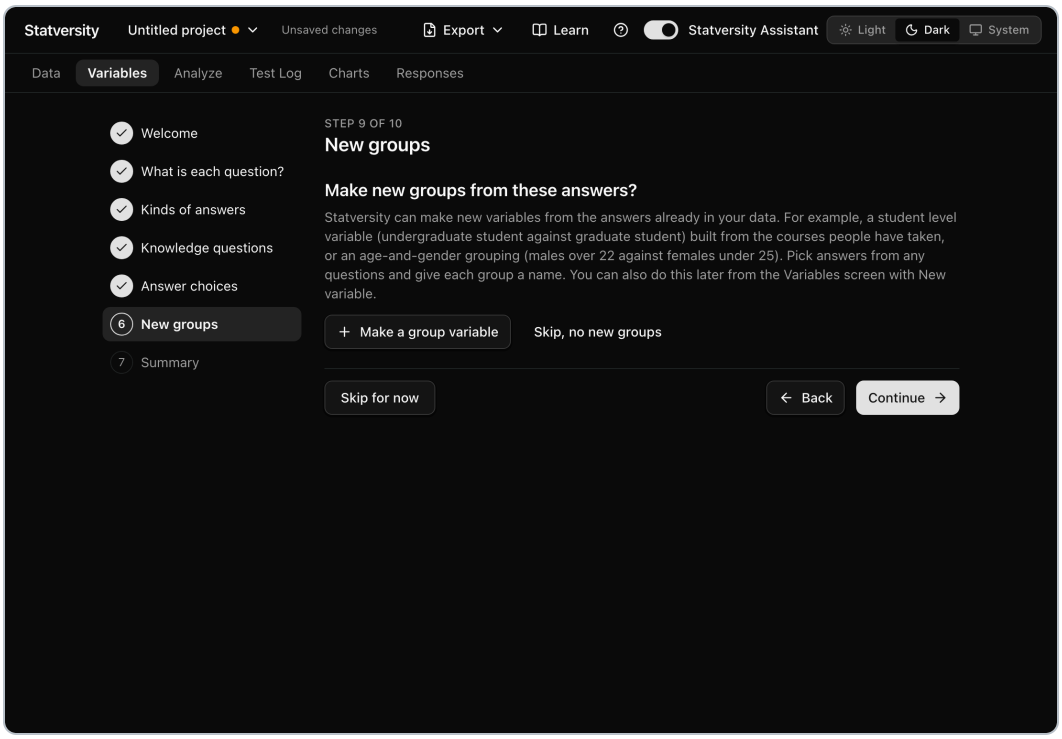


Figure 107. New groups. We did not make any.

20. The **Summary** shows everything you told Statversity. Q2 is a **Group or cohort** and **nominal**. Q4 is a **Test question**. SC0 is a **Test total score** and **continuous**. Under **Test scoring** it says: “20 questions will be scored right (1) / wrong (0), with blank answers left blank, and added up into a total.” At the bottom it says “21 variables to update”. Click **Finish**.

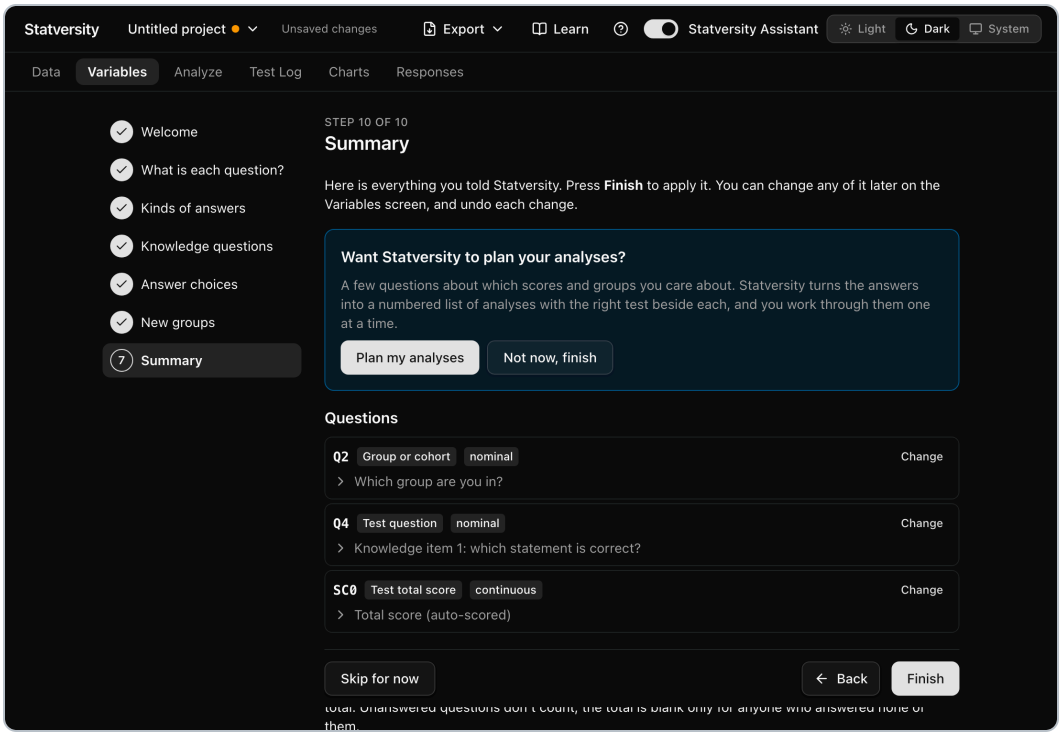


Figure 108. Summary of the interview, including the Test scoring paragraph.

21. You land on the **Variables** screen. Scroll through it if you like. Statversity has scored the 20 test questions with the answer key and added the 0/1 columns and the total that the Advisor offers next as **Q4 total (number correct)**.

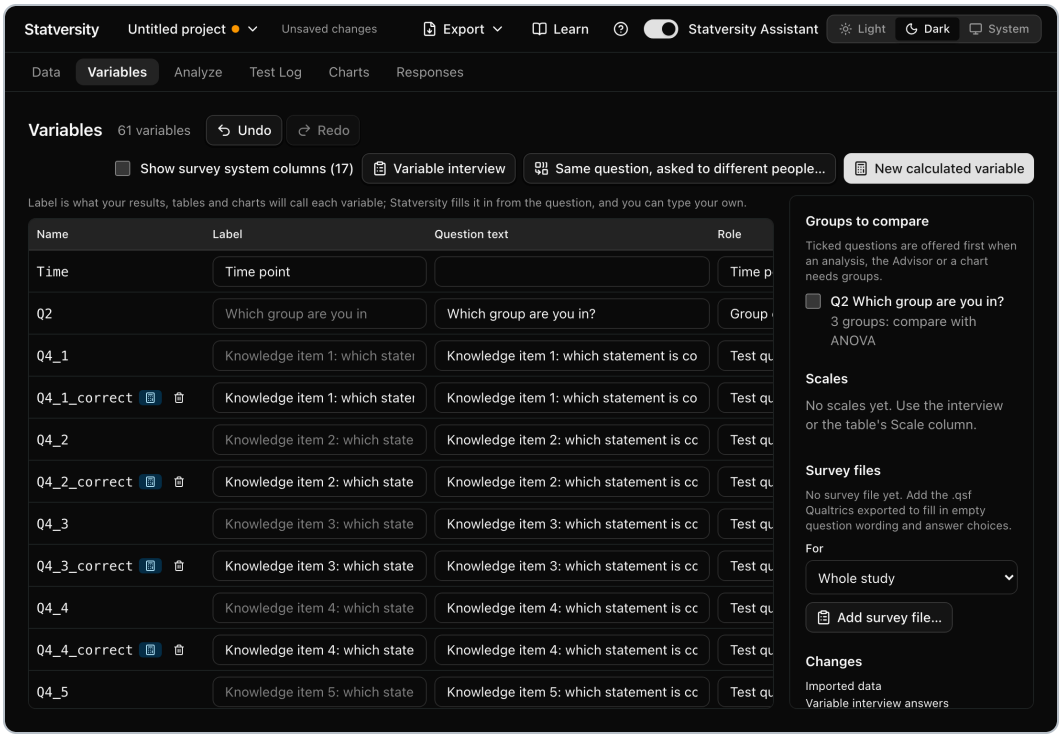


Figure 109. The Variables screen after the interview.

Part 3: Ask the Test Advisor

22. Click **Analyze**. The Test Advisor appears. Read the line under **Which score or answer do you want to analyse?**. It says: “From your data: your outcome is a score, 3 groups, 3 time points, responses not linked across time, 352 responses.” Statversity has already worked out the shape of your data.
23. In **Which score or answer do you want to analyse?**, choose **Q4 total (number correct)**. This is the total of the 20 correct/incorrect columns Statversity just made. The Test Advisor asks **What do you want to know?** (Chapter 7 explains the Advisor in full.)

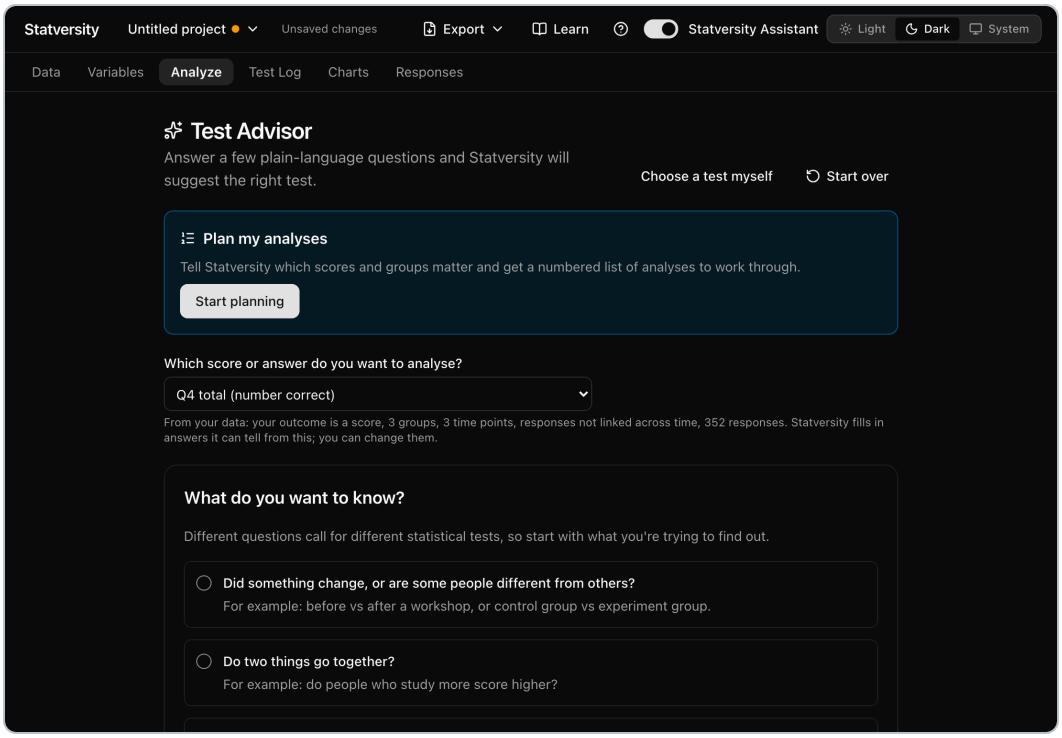


Figure 110. The Test Advisor asking What do you want to know?, with Q4 total (number correct) chosen as the score.

24. Under **What do you want to know?**, choose **Did something change, or are some people different from others?**, then click **Continue**.

> **IN PLAIN LANGUAGE**
>
> **What this question is really asking.** The four choices are four families of questions. “Did something change, or are some people different from others?” is the family for comparing: before versus after, or one group versus another. We want to know whether the three groups differ, so this is the one. The others are for relationships (“do two things go together?”), prediction (“can some things help me estimate something else?”) and surveys (“do my rating questions measure one idea?”).

25. The Advisor now shows what you have said so far, and asks **What are you comparing?** Choose **Different groups of people.**, then click **Continue**.

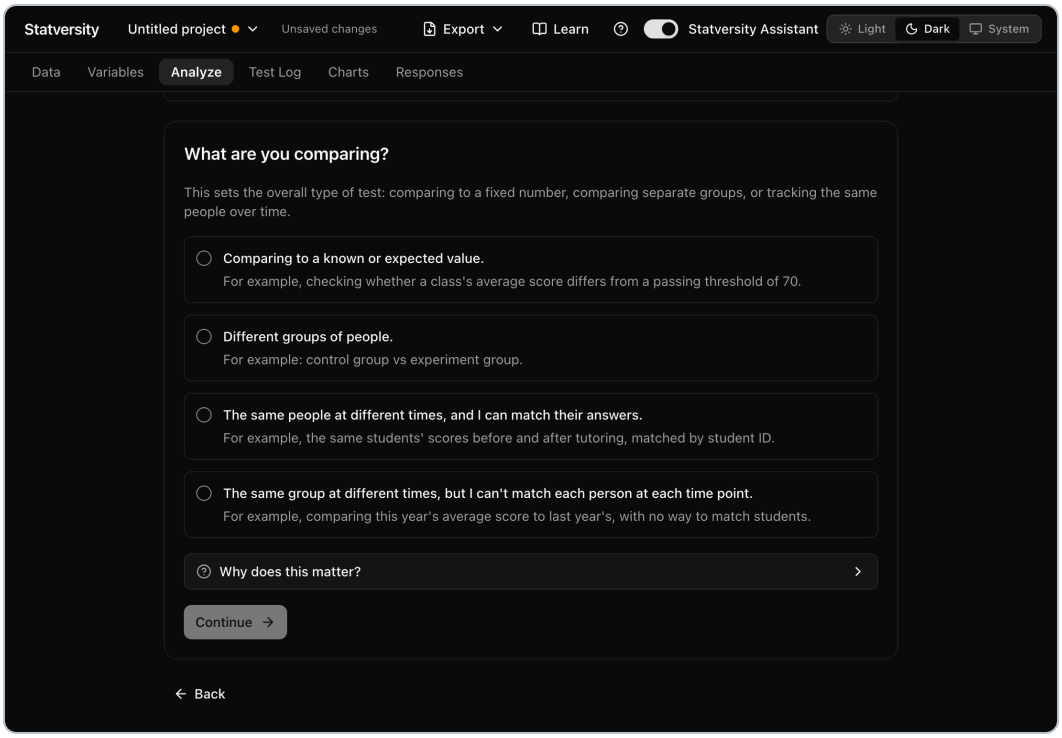


Figure 111. What are you comparing? Different groups of people is the right answer for three separate groups.

> **IN PLAIN LANGUAGE**
>
> **Why not "the same people at different times"?** Because that choice needs you to match each person's answers at one time to their answers at another. Our data are not linked, so we cannot. Instead we compare the groups at one time point: the Post test.

26. Under **How many ways are people split into groups?**, choose **One way, e.g. by teaching method.**, then click **Continue**.

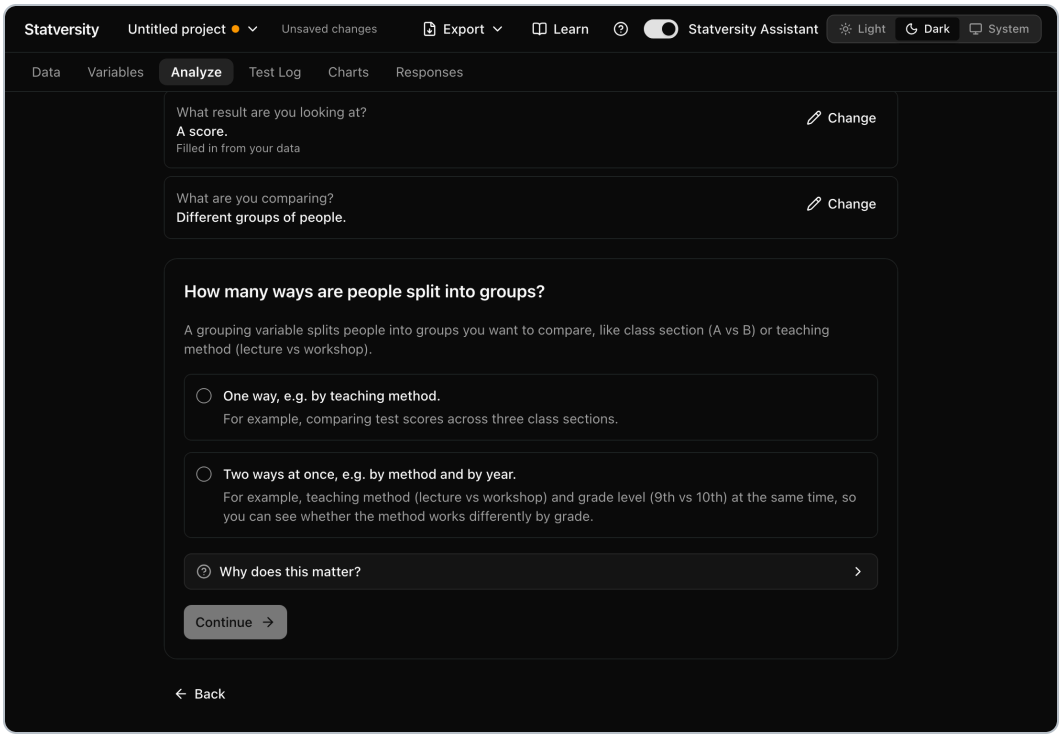


Figure 112. How many ways are people split into groups? One way is chosen.

> **IN PLAIN LANGUAGE**
>
> **One way or two ways?** We are splitting people by one thing only: which group they were in. If we had split them by group and by, say, year of study at the same time, that would be two ways. Notice that Statversity filled in the next question itself: "How many groups? Three or more groups. Filled in from your data."

27. Under **Do you want to control for another variable (a covariate), such as a pretest score?**, choose **No**, then click **Continue**.

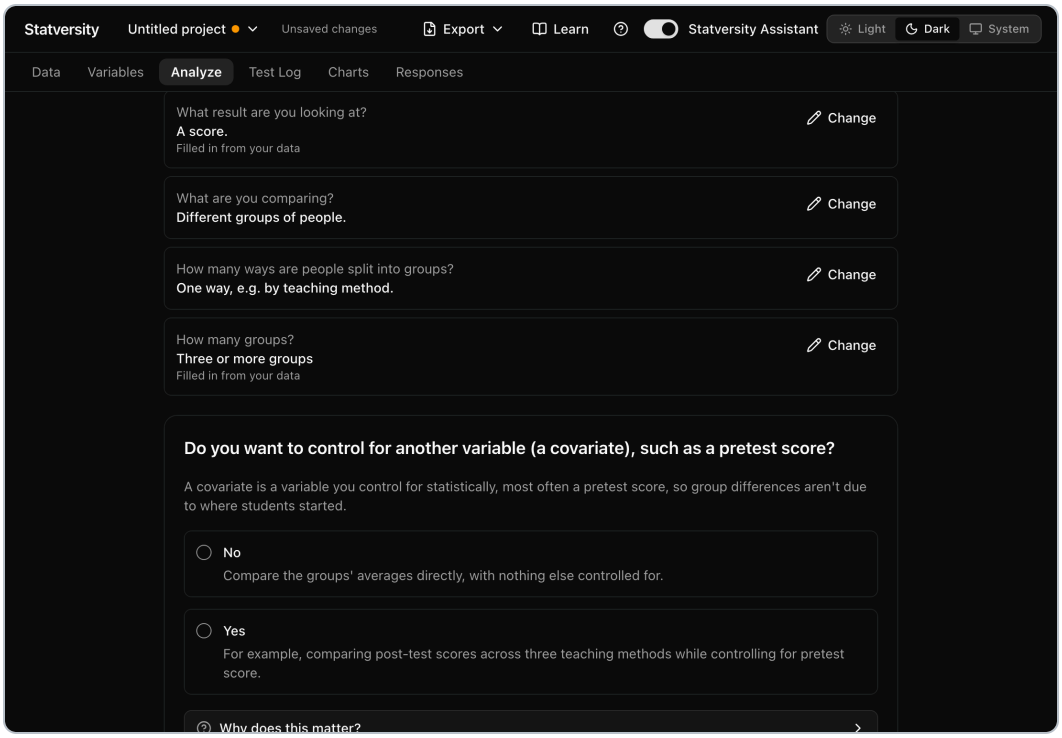


Figure 113. The covariate question. We chose No.

> **IN PLAIN LANGUAGE**
>
> **What is a covariate?** It is another measurement you adjust for, usually a pretest, so that group differences are not just differences in where people started. Our pretest is in a different file from the posttest with no link between people, so we cannot adjust for it. The answer is No, which compares the groups' averages directly.

28. Statversity shows its recommendation: **Statversity recommends: One-way ANOVA**. Under it, it lists a backup test (the Kruskal-Wallis test), the effect sizes it will report (eta squared, omega squared and Cohen's f), and the assumptions it will check: **Normality (bell-shaped scores)** and **Equal spread across groups**. Remember that backup test, because you will meet it again in a moment.

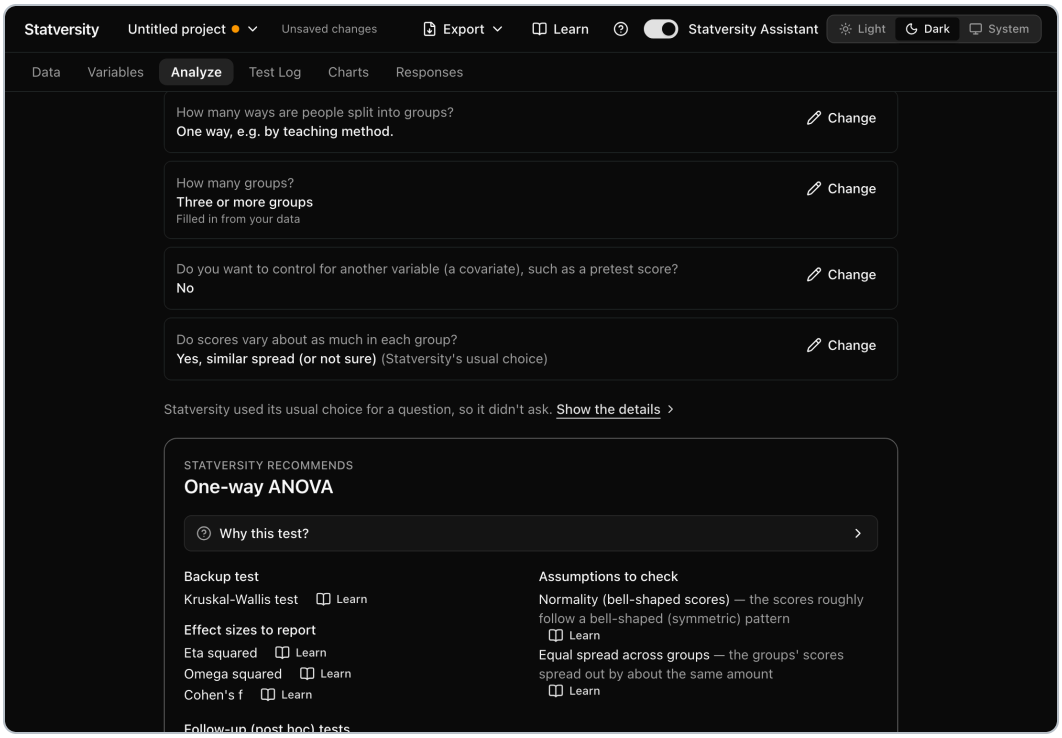


Figure 114. Statversity recommends One-way ANOVA, and names Kruskal-Wallis as the backup test.

Part 4: Choose your variables

29. Click **Set up this analysis**. **Choose your variables** opens.

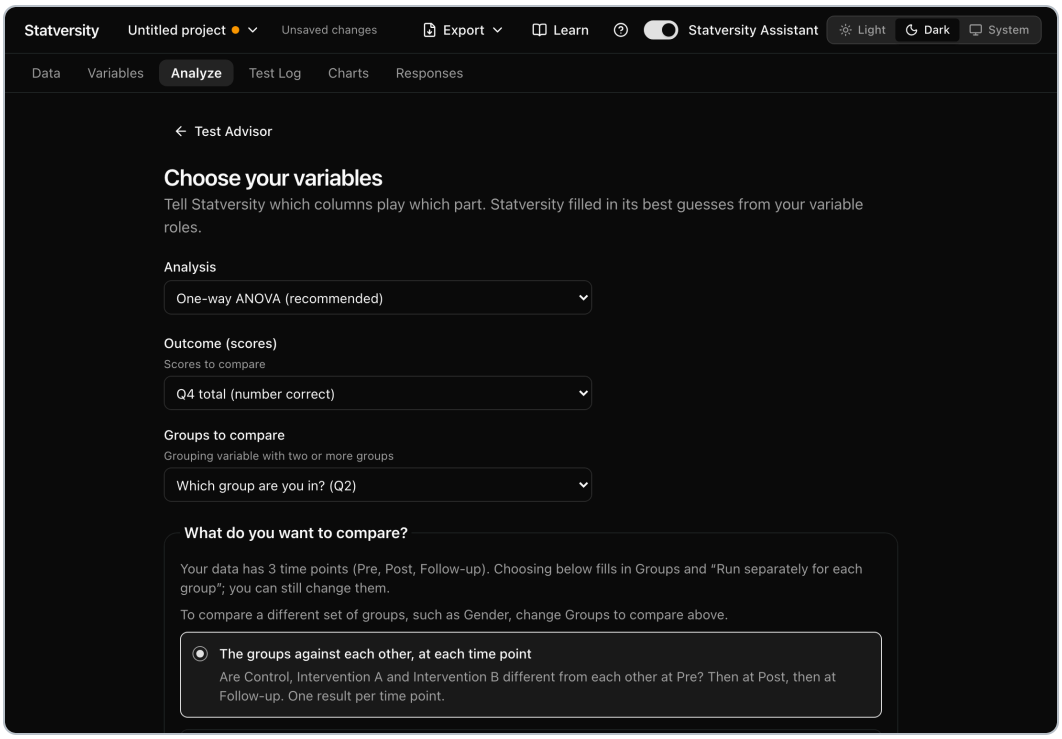


Figure 115. Choose your variables. The analysis, the outcome and the groups are already filled in.

30. Check what Statversity filled in. **Analysis** is **One-way ANOVA (recommended)**. **Outcome (scores)** is **Q4 total (number correct)**. **Groups to compare** is **Which group are you in? (Q2)**. No change is needed to any of these.
31. Scroll down to **Which responses to use**. Your data has three time points, but our question is about the Post test, so we will tell Statversity to use only the Post responses. In **Use only some responses**, choose **Time point**.
32. Under **Keep the responses where Time point is:**, tick **Post (119)**. The page now says "Your filter keeps Post only, so the groups are compared at Post." and "119 responses of 352 will be used." Leave **Run separately for each group** on **Don't split (one result)**.

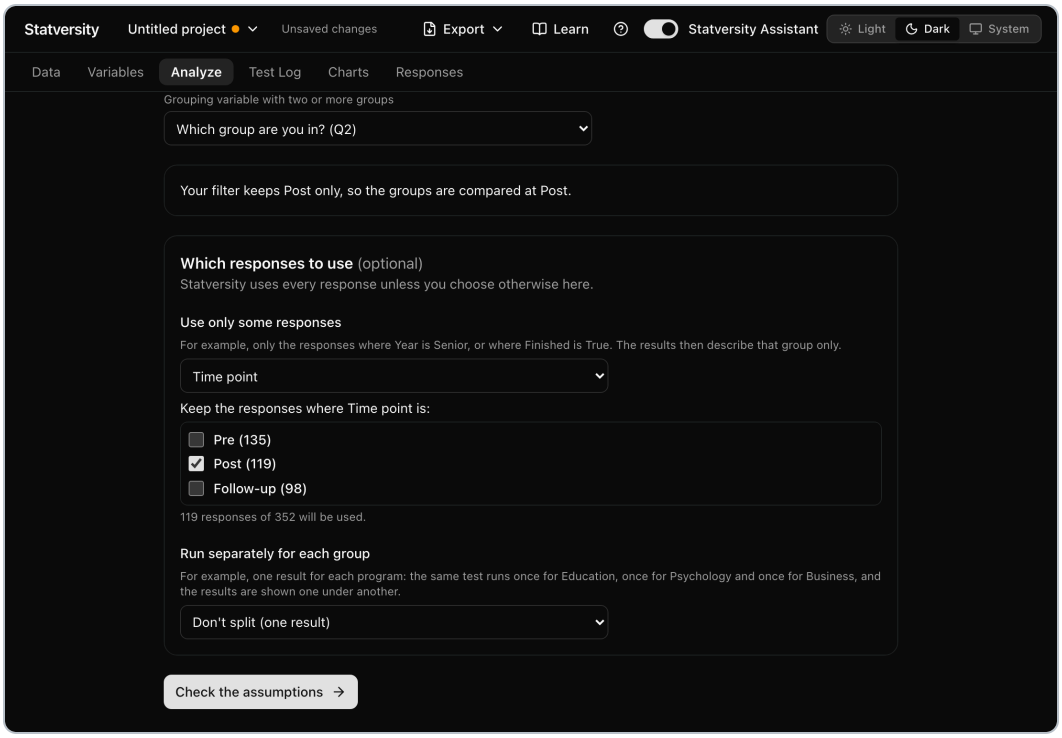


Figure 116. The variables are set, with only Post (119) ticked.

> **CAUTION**
>
> **Why filter to Post?** If you left all 352 responses in, Statversity would mix the Pre, Post and Follow-up responses into the same three groups, and the answer would blur together "who started better" with "who did better after the intervention". Comparing groups at one time point keeps the question clean.

Part 5: Check the assumptions

An **assumption** is a condition a test needs in order to give trustworthy answers. Statversity checks each one for you, shows a plain-English verdict, and shows the plots behind it. Chapter 23 explains the idea in more detail.

33. Click **Check the assumptions**. The first of four checks appears: **Normality: Control**. Normality means that scores follow a roughly bell-shaped pattern, with most values in the middle and fewer at the extremes. Statversity used the **Shapiro-Wilk test** on the 38 Control scores and reports p = .197. The verdict says **Looks fine**. Below it are two pictures: a Q-Q plot and a histogram.

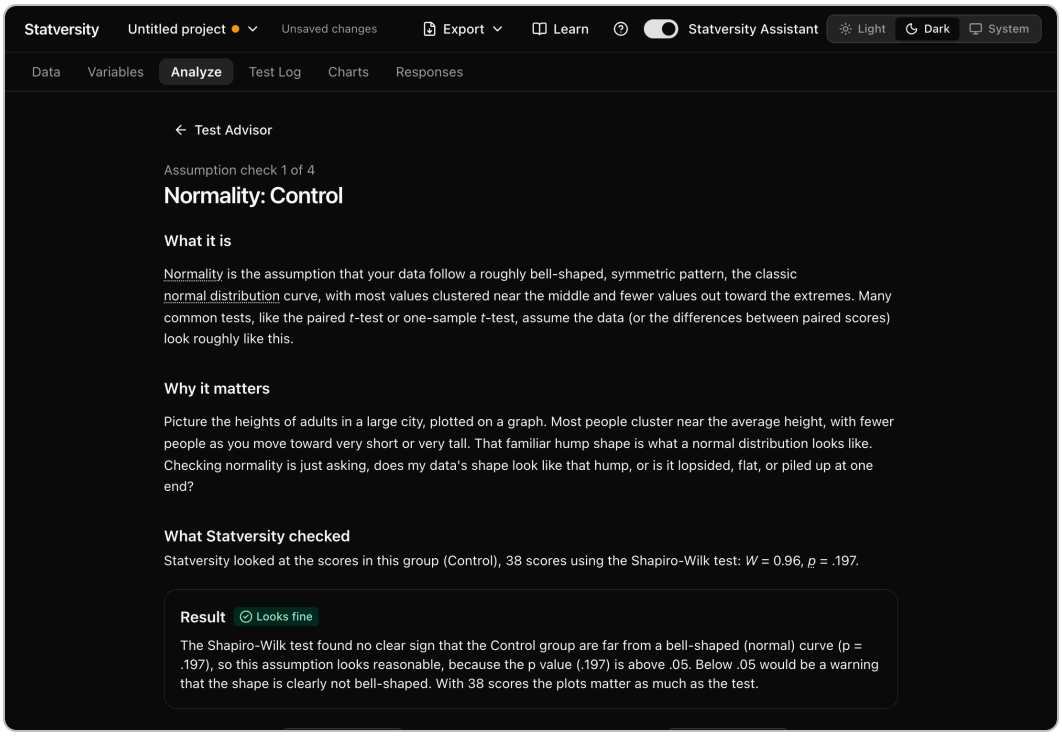


Figure 117. Normality: Control. The verdict is Looks fine (p = .197).

IN PLAIN LANGUAGE

How to read the verdict. Here a *large* p-value is good news. The test starts by assuming the data are bell-shaped, and the p-value says how surprising your data would be if that were true. p = .197 is not at all surprising, so there is no reason to doubt the bell shape. If p dropped below .05, Statversity would flag a warning.

Reading the plots. In the Q-Q plot, the dots should hug the dashed diagonal line if the data are bell-shaped. In the histogram, the bars should look roughly like a hump.

34. Click **Next check**. **Normality: Intervention A** appears: 41 scores, p = .474, **Looks fine**.

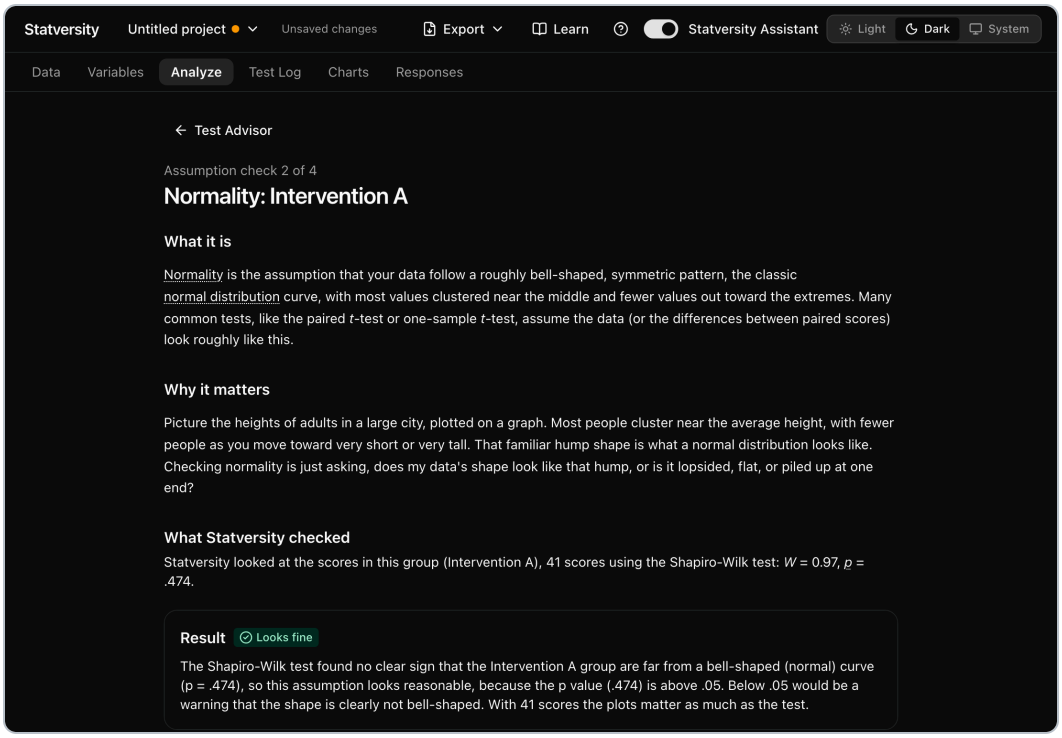


Figure 118. Normality: Intervention A. The verdict is Looks fine (p = .474).

35. Click **Next check**. **Normality: Intervention B** appears, and this time the badge is orange and says **Concern**. Statversity checked the 40 Intervention B scores and found W = 0.94, p = .034, which is below .05. The text says the Shapiro-Wilk test suggests the group are not bell-shaped, that with a sample this size that can affect the results, and that you should check the plots and consider the rank-based (nonparametric) alternative.

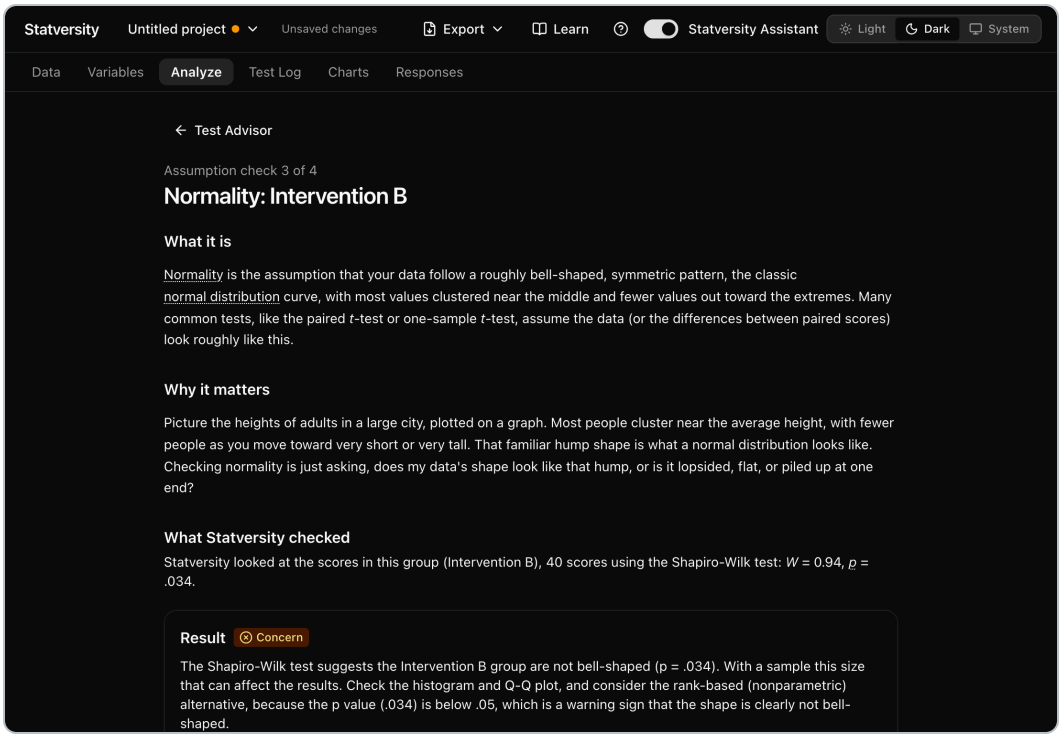


Figure 119. Normality: Intervention B. The badge says Concern (p = .034).

> **IN PLAIN LANGUAGE**
>
> **This is Statversity protecting you.** A one-way ANOVA works well when scores are roughly bell-shaped. If one group is lopsided or has a few extreme scores, the ANOVA's answer can be less trustworthy. Statversity noticed this for you. A researcher working by hand would have to remember to check, and many forget. Statversity flags it, shows you the evidence, and offers a safer test.

36. Click **Next check**. The fourth check, **Equal spread (homogeneity of variance): all groups**, looks at whether the three groups' scores are scattered by similar amounts. Statversity used Levene's test (Brown-Forsythe version) on all 119 scores: $F(2, 116) = 0.92$, $p = .401$, **Looks fine**. The page explains it with an archery picture: three archers might all hit the middle on average, but one might be much more scattered than the others.

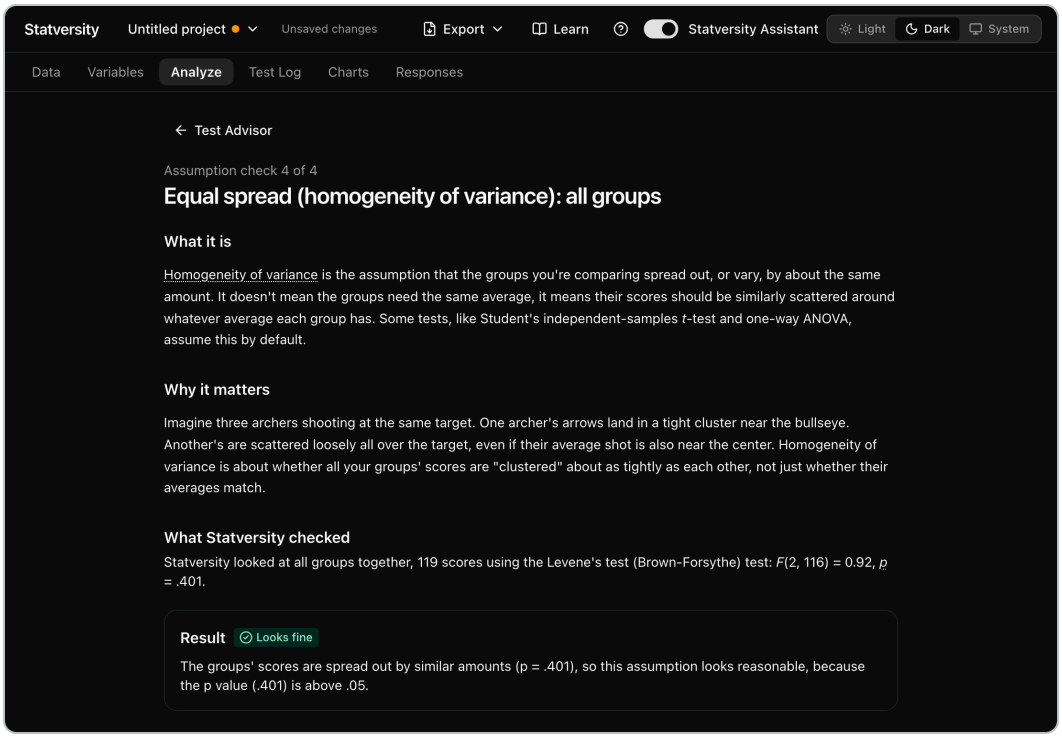


Figure 120. Equal spread across groups. The verdict is Looks fine (p = .401).

Part 6: Decide which test to use

37. Click **Decide which test to use**. The screen **Which test should you use?** collects all four checks in one list: three **Looks fine** and one **Concern**.

Below the list a blue box says: “Statversity suggests: Kruskal-Wallis test. At least one normality check failed. Look at the plots: if they show clear skew or outliers, the rank-based (nonparametric) test is the safer choice.” Under that are two choices. **Use the recommended test: One-way ANOVA** is not selected. **Use the nonparametric alternative: Kruskal-Wallis test** is selected and tagged **Suggested**.

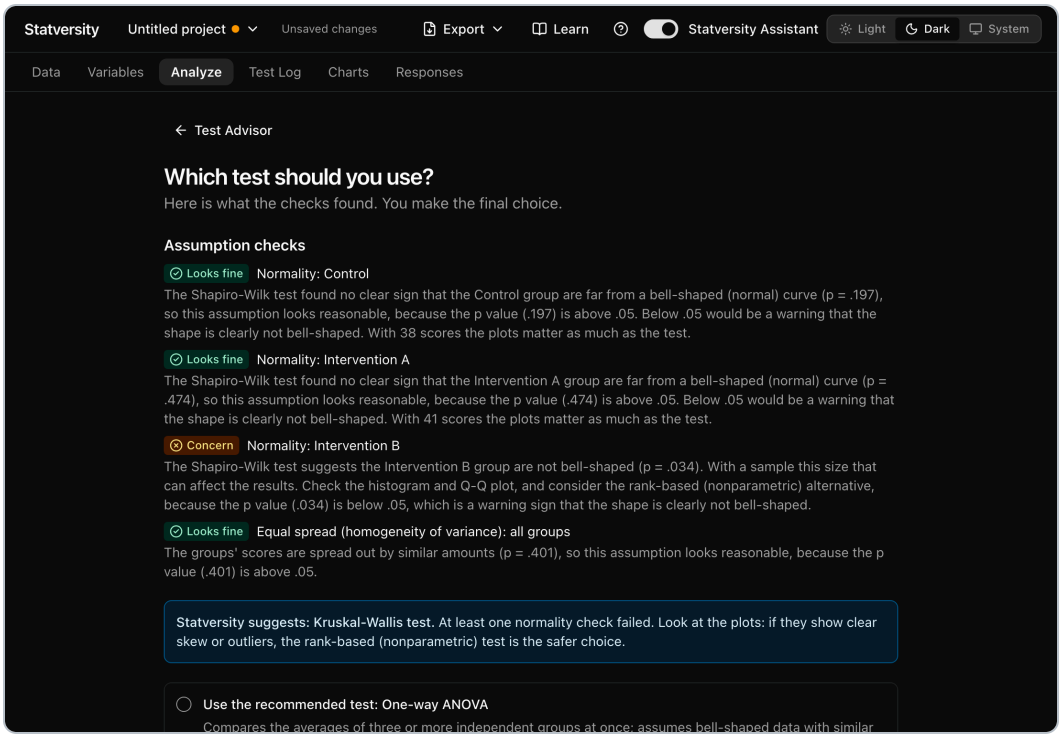


Figure 121. Which test should you use? Kruskal-Wallis test is suggested because of the Intervention B concern.

You make the final choice. We accept Statversity's suggestion.

> **IN PLAIN LANGUAGE**
>
> **What is the Kruskal-Wallis test, in plain words?** Imagine lining up all 119 people from the lowest score to the highest, and giving each a rank: 1st, 2nd, 3rd, and so on. Then split the line back into the three groups and ask: do the ranks in one group tend to be higher than in the others? If the groups really did not differ, each group's ranks would be a random mix of low and high, and each group's *average rank* would be about the same. The Kruskal-Wallis test measures how far apart the groups' average ranks are and asks how surprising that would be by chance alone.
>
> Because it only uses the order of the scores and not their exact sizes, a few extreme scores or a lopsided shape have far less pull on it than on an average. That is why it is called a **nonparametric** or **rank-based** test: it does not need the bell shape. The price is that it is a little less powerful than ANOVA when the bell shape *is* present, which is why Statversity offers it when a check raises a concern instead of using it every time.

38. Make sure **Use the nonparametric alternative: Kruskal-Wallis test** is selected, then click **Run and see the results**.

What Statversity found

The Results screen opens with the title **Kruskal-Wallis test**, then “Time point: Post only, n = 119” and “119 responses analysed (Control: 38; Intervention A: 41; Intervention B: 40).”

39. Read the **What this means** box at the top. Statversity writes: “Scores tended to be highest in the Intervention B group and lowest in the Control group. Differences this large are unlikely to be due to chance alone (p < .001); Dunn’s test in the ‘Which groups differ?’ section shows which pairs of groups differ. The size of the effect was large by common benchmarks.” Beneath it are the two headline numbers: Kruskal-Wallis H(2) = 22.10, p < .001 and an effect size, rank epsilon squared, ε² = .19, 95% CI [.09, .35].

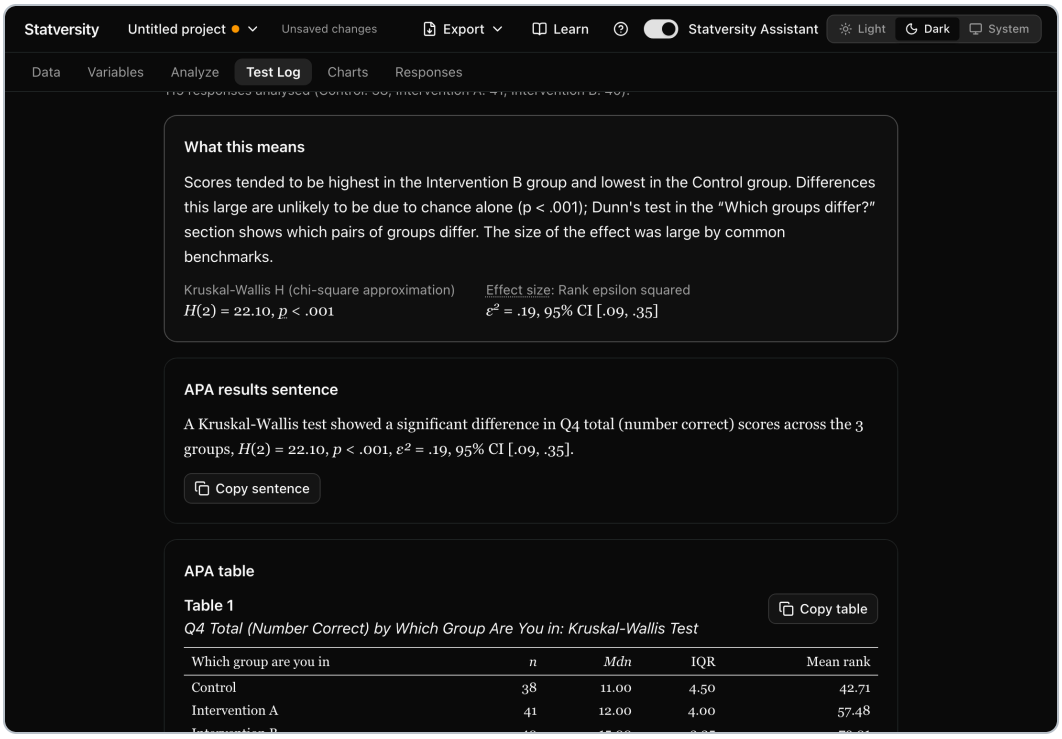


Figure 122. The top of the results: the What this means box, the test statistic and the effect size.

40. Scroll down to the **APA results sentence** and **APA table**. The sentence is ready to copy, and the table gives n, median, interquartile range and mean rank for each group.

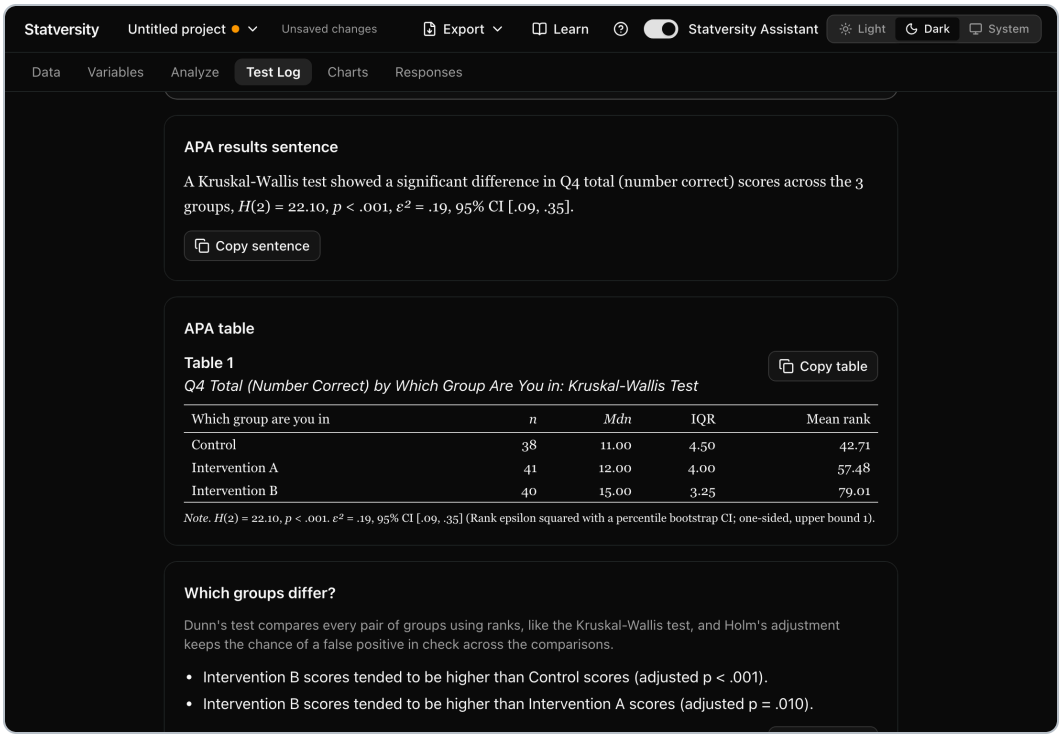


Figure 123. The APA results sentence and Table 1.

41. Scroll to **Which groups differ?**. This is the follow-up, or *post hoc*, section, which Statversity added because the main test was significant.

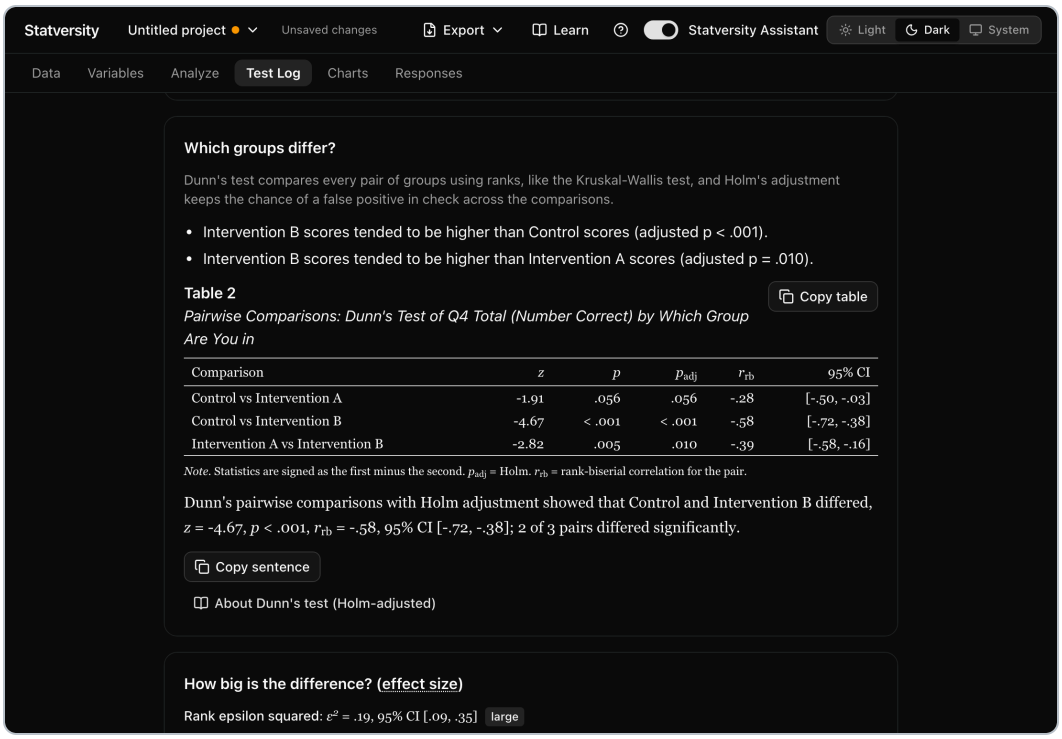


Figure 124. Which groups differ? with Dunn's test and Holm adjustment.

42. Scroll to **Descriptive statistics**. (This result has no chart.)

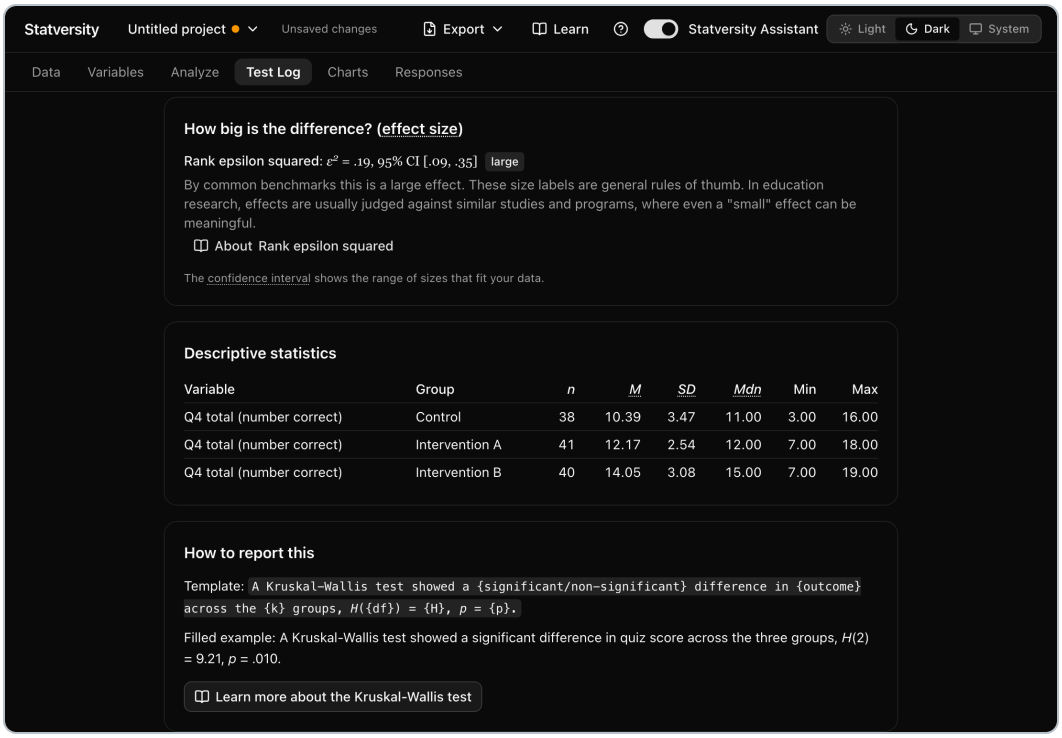


Figure 125. Descriptive statistics for each group.

43. Click **Test Log**. The analysis has been saved there with its inputs and results, so you can come back to it.

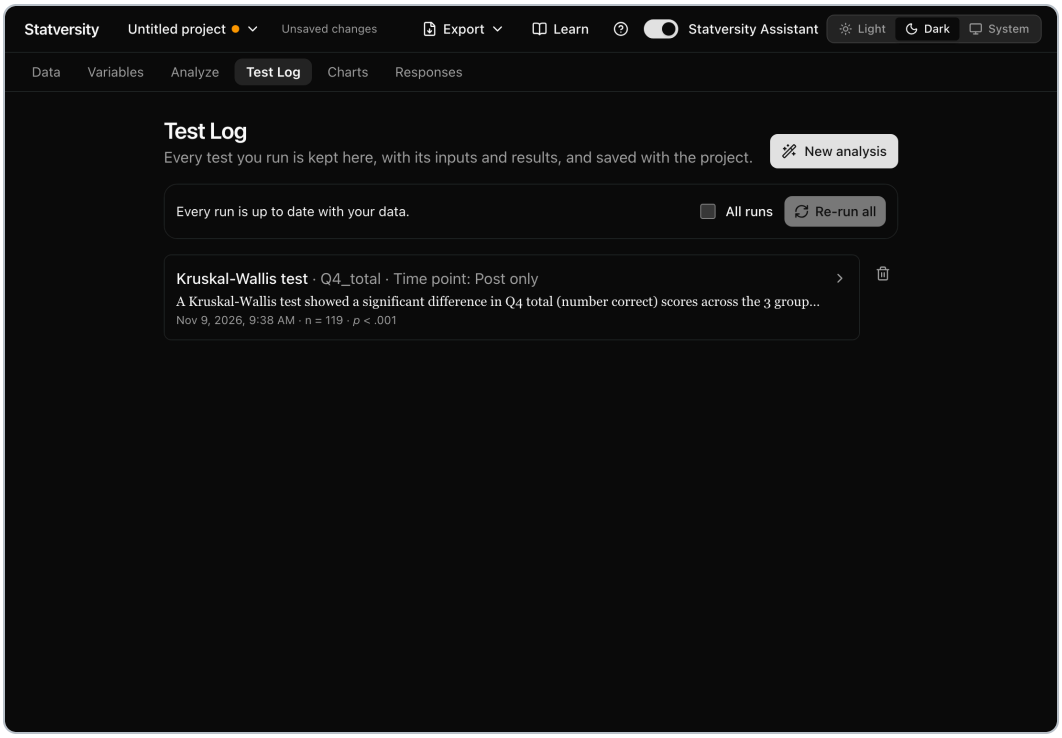


Figure 126. The Test Log, with the Kruskal-Wallis result saved.

The numbers

What	Value
Test	Kruskal-Wallis test (Post only)
Statistic	H(2) = 22.10
p	< .001
Effect size	Rank epsilon squared ε^2 = .19, 95% CI [.09, .35], labelled large
Control	n = 38, M = 10.39, SD = 3.47, Mdn = 11.00, mean rank 42.71
Intervention A	n = 41, M = 12.17, SD = 2.54, Mdn = 12.00, mean rank 57.48
Intervention B	n = 40, M = 14.05, SD = 3.08, Mdn = 15.00, mean rank 79.01

The scores are the number of the 20 test questions answered correctly.

The follow-up comparisons, using Dunn’s test with Holm’s adjustment, are:

Pair	z	p	Adjusted p	Rank-biserial r
Control vs Intervention A	-1.91	.056	.056	-.28
Control vs Intervention B	-4.67	< .001	< .001	-.58
Intervention A vs Intervention B	-2.82	.005	.010	-.39

The Results page also draws a **Comparison figure**. Because this is a rank test, the bars are group medians with the interquartile range, and Dunn's test supplies the brackets.

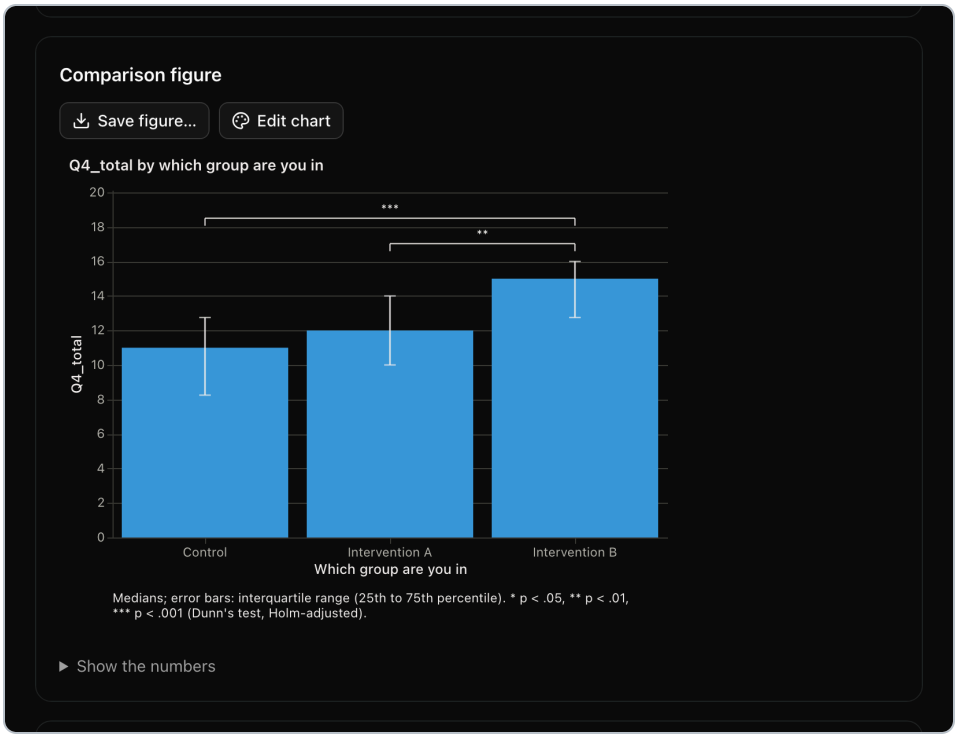


Figure 127. The Comparison figure for the Kruskal-Wallis result: medians, with brackets for Control against Intervention B and Intervention A against Intervention B.

Only the pairs that differ significantly get a bracket. Control against Intervention B has three stars ($p < .001$) and Intervention A against Intervention B has two ($p < .01$), both using Holm-adjusted p values. Control against Intervention A (adjusted $p = .056$) has no bracket because it did not reach .05. Chapter 8 explains how to read and edit this figure.

What Dunn's test and Holm's adjustment do

The Kruskal-Wallis test answers one question: "Is there a difference somewhere among these three groups?" It does not say *where*. With three groups there are three possible pairs (Control and A, Control and B, A and B), and **Dunn's test** compares each pair using the same ranks as before.

There is a catch. Each comparison carries a small chance of a false alarm. Run three of them and the chances add up, so the odds of at least one false alarm are higher than 5%. **Holm's adjustment** corrects for this by making the p-values a bit stricter. That is the "adjusted p" column. Always read the adjusted column, not the plain one.

> **IN PLAIN LANGUAGE**
>
> **An analogy.** If you flip a coin once, getting heads is unremarkable. If you flip it ten times, you would be surprised if heads *never* came up. The more comparisons you make, the more likely it is that one looks impressive purely by luck. Holm's adjustment is Statversity asking for extra proof when you have looked at more than one pair.

The rank-biserial correlation in the last column (r_{rb}) is an effect size for each pair. It runs from -1 to 1, and the sign only tells you the direction: here a negative value means the second group in the pair tended to score higher. Bigger numbers without the sign (.58 is bigger than .28) mean a bigger difference between the pair.

What this means in plain English

The three groups did not score the same. Intervention B did best, with an average of about 14 questions right out of 20. Intervention A came next at about 12, and Control was lowest at about 10. The Kruskal-Wallis test says that a gap this large among the three groups would be very unlikely if the interventions made no difference at all ($p < .001$).

Dunn's follow-up tests show where the differences are. Intervention B scored higher than Control, and the evidence is strong (adjusted $p < .001$). Intervention B also scored higher than Intervention A (adjusted $p = .010$). Intervention A and Control are the one pair Statversity could not separate: the adjusted p-value of .056 is just above the .05 line. That does not prove A did nothing. It means that with these numbers we cannot rule out luck for that one pair.

The effect size says how big the overall difference is. $\varepsilon^2 = .19$ is labelled **large** by common benchmarks, and the confidence interval of .09 to .35 means that, given the data, the true size could plausibly be anywhere from a medium to a very large effect. As the app itself reminds you, these labels are general rules of thumb, and in education research a result is usually judged against similar programs.

> **CAUTION**
>
> **Remember what this study can and cannot say.** The groups here were compared at one time point, and the three time points contained different numbers of people who could not be matched. Statversity cannot tell you whether the groups were already different before the interventions began. If the groups were not assigned at random, some of the gap could reflect who ended up in each group, not what each intervention did.

What you would write in a paper

Statversity gives you the sentence ready to paste:

> A Kruskal-Wallis test showed a significant difference in Q4 total (number correct) scores across the 3 groups, $H(2) = 22.10$, $p < .001$, $\varepsilon^2 = .19$, 95% CI [.09, .35].

To adapt it, replace "Q4 total (number correct)" with the name your readers will recognize, such as "post-test knowledge score", and add one more sentence for the follow-up from the same screen: "Dunn's pairwise comparisons with Holm adjustment showed that Control and Intervention B differed, $z = -4.67$, $p < .001$, $r_{rb} = -.58$, 95% CI [-.72, -.38]; 2 of 3 pairs differed significantly." In your methods section, say that you chose the Kruskal-Wallis test because the normality check for one group raised a concern, and say which groups you compared at which time point.

What if

What if the normality check had passed for all three groups? Statversity would have kept the one-way ANOVA recommendation. You would have run it as the recommended test, and Statversity would have reported eta squared, omega squared and Cohen's f as the effect sizes (the three listed on the recommendation screen), with a follow-up test suited to ANOVA.

What if the spread check had raised a concern instead? The "Equal spread" check protects a different weakness of ANOVA. When groups are scattered by different amounts, Statversity would again flag it, and the **Which test should you use?** screen would again suggest the rank-based Kruskal-Wallis test and explain why. (The one-way ANOVA result also reports Welch's F, a version that does not assume equal spread, and the Advisor recommends Welch's ANOVA instead if you open **Show the details** under its recommendation and change the skipped spread question to "No, quite different spread across groups".) The point is the same: read the verdict, then make the final choice yourself.

What if the effect size had been small? Suppose ε^2 were .02 and p were still below .05. The groups would differ reliably, but the difference would be tiny, and with a small effect the confidence interval would sit near zero. In that case you would report the p-value but would not describe the interventions as making a big difference. Reading both numbers is what chapter 23 is about.

What if the same people had been matched across Pre, Post and Follow-up? On the **Link people across time** screen you would choose **Link people by an ID** and pick the column that holds the student code. Statversity could then treat the data as repeated measurements on the same people, which is a different and usually more powerful kind of analysis. The blue note on that screen mentions that the paired tests need this link.

What if you want to adjust for the pretest? On the Advisor question about a covariate you would choose **Yes**, which the screen describes as “comparing post-test scores across three teaching methods while controlling for pretest score”. That needs the pretest score to be in the same row as the post-test score, which means linking people by an ID.

What if you looked at Follow-up instead of Post? Return to **Choose your variables** and in **Use only some responses** tick **Follow-up (98)** instead of **Post (119)**. Statversity would run the same checks on that smaller set of people. Because fewer people remain, small differences are harder to detect.

> **TRY IT YOURSELF**
>
> **Try it yourself.** Repeat the analysis on the **Follow-up** time point. Open **Choose your variables**, tick **Follow-up (98)** instead of **Post (119)**, and work through the assumption checks again. Before you look at the results, predict from the means or medians which group will be highest. Then answer: did the same group win? Did the normality check for Intervention B raise a concern again? Does the pair that was not significant at Post (Control and Intervention A) stay that way?

CHAPTER 13

Cleaning up a real export

The question

Does a six-question classroom survey hang together as one scale, and can you get a raw survey export into shape first, without losing track of what you removed or why?

This is a **scale reliability** study. Students rated six statements about their class, such as “I enjoy coming to this class” and “I often feel lost in this class”. You want to add those six answers into one “classroom experience” score per student. Before you may do that, you need evidence that the six statements really measure the same thing. The tool for that is **Cronbach’s alpha**, a single number between 0 and 1 that says how consistently a set of questions moves together. Higher means the questions agree with each other more.

But the interesting part of this chapter comes before the analysis. Real survey exports are messy. They carry extra columns you did not ask for, columns that identify people, test responses you made while building the survey, and special codes for “no answer”. This chapter walks through every choice on Statversity’s **Survey clean-up** screen in plain language, and then runs the reliability check on the cleaned data.

The data

The practice dataset is called **Cleaning up a real export**. It is one survey, a “Course Experience Survey” with 113 responses, exported from Qualtrics (a popular survey tool) in five different messy ways. You import only one of the five files. We use the first and most standard one, a CSV file with three header rows. Statversity also supplies the survey’s design file (a `.qsf` file), which is how it knows the wording of each question.

To open it:

1. On the Home screen, click **Open practice data**.
2. Scroll to the card titled **Cleaning up a real export**.

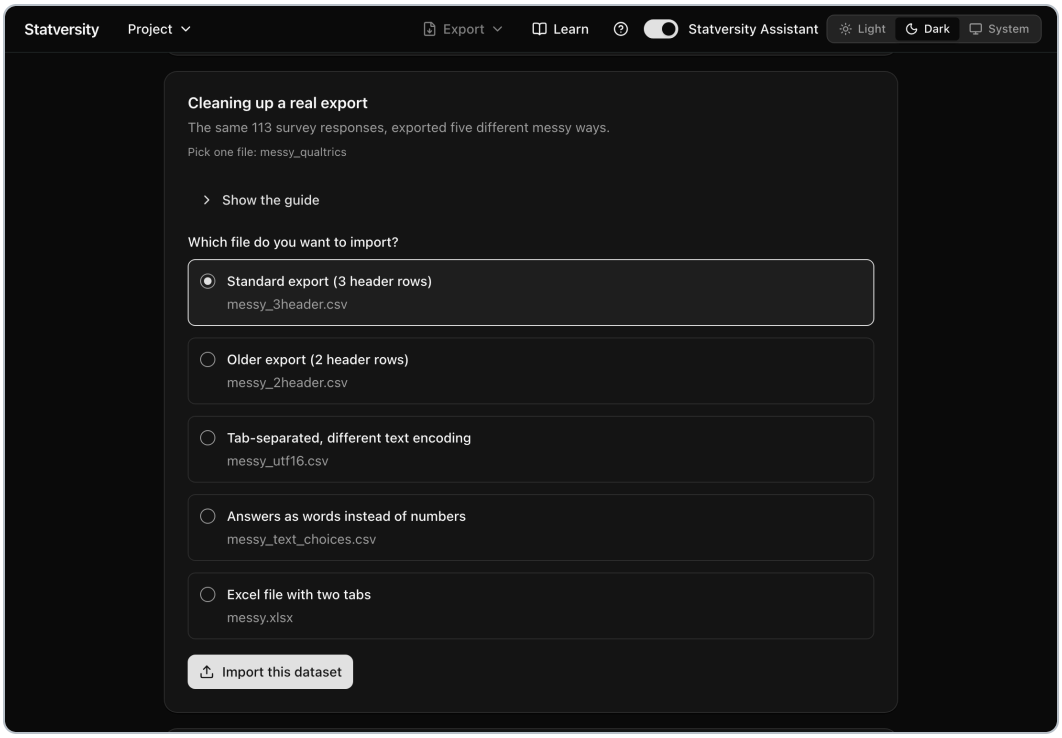


Figure 128. The Practice data screen showing the Cleaning up a real export card, with Standard export (3 header rows) selected under Which file do you want to import?

The card asks **Which file do you want to import?** and lists the five versions, from **Standard export (3 header rows)** to **Excel file with two tabs**. The first one is already selected, and that is the one to use. The other four are worth trying later; each shows Statversity handling a different kind of mess.

Steps: importing

1. Leave **Standard export (3 header rows)** selected and click **Import this dataset**.
2. The **Choose files** screen opens with the file and the survey file already attached.

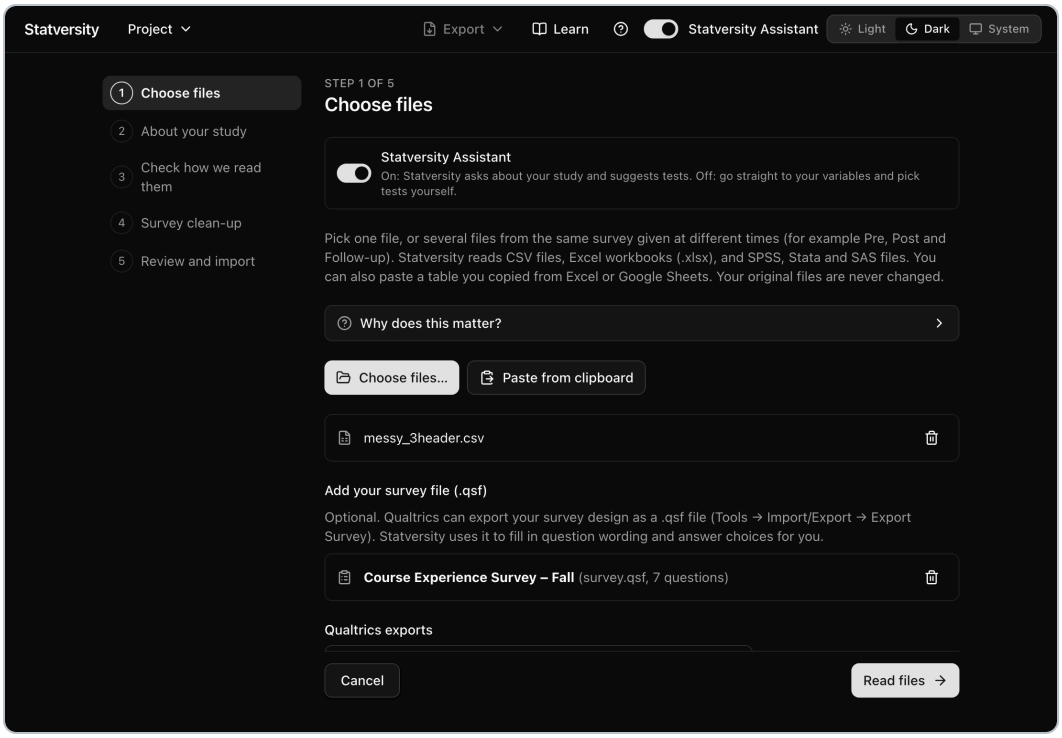


Figure 129. The Choose files screen with messy_3header.csv and the Course Experience Survey design file (7 questions) attached.

On the screenshot, you can see the data file `messy_3header.csv` and, below it, **Course Experience Survey – Fall (survey.qsf, 7 questions)**. The **Qualtrics exports** menu says **Detect automatically (recommended)**. Importing is covered in detail in chapter 5, so here we only follow the path.

3. Click **Read files**. The next screen is **About your study**.
4. Click **Skip these questions**. You can answer them if you like, but they are not needed for this walkthrough. The next screen is **Check how we read them**.

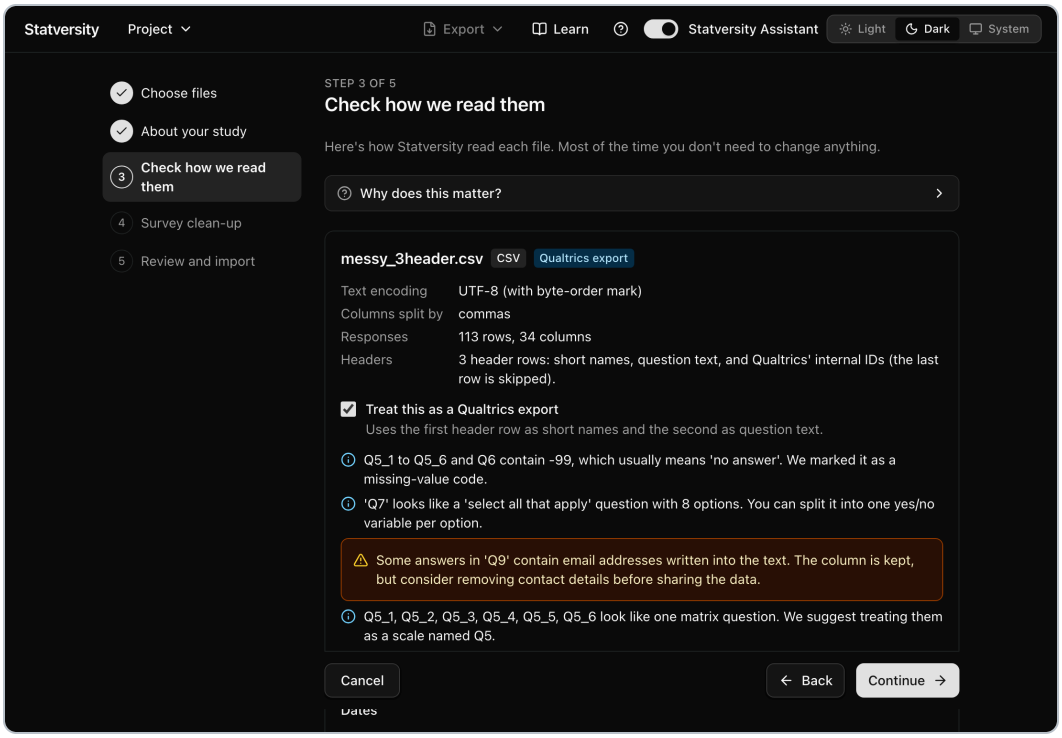


Figure 130. Check how we read them, showing 113 rows and 34 columns, three header rows, and several notices about -99, Q7, Q9 and the Q5 matrix.

This screen is Statversity telling you what it detected. Read it once, slowly, because it previews everything the clean-up screen will ask. It reports:

- **113 rows, 34 columns**, so 113 responses and 34 columns of information.
- **3 header rows**: short names, question text, and Qualtrics' internal IDs. The last row is skipped, so you do not get a meaningless row of codes in your data.
- A notice that **Q5_1 to Q5_6 and Q6 contain -99**, "which usually means 'no answer'", and that Statversity marked it as a missing-value code.
- A notice that **Q7** looks like a "select all that apply" question with 8 options.
- An orange warning that some answers in **Q9** contain email addresses written into the text. The column is kept, but you are reminded to consider removing contact details before sharing the data.
- A suggestion that **Q5_1 to Q5_6** look like one matrix question and could be treated as a scale named Q5.

You do not need to change anything here, since Statversity's guesses are right.

5. Click **Continue**. The screen shows **Survey clean-up**.

The Survey clean-up screen, one section at a time

The screenshot below shows the top of the screen. It says “A few choices about what to keep. Statversity picked safe defaults; check each one and change anything that doesn’t fit your study.”

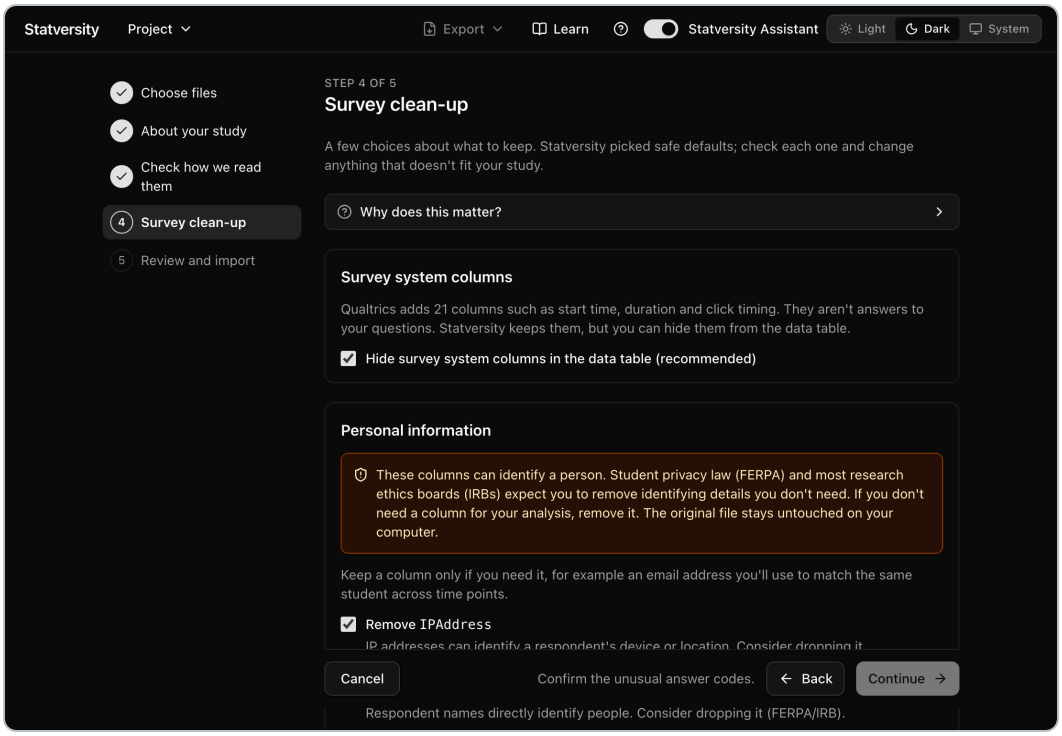


Figure 131. The top of Survey clean-up showing the Survey system columns and Personal information sections. The Continue button is greyed out with the message Confirm the unusual answer codes.

Notice the bottom bar. **Continue** is greyed out and beside it is a message, **Confirm the unusual answer codes.** You cannot move on until you have done that. We will get there in section five. The screen has six sections. You will scroll down through them in order.

Section 1: Survey system columns

6. Scroll to **Survey system columns** and read it.

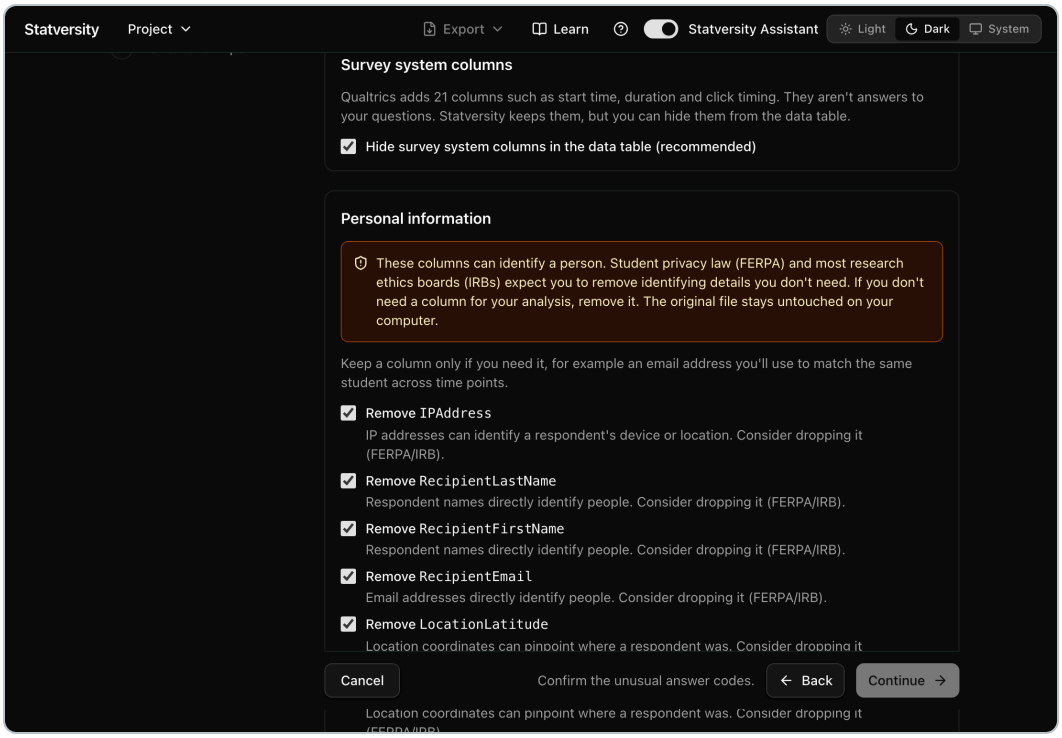


Figure 132. The Survey system columns section: Qualtrics adds 21 columns, with Hide survey system columns in the data table (recommended) ticked.

The problem in the raw export. When Qualtrics saves your survey data, it adds its own bookkeeping columns: when the person started, when they finished, how long it took, click timing, and so on. The screen says Qualtrics adds 21 columns of this kind. They are not answers to your questions.

What the choice does. The box **Hide survey system columns in the data table (recommended)** is ticked. Hiding does not delete anything. The columns are kept, but they are tucked out of sight in the data table so you are not scrolling past 21 columns you will rarely use. You can show them again later from the Variables screen. Leave this ticked.

> **IN PLAIN LANGUAGE**
> Hiding is reversible and deleting is not. When Statversity says “keeps them, but you can hide them”, it is giving you a tidy view without making a decision you could regret.

Section 2: Personal information

7. Scroll to **Personal information** and read it.

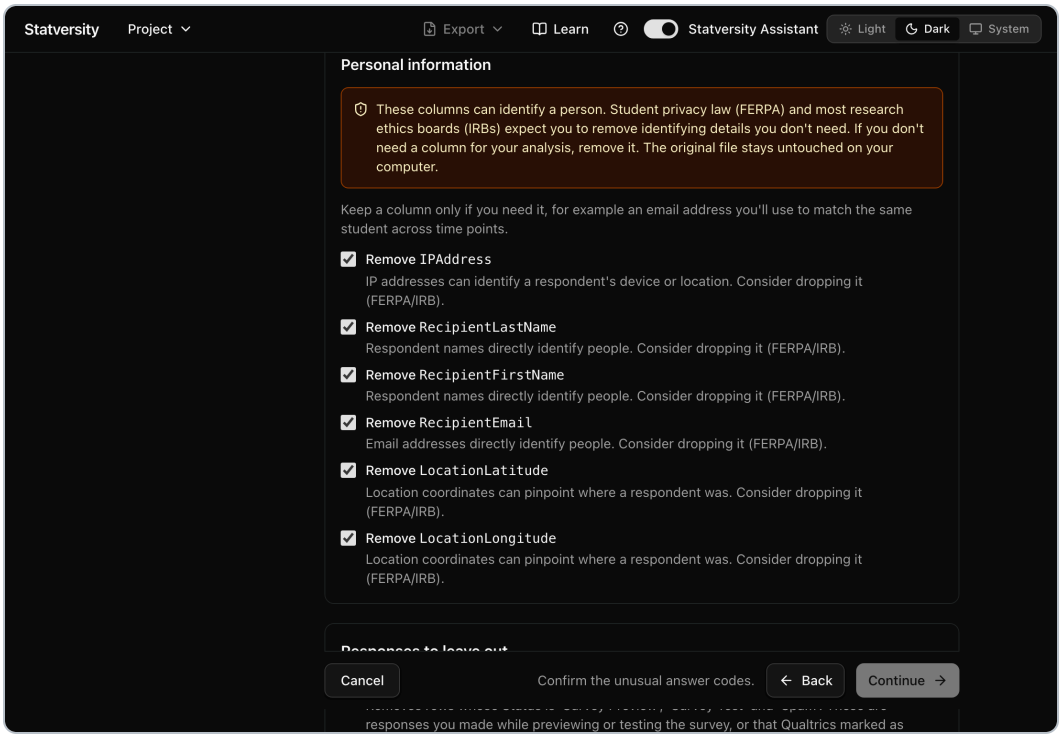


Figure 133. The Personal information section, listing Remove boxes for IPAddress, RecipientLastName, RecipientFirstName, RecipientEmail, LocationLatitude and LocationLongitude, all ticked.

The problem in the raw export. Qualtrics records the respondent's IP address, their name and email address (if you emailed them a survey link), and the map coordinates of where they were. Each of these can identify a person. The orange box explains that student privacy law (FERPA) and most research ethics boards (IRBs) expect you to remove identifying details you do not need.

What the choices do. There is one box per column: **Remove IPAddress**, **Remove RecipientLastName**, **Remove RecipientFirstName**, **Remove RecipientEmail**, **Remove LocationLatitude** and **Remove LocationLongitude**. Each ticked box means the column is dropped when the data is imported. Your original file on your computer is never changed; only Statversity's working copy loses the column.

8. Check that all six boxes are ticked. They are, by default.

You would untick one only if you truly need it. The screen gives the example of an email address you will use later to match the same student across time points (chapter 14 covers matching). For this reliability analysis you need none of them.

> **CAUTION**
> These six are separate columns. A name or email written into an open-ended text answer is not one of them. That is why the earlier screen warned you separately about email addresses in Q9. Look through free-text columns yourself before sharing data.

Section 3: Responses to leave out

9. Scroll to **Responses to leave out** and read it.

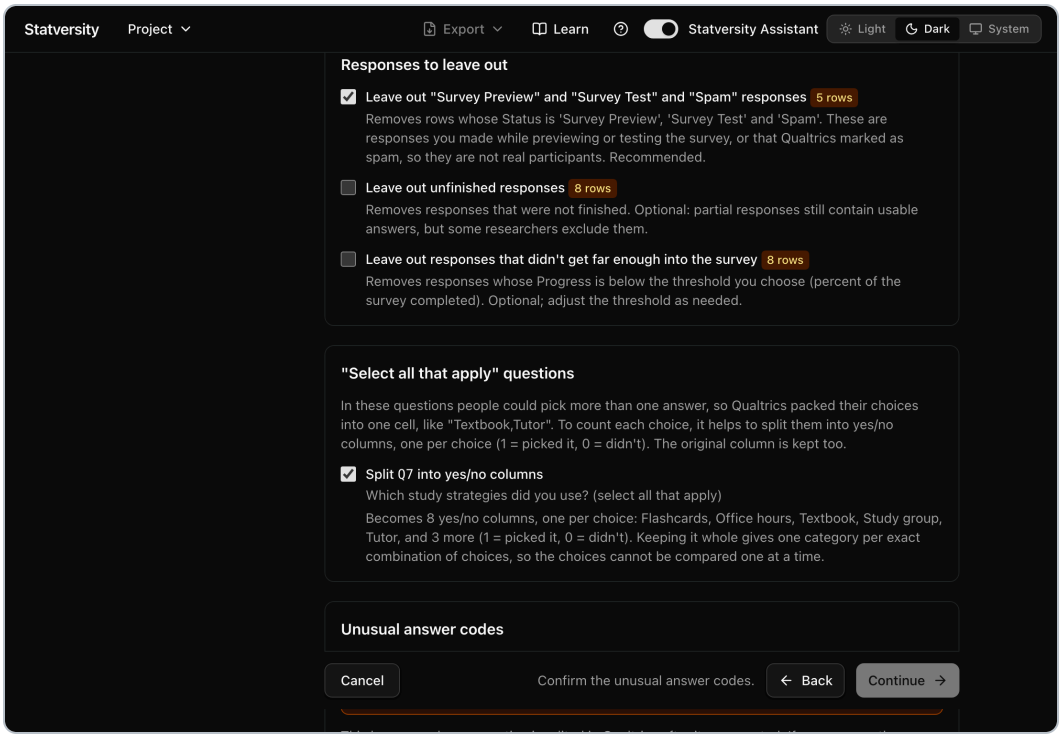


Figure 134. The Responses to leave out section with three options: Survey Preview, Survey Test and Spam responses (5 rows, ticked); unfinished responses (8 rows, unticked); and responses that didn't get far enough into the survey (8 rows, unticked).

The problem in the raw export. The file contains rows that are not real participants. When you preview or test your survey, Qualtrics records those clicks as responses with a status of “Survey Preview” or “Survey Test”. Qualtrics can also flag responses as “Spam”.

What the choices do. Each option shows a badge with the number of rows it would remove:

- **Leave out “Survey Preview” and “Survey Test” and “Spam” responses** (**5 rows**). Ticked by default and marked Recommended. These are not participants.
- **Leave out unfinished responses** (**8 rows**). Unticked. These are people who started but did not finish. They still contain usable answers, so excluding them is optional; some researchers do and some do not.
- **Leave out responses that didn’t get far enough into the survey** (**8 rows**). Unticked. This one uses the Progress value (percent of the survey completed) and a threshold you choose.

10. Check that **Leave out “Survey Preview” and “Survey Test” and “Spam” responses** is ticked, and leave the other two unticked.

The five removed rows are the difference between the 113 responses in the file and the 108 you will analyze. We return to that number at the end.

> **IN PLAIN LANGUAGE**
> Whether to drop unfinished responses is a judgment call about your study, not a technical question. Dropping them gives you a cleaner dataset but a smaller one. Whatever you decide, report it in your methods section.

Section 4: “Select all that apply” questions

11. Scroll to **“Select all that apply” questions** and read it.

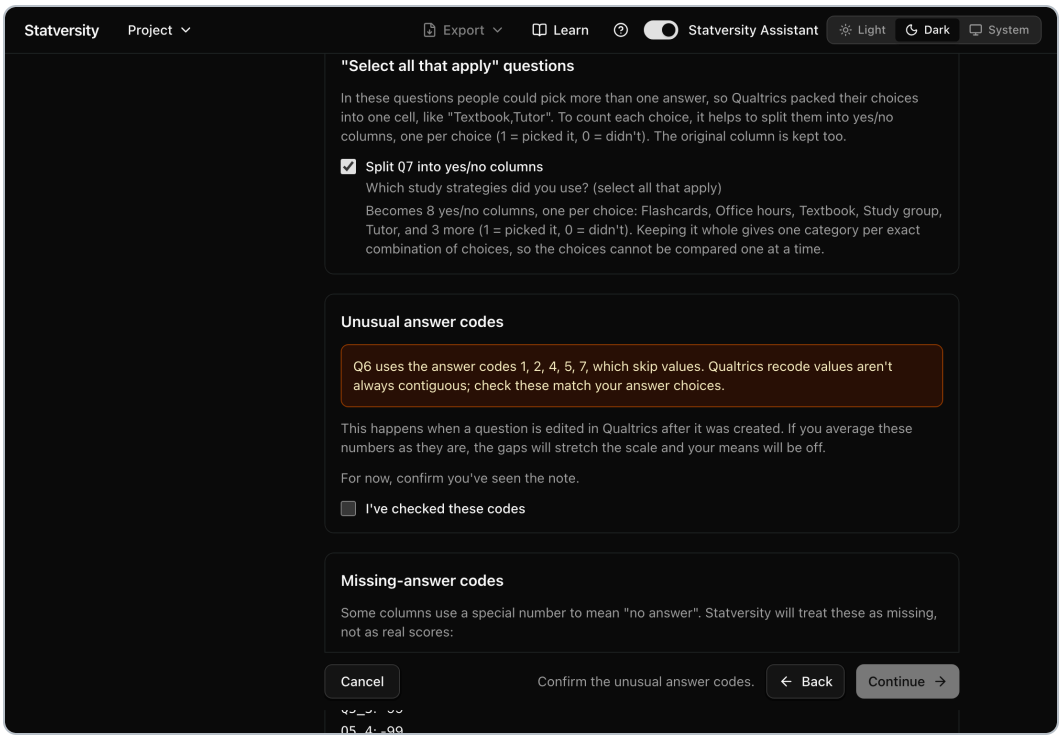


Figure 135. The Select all that apply section: Split Q7 into yes/no columns is ticked, with the question Which study strategies did you use? and a note that it becomes 8 yes/no columns.

The problem in the raw export. In a question like “Which study strategies did you use? (select all that apply)”, a student can tick several boxes. Qualtrics packs all of someone’s choices into one cell, like “Textbook,Tutor”. If you tried to count how many students used a tutor, you would have to untangle thousands of different combinations.

What the choice does. Split Q7 into yes/no columns (ticked by default) replaces that one packed cell with one column per option. The screen says it “becomes 8 yes/no columns, one per choice: Flashcards, Office hours, Textbook, Study group, Tutor, and 3 more”, where 1 means the person picked it and 0

means they did not. The original packed column is kept too. The text underneath explains the alternative: if you keep it whole, you get one category for every exact combination of choices, so the choices cannot be compared one at a time.

Leave it ticked. We will not analyze Q7 in this chapter, but splitting costs nothing and keeps your options open.

Section 5: Unusual answer codes

12. Scroll to **Unusual answer codes** and read it.

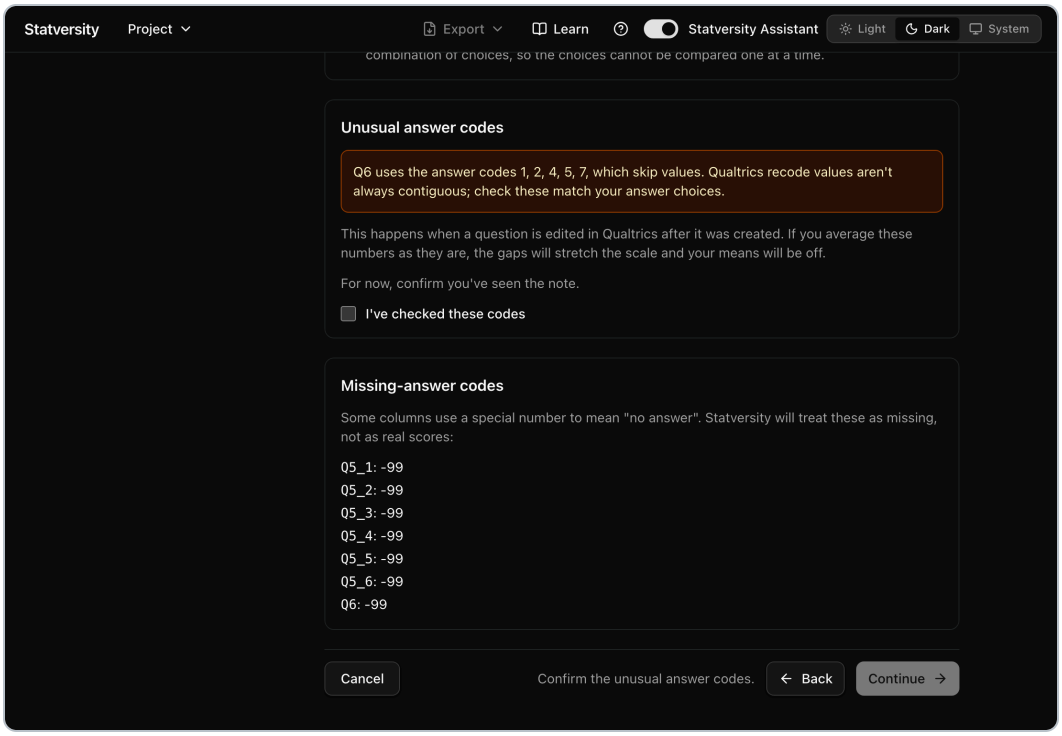


Figure 136. The Unusual answer codes section with an orange note about Q6 and an unticked box labelled I've checked these codes.

The problem in the raw export. The orange note reads: “Q6 uses the answer codes 1, 2, 4, 5, 7, which skip values.” Q6 asks “How satisfied are you with the course overall?” The answers are stored as numbers, but they are not the tidy 1, 2, 3, 4, 5 you would expect. The explanation underneath is that this happens when a question is edited in Qualtrics after it was created, and that if you average numbers with gaps in them, “the gaps will stretch the scale and your means will be off”.

What the choice does. Statversity cannot fix this for you, because only you know what each number meant in your survey. So it asks you to confirm that you have seen the note. This is the box that blocks **Continue**.

13. Do not tick it yet. First scroll one more section to see the last part of the screen.

Section 6: Missing-answer codes

14. Scroll to **Missing-answer codes** and read it.

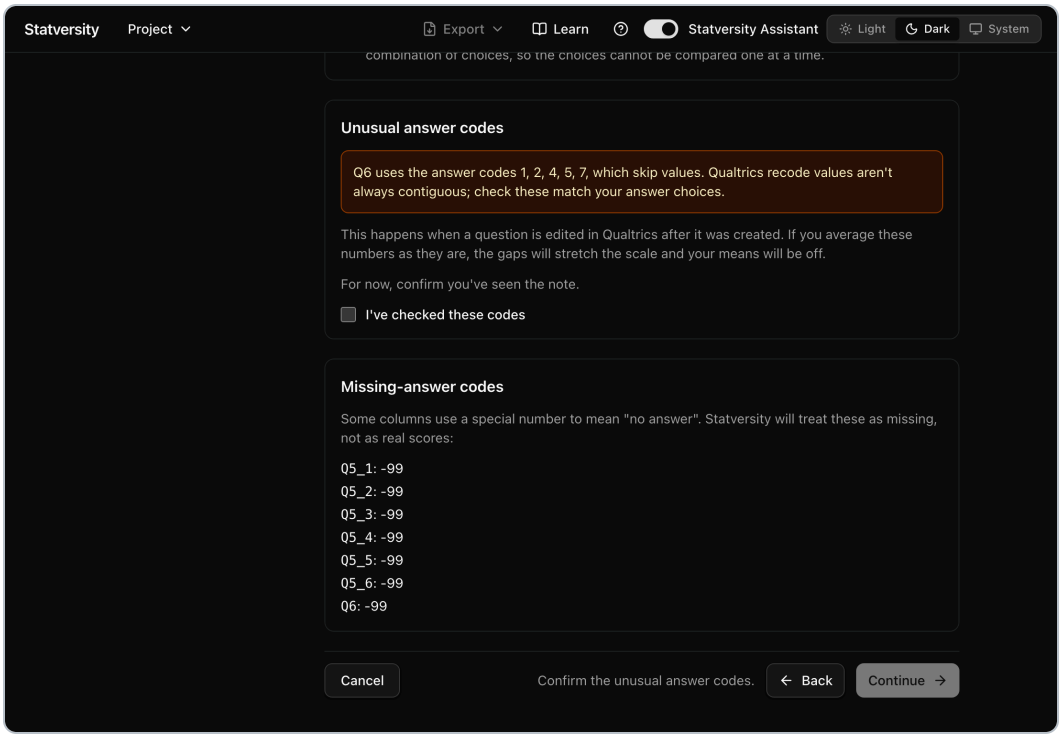


Figure 137. The Missing-answer codes section listing Q5_1 through Q5_6 and Q6, each with the code -99, under the Unusual answer codes section.

The problem in the raw export. Some people skipped some questions. In this file, a skipped answer was recorded as the number -99 instead of being left blank. If Statversity averaged a -99 as though it were a real rating, one skipped question would drag that student's score far below 1 and wreck everything.

What the choice does. The section reads: "Some columns use a special number to mean 'no answer'. Statversity will treat these as missing, not as real scores." It lists **Q5_1: -99** through **Q5_6: -99**, and **Q6: -99**. There is nothing to tick; this is Statversity telling you what it already decided, based on the pattern in your data. Your job is to read the list and check that -99 really does mean "no answer" in your survey. (It does here.)

> **IN PLAIN LANGUAGE**
> "Missing" has a precise meaning in statistics: "no usable value here". Marking -99 as missing means Statversity skips it when it computes averages and counts, instead of treating it as a very negative rating.

Confirming the codes

15. Scroll back up a little and tick **I've checked these codes**.

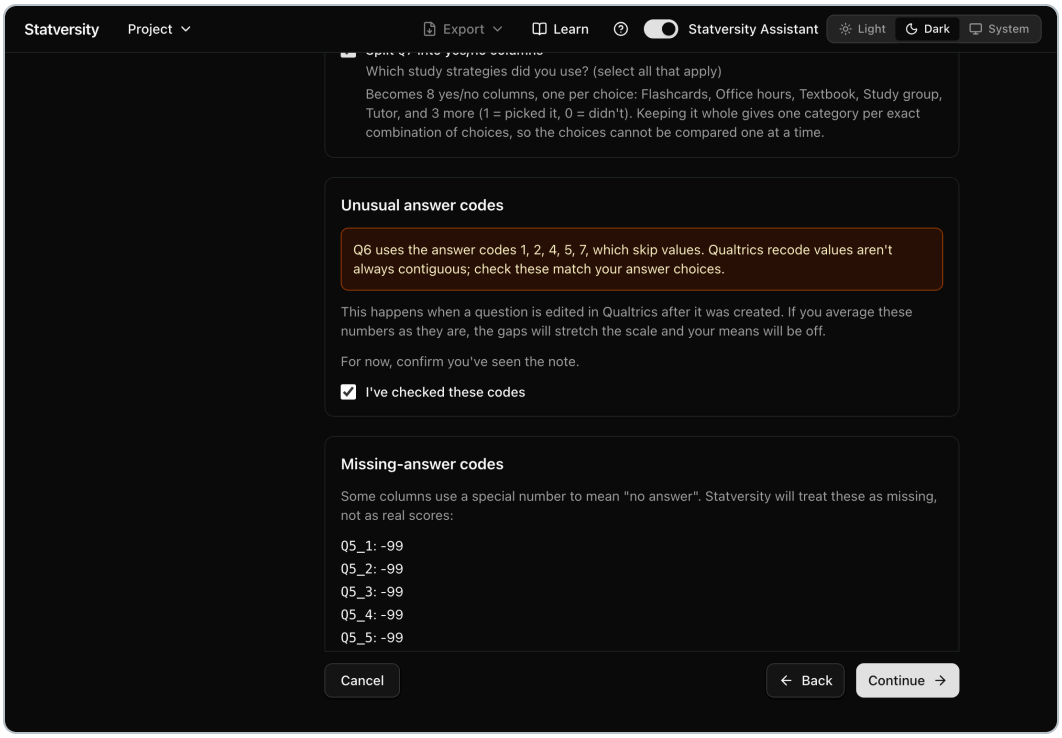


Figure 138. The same area after ticking I've checked these codes. The Continue button is now bright and enabled.

On the screenshot above, the box is ticked and **Continue** is no longer greyed out. The message that sat beside it, "Confirm the unusual answer codes.", has gone. This is a deliberate speed bump. Statversity will not let you carry on until you have acknowledged that Q6's codes skip values, because the mistake it prevents (wrong means) is silent: nothing would visibly break.

16. Click **Continue**. The screen shows **Review and import**.

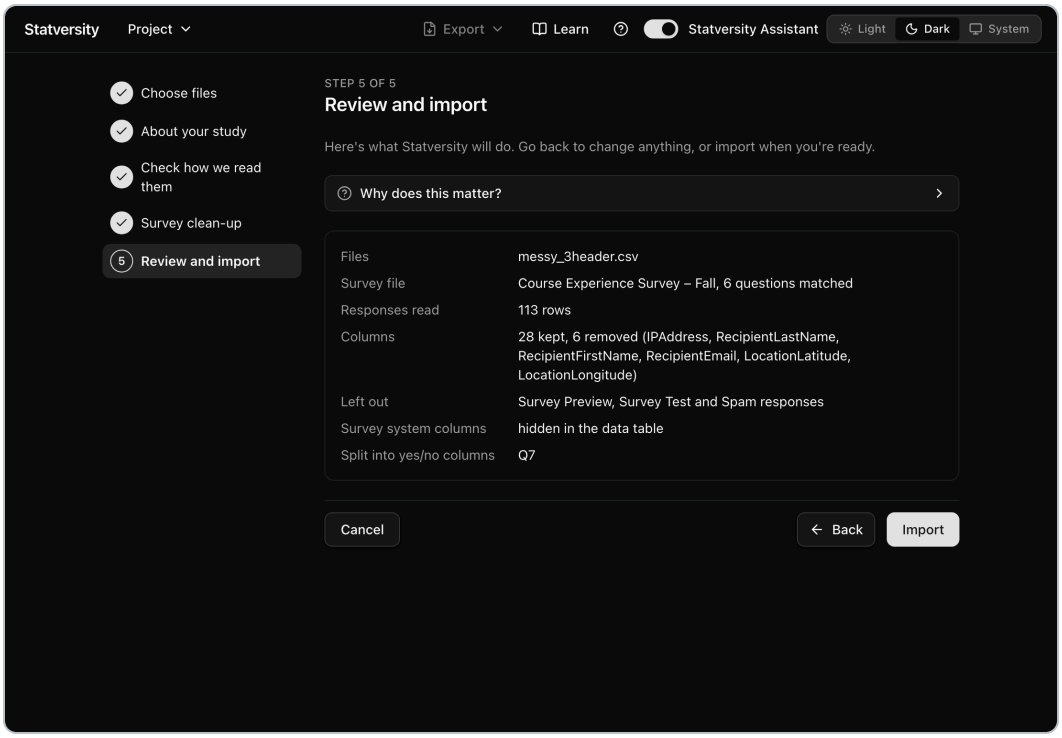


Figure 139. Review and import: 113 rows read, 28 kept and 6 removed columns, Survey Preview, Survey Test and Spam responses left out, survey system columns hidden, and Q7 split into yes/no columns.

This is your receipt. It lists what Statversity is about to do: the file (**messy_3header.csv**), the survey file ("Course Experience Survey – Fall, 6 questions matched"), **113 rows** read, **28 kept, 6 removed** columns (named one by one), the response types **left out**, system columns **hidden in the data table**, and **Q7** split into yes/no columns. If anything is wrong, **Back** takes you to change it.

17. Click **Import**. The screen shows **Welcome**, the start of the Variable Interview.

The Variable Interview

The interview (chapter 6 explains every screen) asks Statversity's guesses to be confirmed. Here most of it is "check and click Continue".

18. Click **Start**. The screen shows **What is each question?**

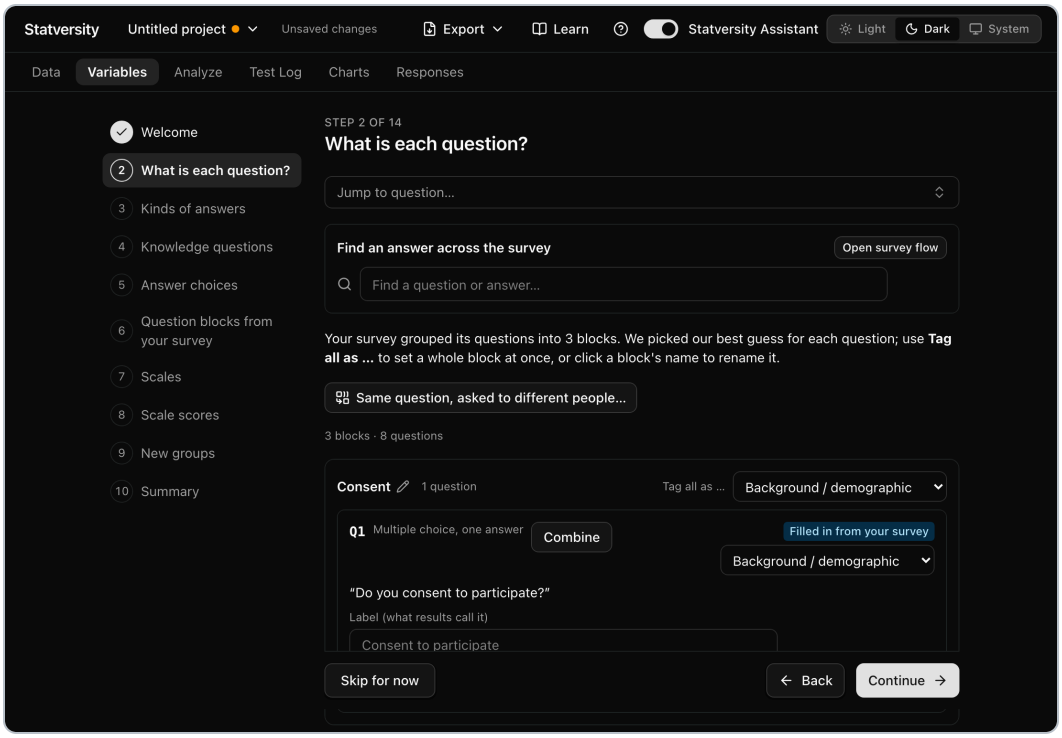


Figure 140. What is each question? shows the survey grouped into 3 blocks: Consent, Classroom experience and Open comments, with Statversity's guess for each question.

The survey's own design file has told Statversity that the questions come in three blocks and what each question is. Each guess is labelled **Filled in from your survey**. You can see Q1 (consent) and the **Q5 (6 items)** matrix, guessed as **Survey (Likert) item**. A Likert item is a rating statement such as "strongly disagree" to "strongly agree".

19. Click **Continue** through the next screens, checking each one as you go: four **Kinds of answers** screens, then **Knowledge questions**, then **Answer choices**.

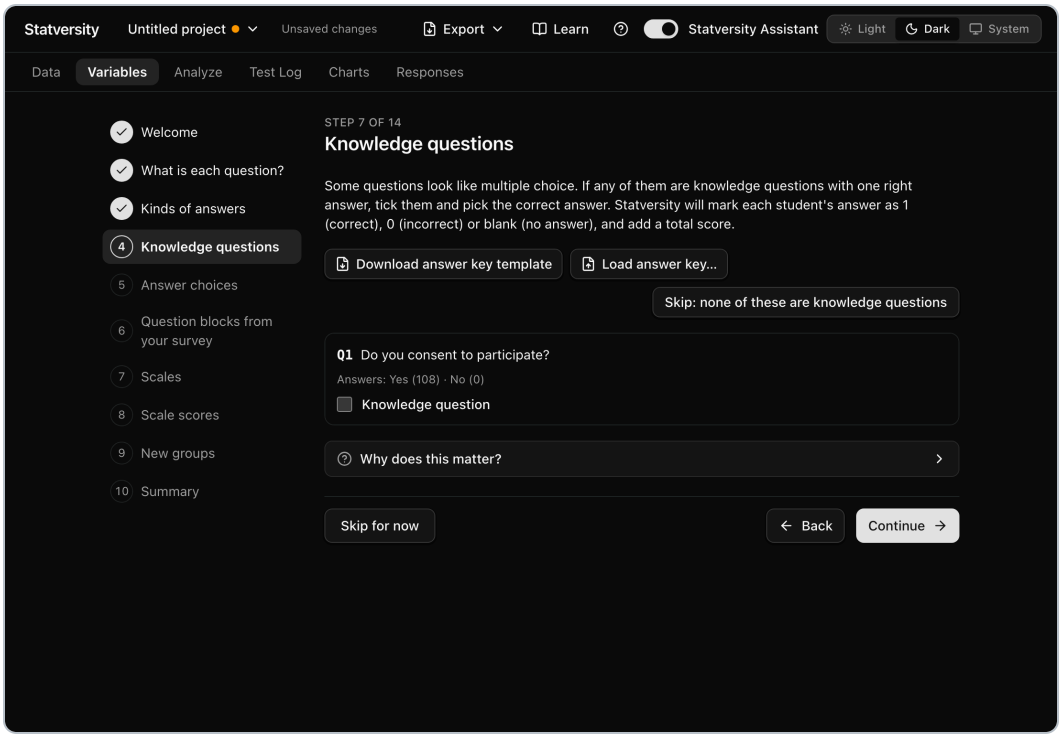


Figure 141. Knowledge questions asking whether any question is a test question with one right answer. Q1 is shown with its Knowledge question box unticked.

Knowledge questions only matters for quizzes with right and wrong answers. This survey asks for opinions, so leave the box unticked.

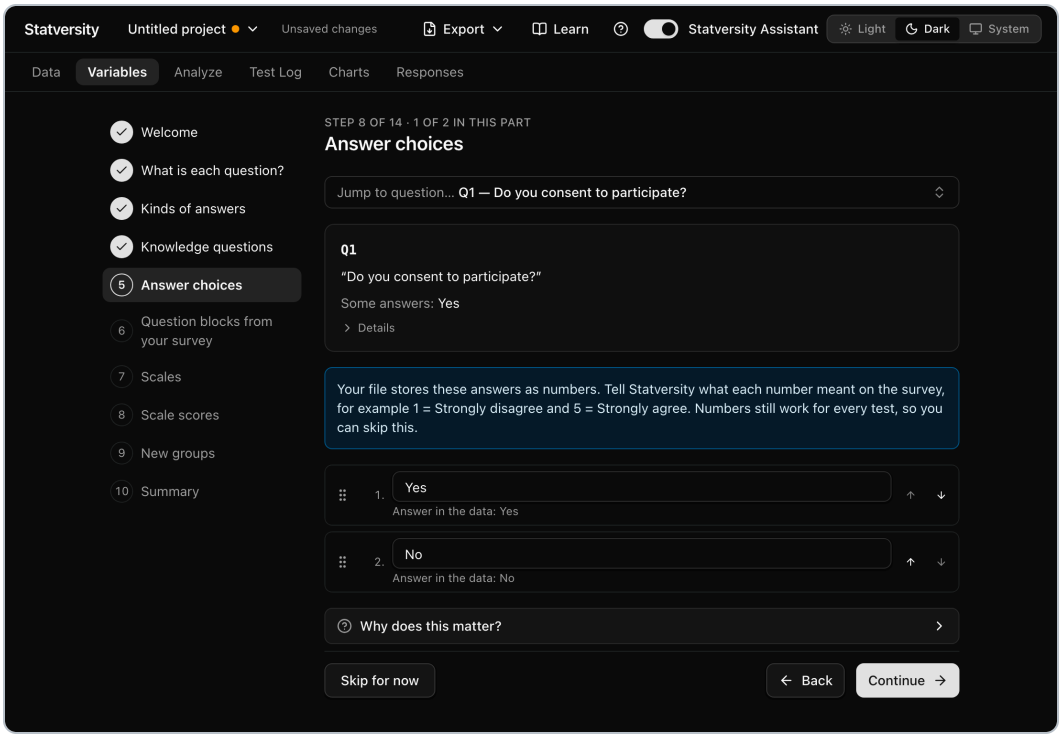


Figure 142. Answer choices for Q1, showing Yes and No.

Answer choices shows the labels attached to each coded answer (for example Yes and No). Nothing to change.

20. Click **Continue** until the screen shows **Question blocks from your survey**.

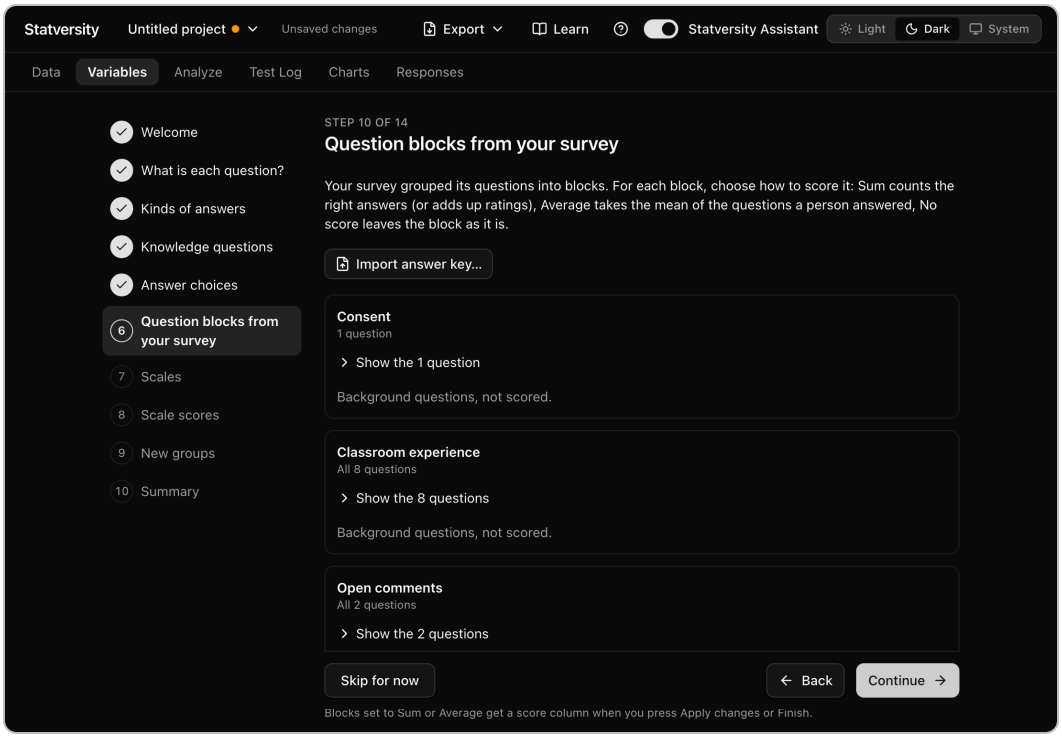


Figure 143. Question blocks from your survey with three blocks, Consent, Classroom experience and Open comments, each set to Background questions, not scored.

Each block is currently “Background questions, not scored”, which is fine. We will build the score in the next step, where we get to tell Statversity about the reversed item.

The reversed item: Q5_4

21. Click **Continue**. The screen shows **Scales**.

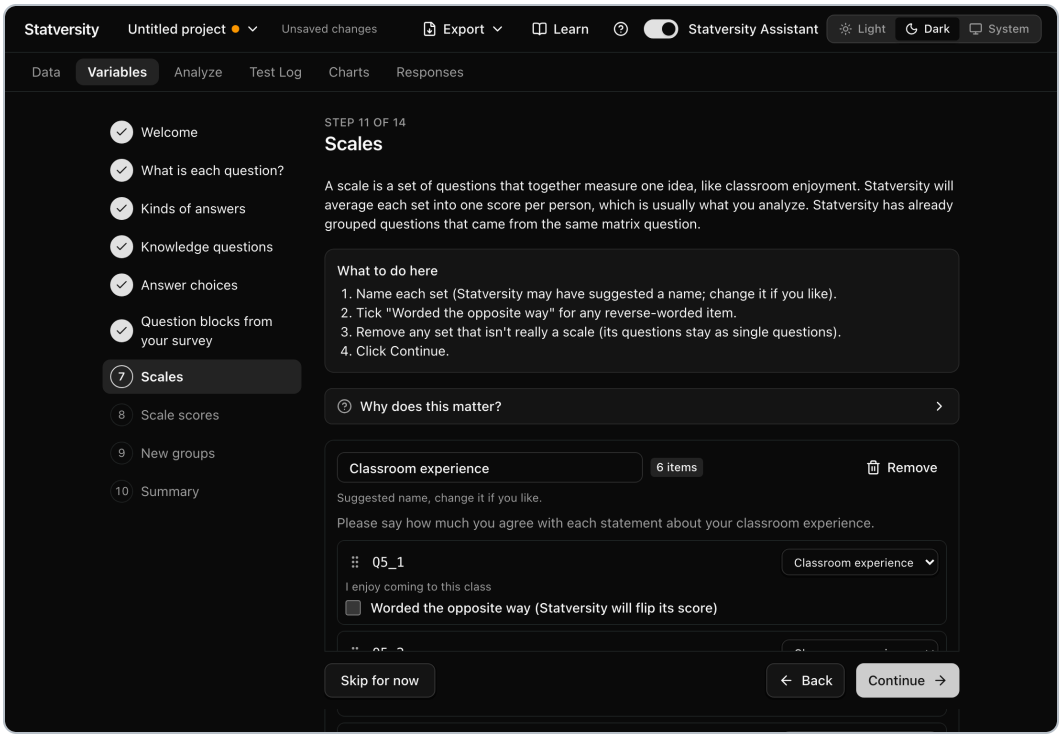


Figure 144. The Scales screen with Classroom experience, 6 items, the six Q5 questions each with a Worded the opposite way checkbox, all unticked.

Statversity has grouped Q5_1 to Q5_6 into a scale and suggested the name **Classroom experience**. Each of the six items has a checkbox, **Worded the opposite way (Statversity will flip its score)**.

Why this matters. Look at the six statements:

- Q5_1: I enjoy coming to this class
- Q5_2: The teacher explains things clearly
- Q5_3: I feel comfortable asking questions
- Q5_4: **I often feel lost in this class**
- Q5_5: The activities help me learn
- Q5_6: I would recommend this class to a friend

Five of these are positive statements, where “strongly agree” means a good experience. Q5_4 is a negative statement: a student who strongly agrees that they feel lost is having a bad experience. If you left it alone, a happy student would score high on five items and low on Q5_4, and the six items would appear to disagree with each other. That would make your scale look unreliable when the real problem is only the direction of the wording.

22. For **Q5_4**, tick **Worded the opposite way**.

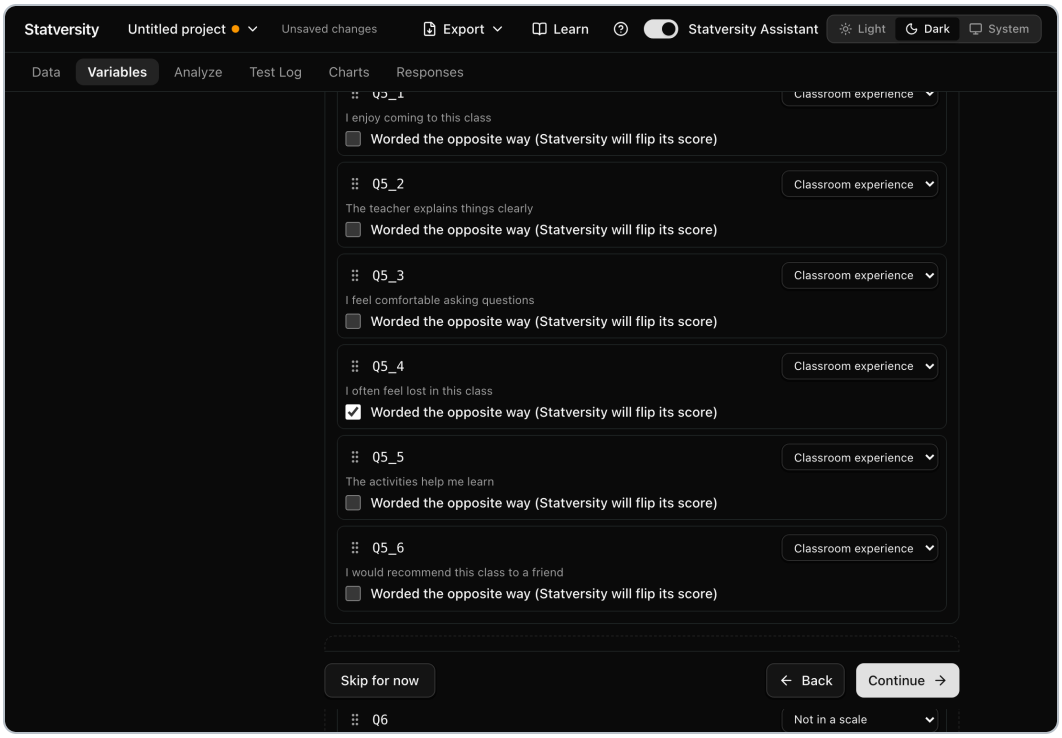


Figure 145. The Scales screen with Q5_4, I often feel lost in this class, ticked as Worded the opposite way. The other five boxes are unticked.

Only Q5_4 is ticked on the screenshot. Statversity will flip its scores, so that on this item a 1 becomes a 5, a 2 becomes a 4, and so on, before it is combined with the others. Statversity does this for the scale score and for Cronbach's alpha. The analysis result will remind you that it did this.

> **CAUTION**
>
> Statversity does not know which of your items are reversed. It cannot read the meaning of the words, so it is your job to tick the box. Look at each item and ask, "Does a high answer mean more of the thing this scale measures?" If not, tick the box.

23. Click **Continue**. The screen shows **Scale scores**.

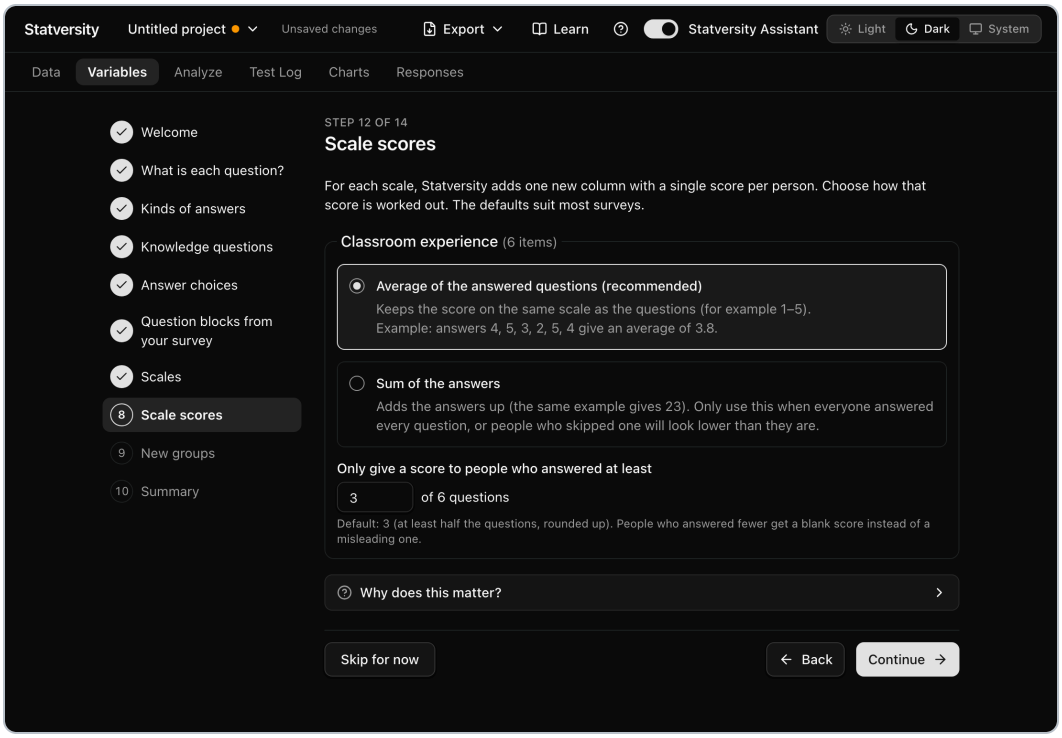


Figure 146. Scale scores with Average of the answered questions (recommended) selected and a minimum of 3 of 6 questions.

Here you choose how the six answers become one score per student. **Average of the answered questions (recommended)** keeps the score on the same 1 to 5 scale as the questions. The alternative, **Sum of the answers**, adds them and is only fair when everyone answered every question. Below, **Only give a score to people who answered at least 3 of 6 questions** means a student who answered fewer than half gets a blank score rather than a misleading one. Leave the defaults.

24. Click **Continue** twice more (past **New groups**) to reach **Summary**.

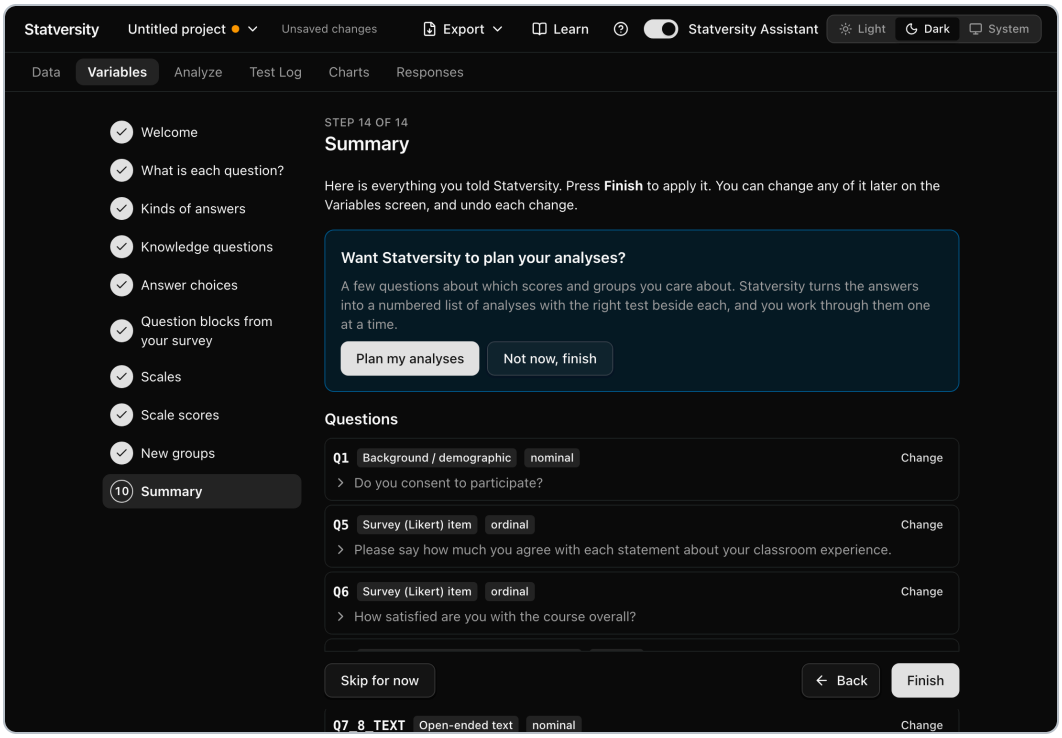


Figure 147. The Summary screen listing each question with its role and level, such as Q5 as a Survey (Likert) item that is ordinal.

The summary lists everything you told Statversity. It also offers to plan your analyses; we will not use that here.

25. Click **Finish**. The screen shows **Variables**.

Running Cronbach’s alpha

26. Click **Analyze**. The Test Advisor opens.
27. In **Which score or answer do you want to analyse?**, choose **Classroom experience (average score)**.

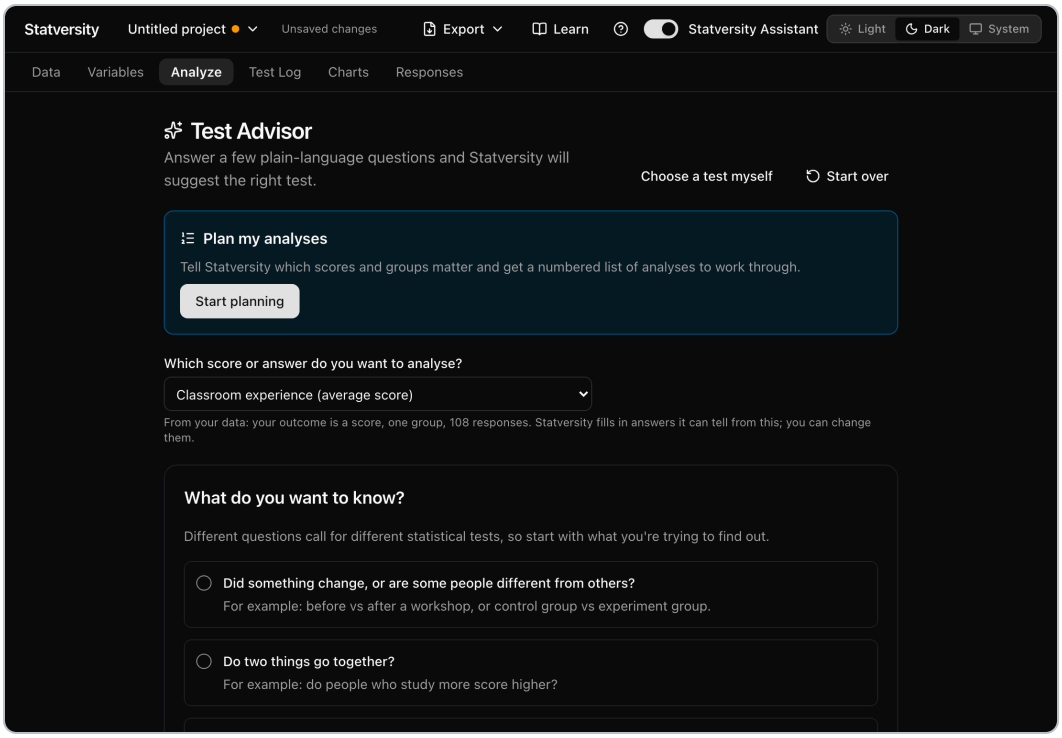


Figure 148. The Test Advisor with Classroom experience (average score) chosen, 108 responses, and the question What do you want to know? below.

Notice the line under the menu: **108 responses**. That is the cleaned count. The 5 test and spam rows are already gone.

28. Under **What do you want to know?**, choose **Do my rating questions consistently measure one idea?** and click **Continue**.

> **IN PLAIN LANGUAGE**
>
> This is the plain-language version of "is my scale reliable?" You are not asking whether students are happy. You are asking whether the six statements behave like six readings of the same thing, so that averaging them is sensible.

29. Under **What kind of questions are they?**, choose **Rating questions (Likert-scale)** and click **Continue**.

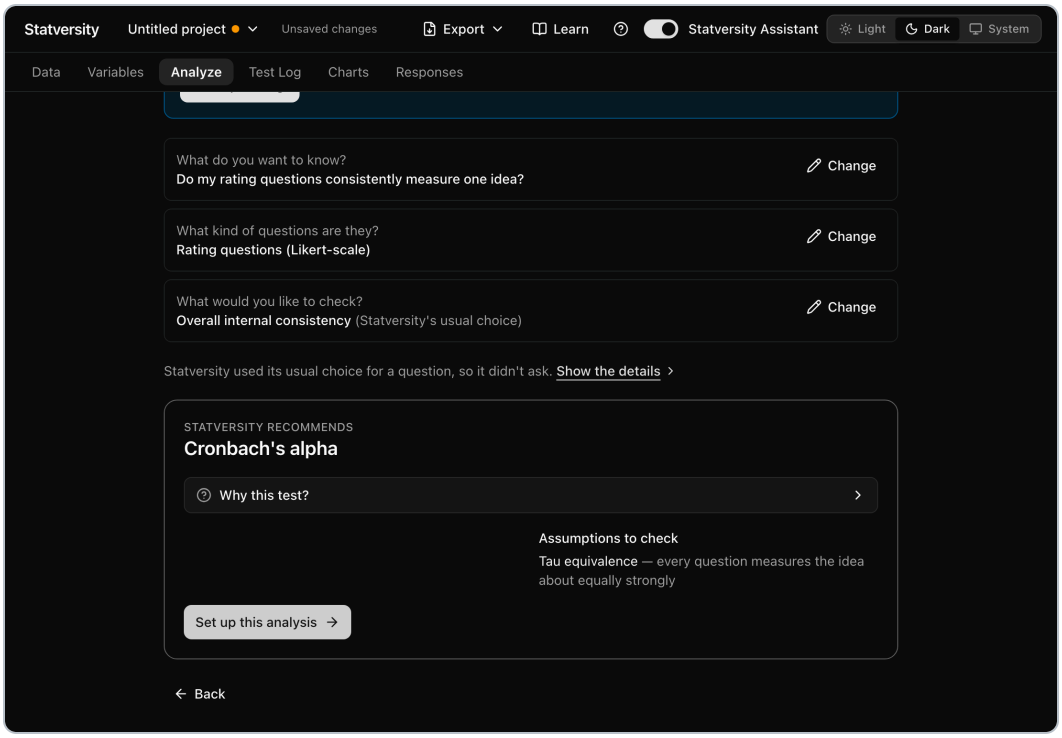


Figure 149. The Advisor recommending Cronbach's alpha, with the assumption Tau equivalence: every question measures the idea about equally strongly.

Statversity recommends **Cronbach's alpha**. Its assumption to check is "Tau equivalence", which means every question measures the idea about equally strongly. That is a technical way of saying that none of the six items should be much weaker than the rest.

30. Click **Set up this analysis**.

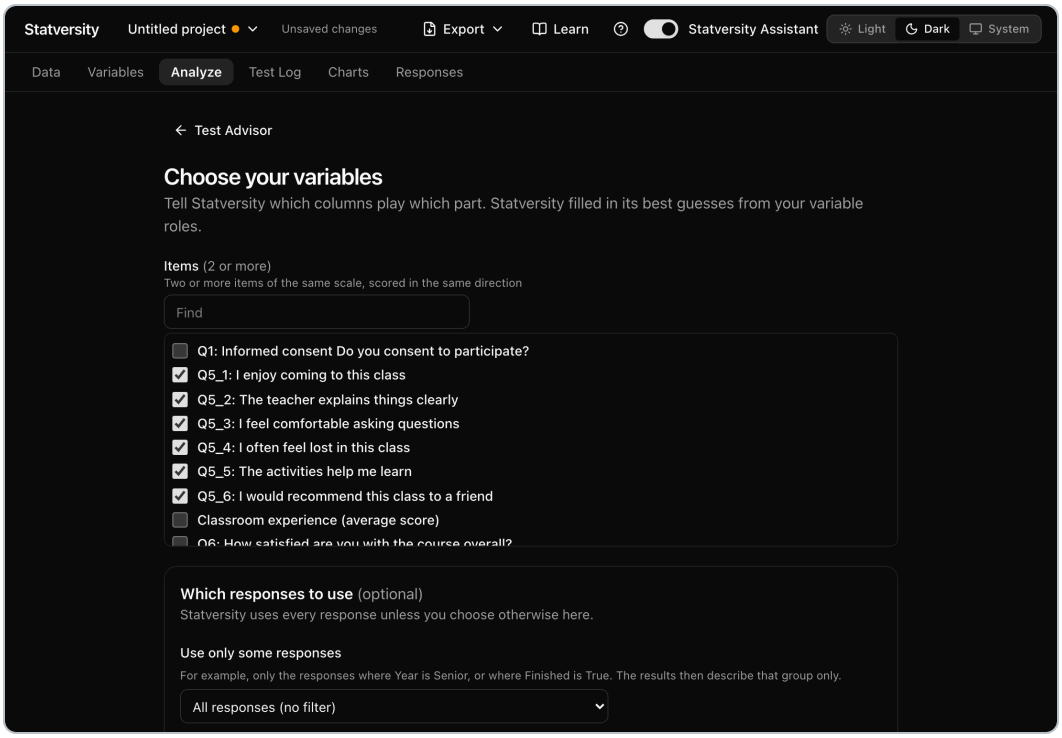


Figure 150. Choose your variables with the six Q5 items ticked under Items.

Items already has exactly the six Q5 statements ticked, as Statversity filled in. The list also contains items like Q1 and the score “Classroom experience (average score)”, which are not ticked and should not be. Alpha needs the six original items, not their average.

31. Check the list, then click **Check the assumptions**. The results appear.

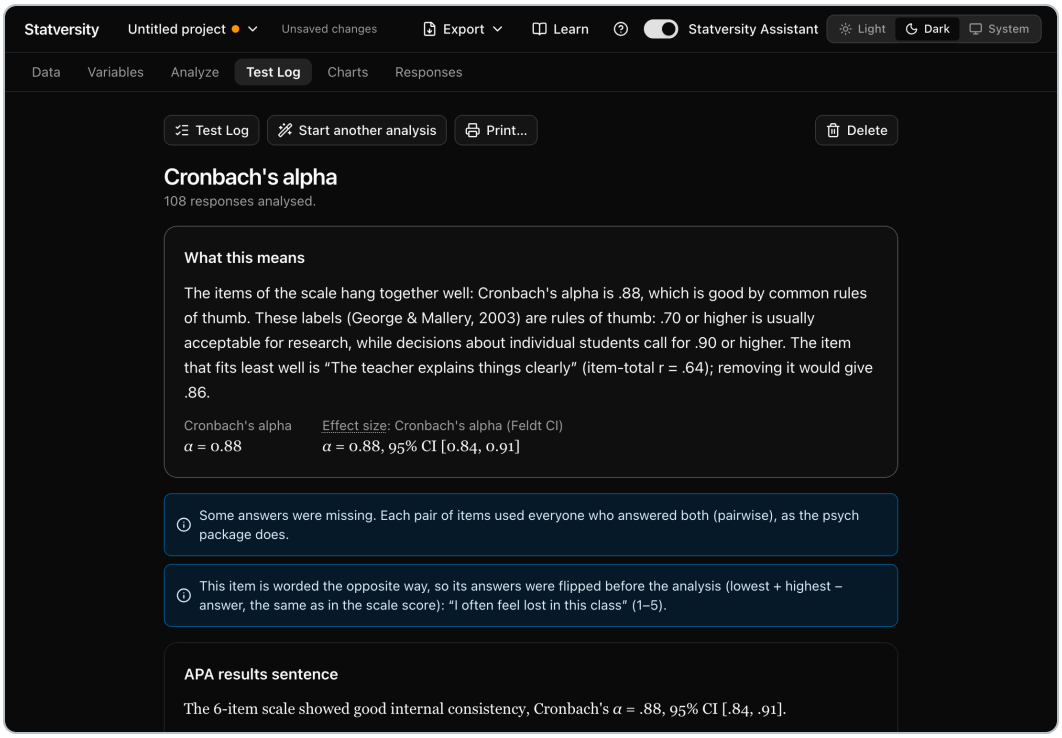


Figure 151. The top of the result: Cronbach's alpha, 108 responses analysed, and What this means saying alpha is .88.

What Statversity found

The result is for 108 responses.

Test	Cronbach's alpha, 6 items
Cronbach's alpha	α = .88
95% confidence interval	[.84, .91]
Responses analysed	108
Verdict	"good by common rules of thumb"
Item that fits least well	"The teacher explains things clearly", item-total r = .64

Beneath the headline, two blue notices explain what Statversity did behind the scenes.

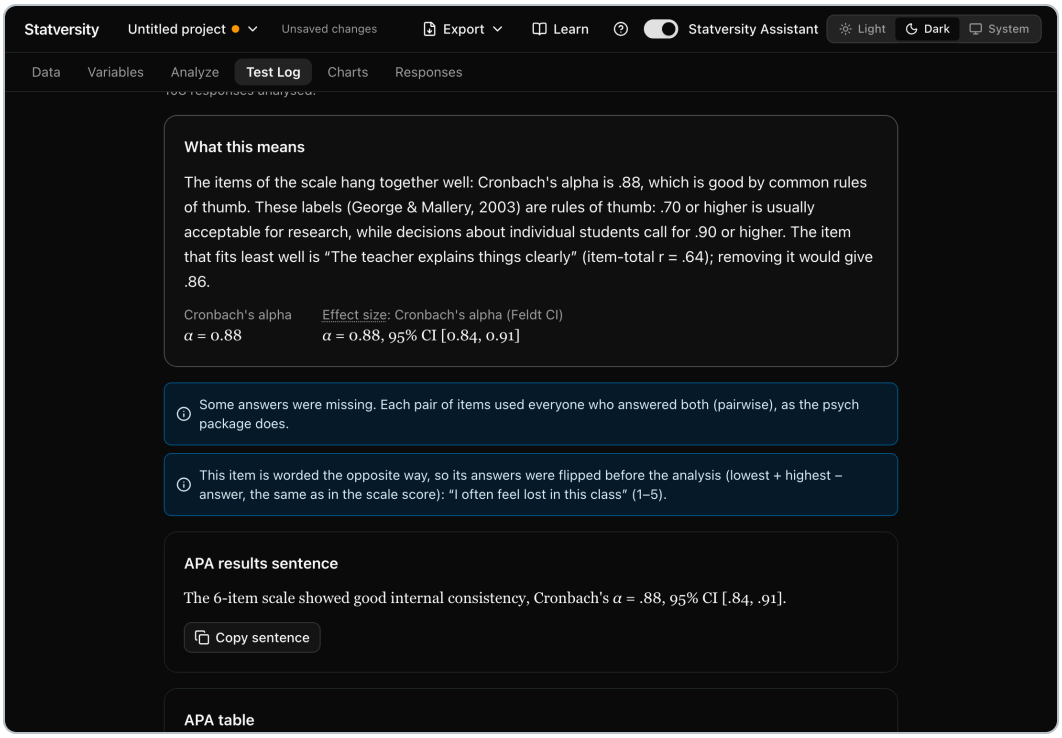


Figure 152. The What this means box with two notices: some answers were missing, so each pair of items used everyone who answered both; and the item I often feel lost in this class was flipped before analysis.

The first says that some answers were missing, so "each pair of items used everyone who answered both (pairwise)". That is the -99 codes at work: skipped answers were treated as missing rather than as numbers. The second says that "this item is worded the opposite way, so its answers were flipped before the analysis", naming **"I often feel lost in this class" (1–5)**. That is the box you ticked for Q5_4, doing its job.

Scroll down to **APA results sentence** and **APA table**.

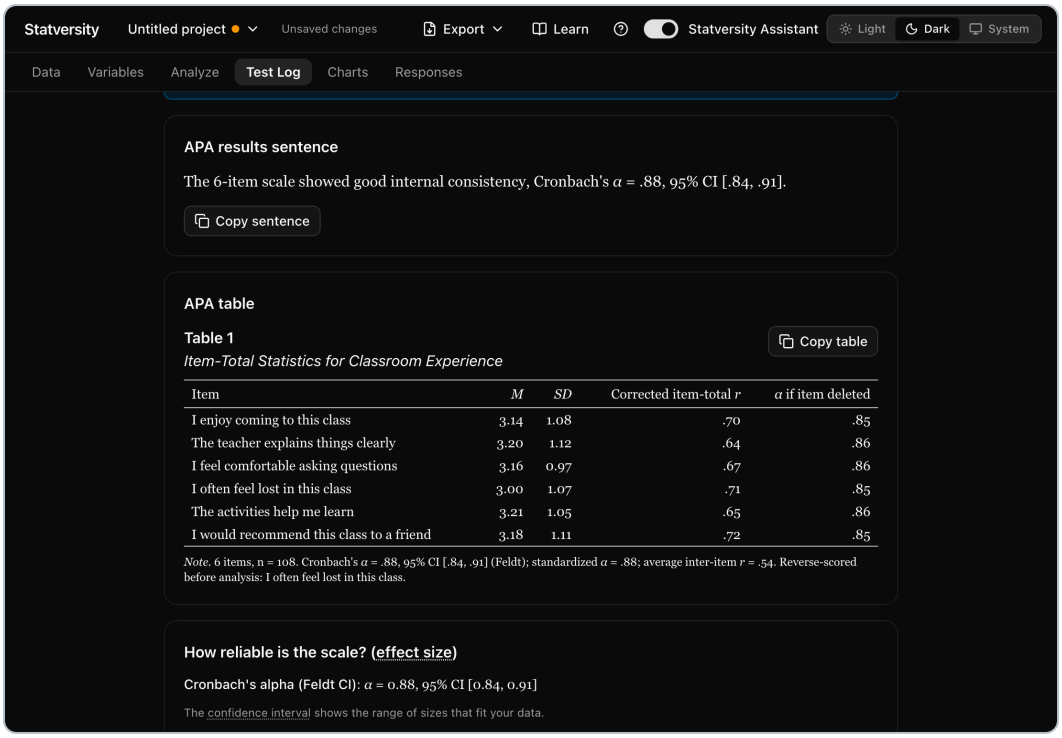


Figure 153. The APA results sentence and the APA table Item-Total Statistics for Classroom Experience, with means, standard deviations, corrected item-total correlations and alpha if item deleted for each item.

The APA table gives one row per item. For each item it shows the average answer (**M**), the spread (**SD**), the **corrected item-total r** (how strongly that item moves with the other five, with .70 or so being strong), and **α if item deleted** (what alpha would be if you removed it). Here the item-total correlations run from .64 to .72 and alpha-if-deleted is between .85 and .86 for every item. Its note confirms **6 items, n = 108**, and that "I often feel lost in this class" was reverse-scored before analysis.

What this means in plain English

The six statements hang together well. An alpha of .88 means that students who rate one statement high tend to rate the others high too, which is what you would expect if the six are measuring one thing, which we are calling classroom experience. (Alpha on its own cannot prove that the items measure a single idea; a factor analysis, shown in chapter 15, checks that more directly.) Statversity uses the common rule of thumb that .70 or higher is acceptable for research. Adding the six answers into one score is reasonable.

Every item contributes. Every "alpha if item deleted" value (.85 to .86) is lower than the full-scale .88, which means that deleting any single item would make the scale slightly worse, not better. There is no item you should throw out.

Also notice what the clean-up bought you. Five test and spam rows were removed before the test, so the 108 people in the analysis are real participants. The -99 codes were treated as missing, so skipped questions did not distort anyone's score. And one reversed item was flipped, so it pulled in the right direction.

What you would write in a paper

Statversity gives you this sentence:

> The 6-item scale showed good internal consistency, Cronbach's α = .88, 95% CI [.84, .91].

In a methods section you would add the clean-up, for example: "Responses with a Qualtrics status of survey preview, survey test or spam (n = 5) were removed, leaving 108 responses. Identifying columns were removed. One negatively worded item was reverse scored before analysis." Change "6-item scale" to the name of your scale, and keep the numbers exactly as Statversity shows them.

What if

- **I forgot to tick Worded the opposite way for Q5_4.** Then a high answer on that item would mean the opposite of a high answer on the other five. The six items would disagree with each other and alpha would come out noticeably lower, which would wrongly suggest the scale is unreliable. You can go back to the Scales step on the Variables screen and tick the box, and Statversity will flip it.
- **The -99 codes had not been recognized.** A -99 would count as a real, very low answer, and it would distort both the averages and alpha. Statversity flagged them on the "Check how we read them" screen and again on the clean-up screen so that you can check.
- **I wanted to drop unfinished responses.** Tick **Leave out unfinished responses** (8 rows) on the clean-up screen. The Review and import screen would list it, and the number of responses analysed would fall.
- **Alpha had been .65.** That would be below the .70 rule of thumb. The table would show which item has the lowest item-total correlation, and you would have to decide whether to revise the scale, rather than treating it as reliable.
- **I want to try the other four files.** Go back to Practice data and choose a different version, such as **Answers as words instead of numbers**, and watch how Statversity reads it. The clean-up screen will look much the same.

TRY IT YOURSELF

Import the same dataset again, but this time choose **Older export (2 header rows)** instead. On the **Check how we read them** screen, find what Statversity says about the headers, and compare it with the 3-header file. Then, on the clean-up screen, also tick **Leave out unfinished responses** (8 rows) and run Cronbach's alpha again. How many responses does the result say were analysed, and is alpha still in the "good" range?

CHAPTER 14

Matching the same student across two files

This chapter is the companion to chapter 11. There you compared “before” and “after” scores without being able to tell which person was which. Here the two files do share a student ID, so Statversity can match each student’s “before” to their own “after.” You will see how to tell Statversity about the ID, how to read the report it gives you about how well the IDs matched, and why the final result says 73 students when the import said 75.

The question

Did students score higher on an assessment after a course than before it?

This is again a before-and-after study, but now the same students can be followed from one time to the next. That is a stronger design. Instead of asking “is the Post group different from the Pre group?” you can ask “did each student change, and by how much on average?” Comparing each person with themselves removes a lot of noise, because some students simply score high and some score low no matter what, and that difference no longer gets in the way.

The practice data has a realistic problem: the IDs were typed by hand, so they do not match perfectly. Real data often looks like this, and this chapter is mostly about handling it calmly.

The data

The practice dataset is called **Matching the same student across two files**. It has two files:

- `pre.csv`, the assessment before the course. It has 80 rows (80 students).
- `post.csv`, the assessment after the course. Most of the same students took it, but not everyone, and a few new IDs appear.

Each file has two questions we care about:

- **Q1**, “Please enter your unique ID (first two initials + birth day, e.g. AB05).” Students typed this themselves, so the IDs are not guaranteed to line up perfectly between the two files. The first few IDs in the Pre file look like PR23, TY23, BQ23, EW13.
- **Q4**, “Assessment score (0-100).” The first few Pre scores are 69, 82.8, 54.6, 77.7.

Open the practice data

1. Start Statversity. The Home screen shows **Welcome to Statversity**. Click **Open practice data**.

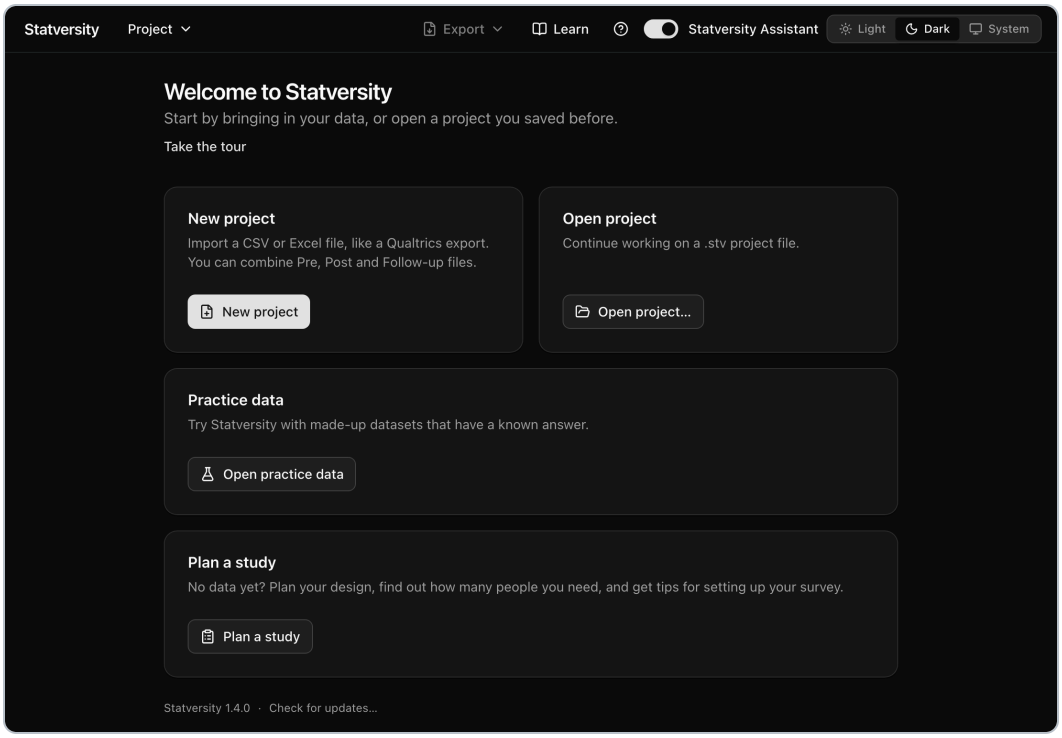


Figure 154. The Home screen with the Open practice data button.

2. On the **Practice data** screen, find the card titled **Matching the same student across two files** and click **Import this dataset** on that card.

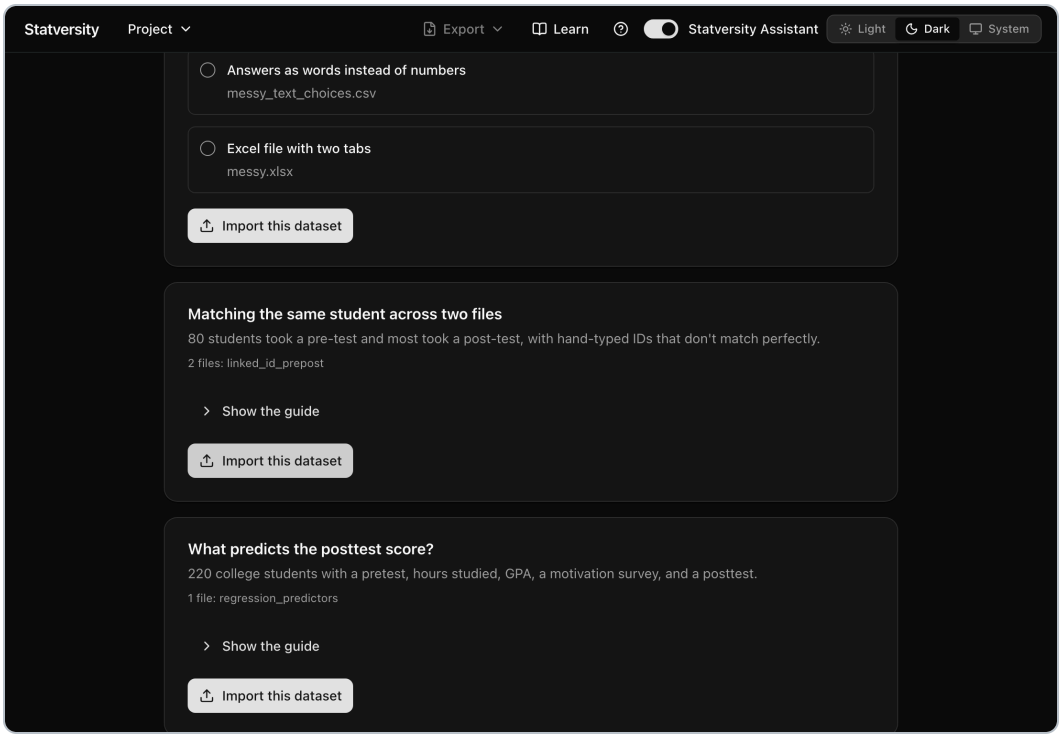


Figure 155. The Practice data screen, with the Matching the same student across two files card.

Steps: bringing the data in

The import is a seven-step wizard. Chapter 5 describes every screen, so we move quickly until the step that matters, **Link people across time**.

3. The screen shows **Choose files** with `pre.csv` and `post.csv` listed. Click **Read files**.

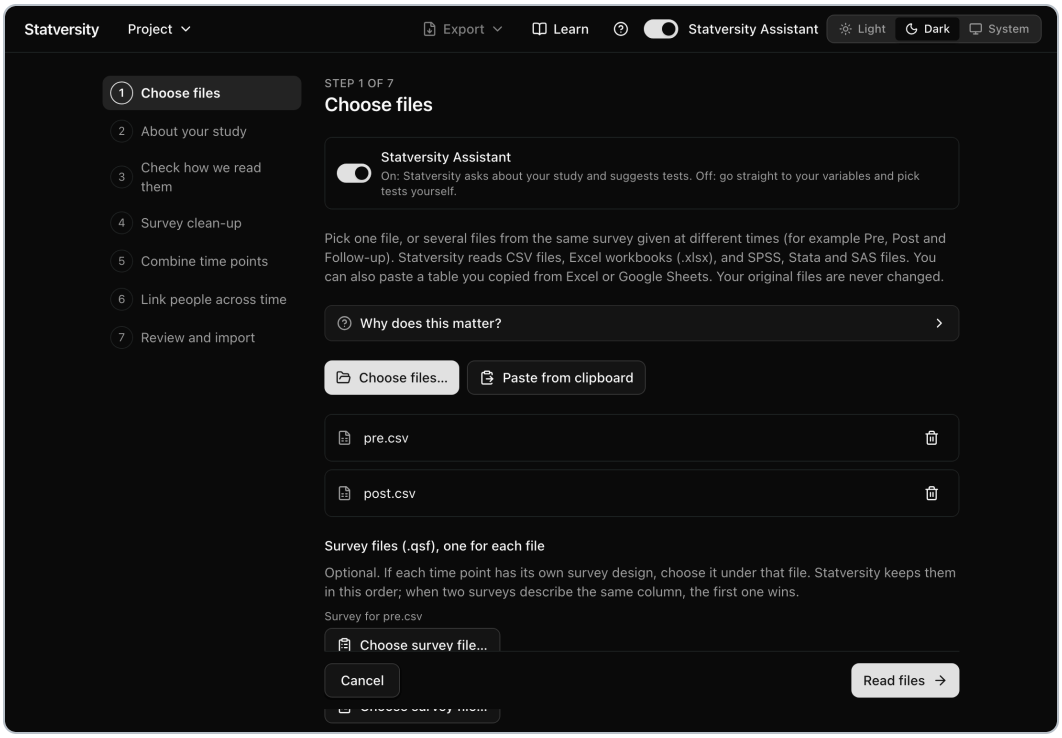


Figure 156. Choose files, with pre.csv and post.csv.

4. The screen shows **About your study**. Click **Skip these questions**.

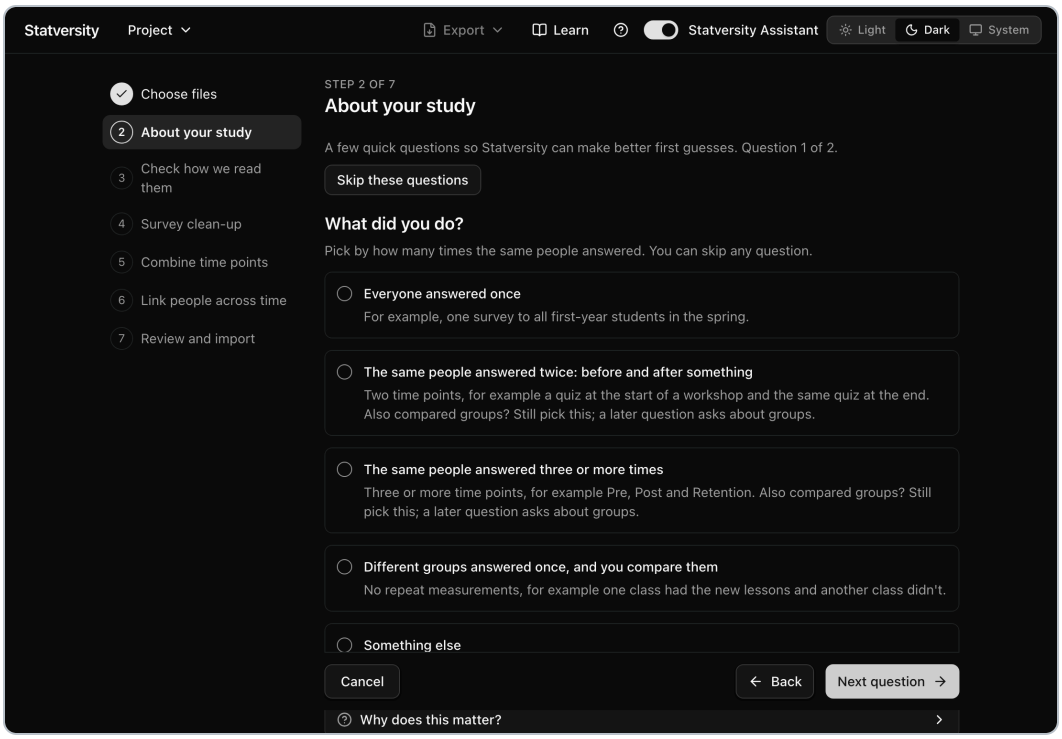


Figure 157. About your study, with the Skip these questions button.

5. The screen shows **Check how we read them**. Statversity reports that `pre.csv` has 80 rows and 19 columns and is a Qualtrics export. The preview of the first answers shows just two columns of real answers, **Q1** (the IDs, such as PR23 and TY23) and **Q4** (the scores, such as 69 and 82.8). Nothing needs changing. Click **Continue**.

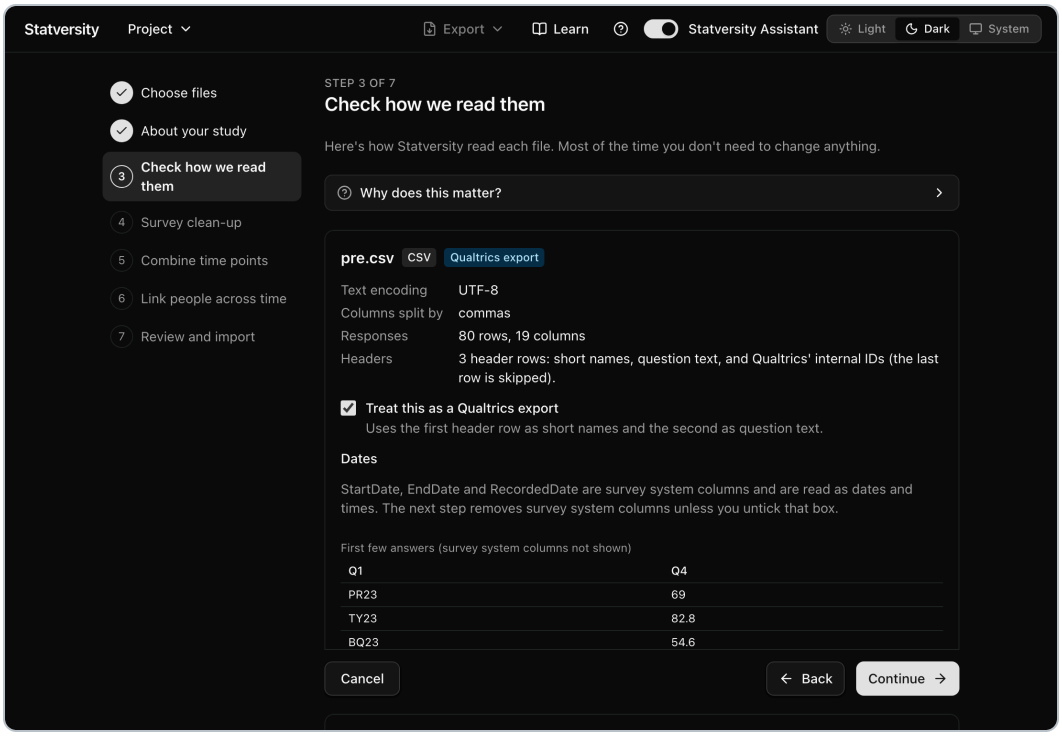


Figure 158. Check how we read them, with the Q1 and Q4 preview.

6. The screen shows **Survey clean-up**. The defaults are right: hide the survey-system columns, and leave out Survey Preview, Survey Test, and Spam responses. There are 0 rows to remove in either file. Click **Continue**.

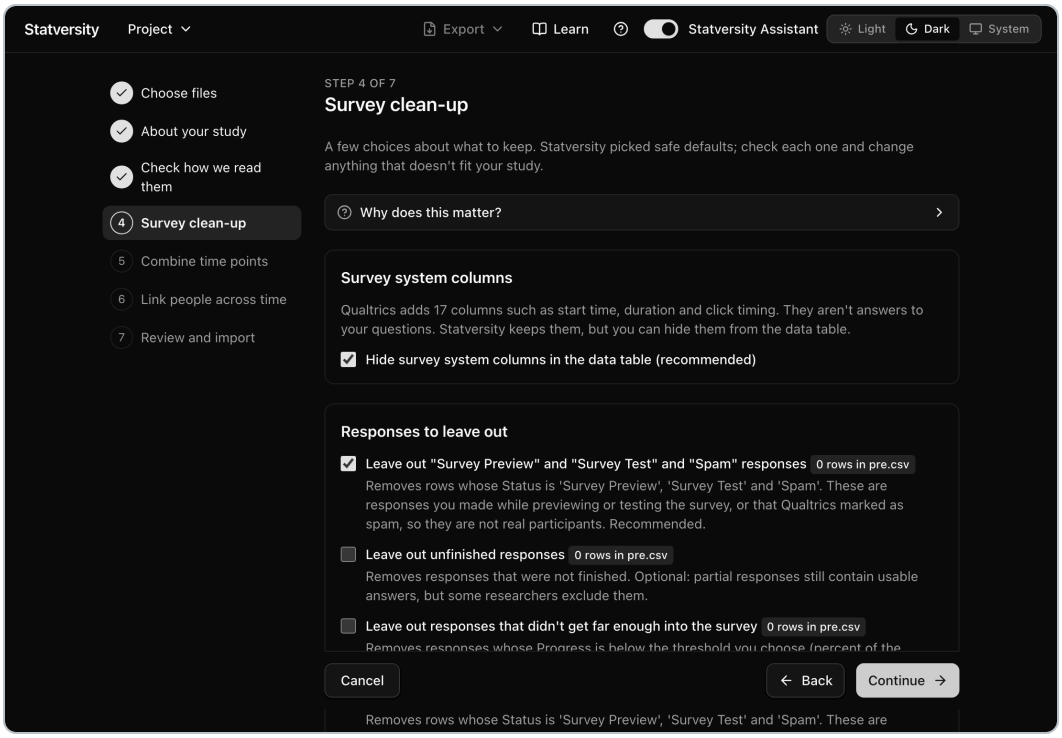


Figure 159. Survey clean-up, with 0 rows to remove in each file.

7. The screen shows **Combine time points**. Statversity will stack the files into one table with a new column called **Time**. The labels read **Pre** for `pre.csv` and **Post** for `post.csv`, and 19 questions matched by their wording. Click **Continue**.

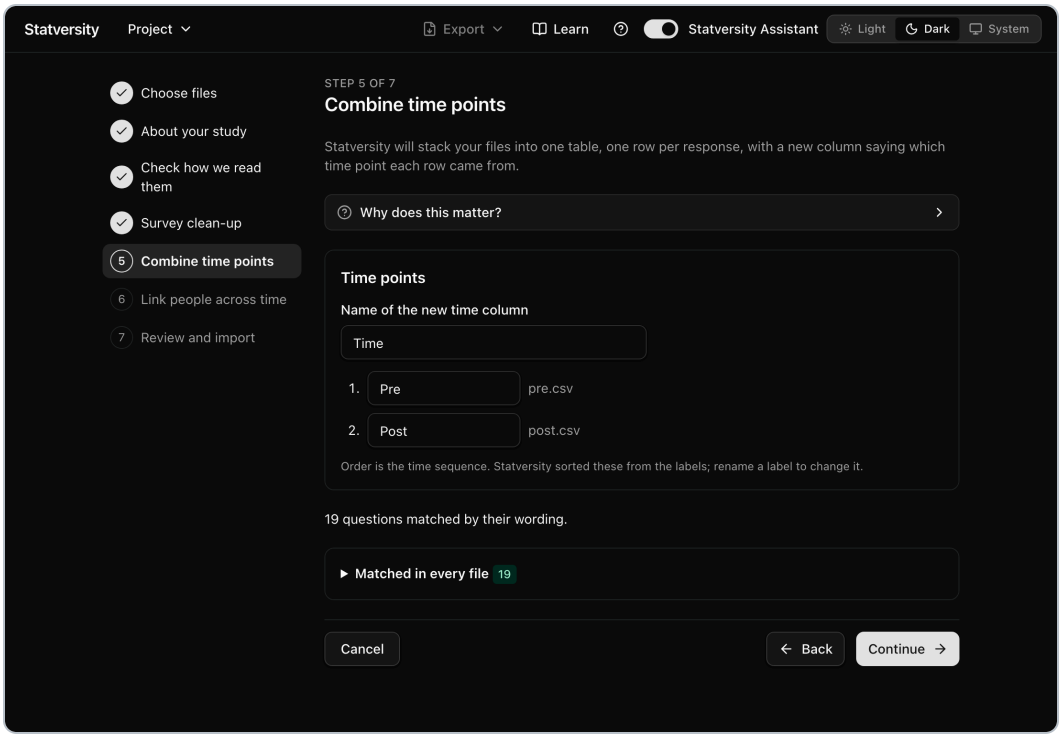


Figure 160. Combine time points, with the labels Pre and Post.

The decision that matters: linking by ID

8. The screen shows **Link people across time**, with **Don’t link (default)** picked. This is the setting you left alone in chapter 11. This time we want the other one.

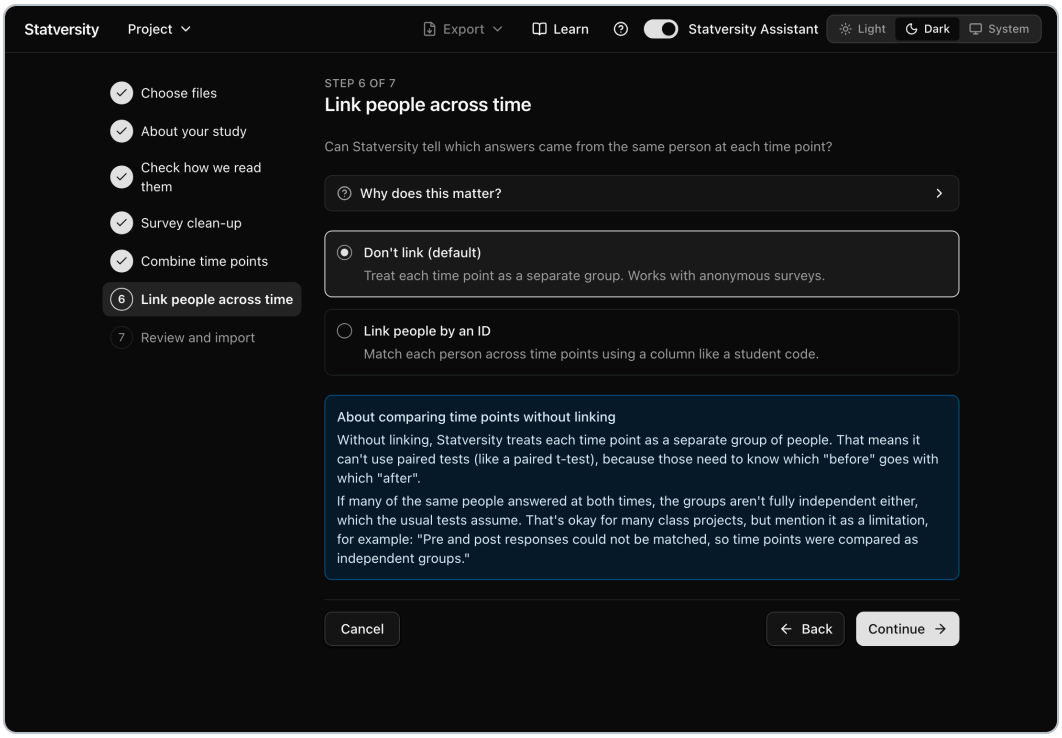


Figure 161. Link people across time, with the default Don't link still picked.

> **IN PLAIN LANGUAGE**
>
> **What linking means.** Linking tells Statversity: "These rows belong to the same person, and here is the column that proves it." Once Statversity knows that, it can line up each student's Pre score with their own Post score. A comparison that uses those pairs is called a **paired** comparison. Without linking, Statversity has to treat the Pre and Post rows as two unrelated crowds.

9. Choose **Link people by an ID** ("Match each person across time points using a column like a student code"). A new box appears asking **Which column identifies each person?** It has an **ID column** dropdown (still on "Choose a column…"), and two ticked boxes, **Ignore extra spaces** and **Ignore capital letters**. The **Continue** button is disabled until you pick a column.

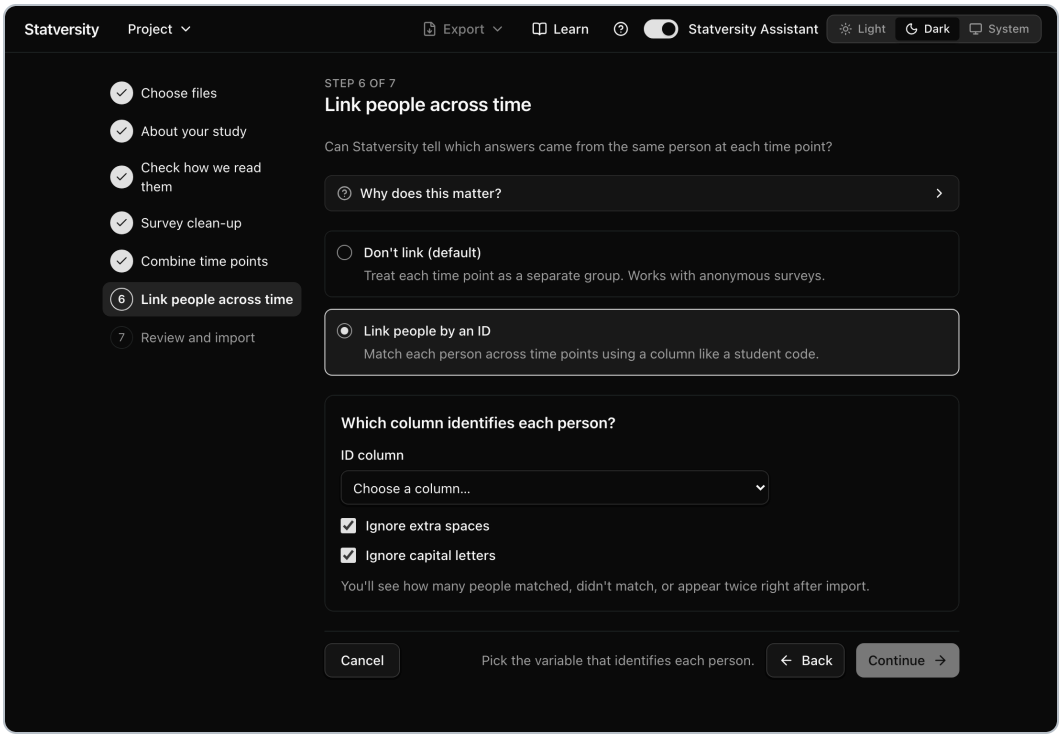


Figure 162. Link people by an ID picked, with the ID column dropdown still empty.

10. In **ID column**, choose **Q1 - Please enter your unique ID (first two initials + birth day,** (the text is cut off on screen, but that is the Q1 column). Leave **Ignore extra spaces** and **Ignore capital letters** ticked.

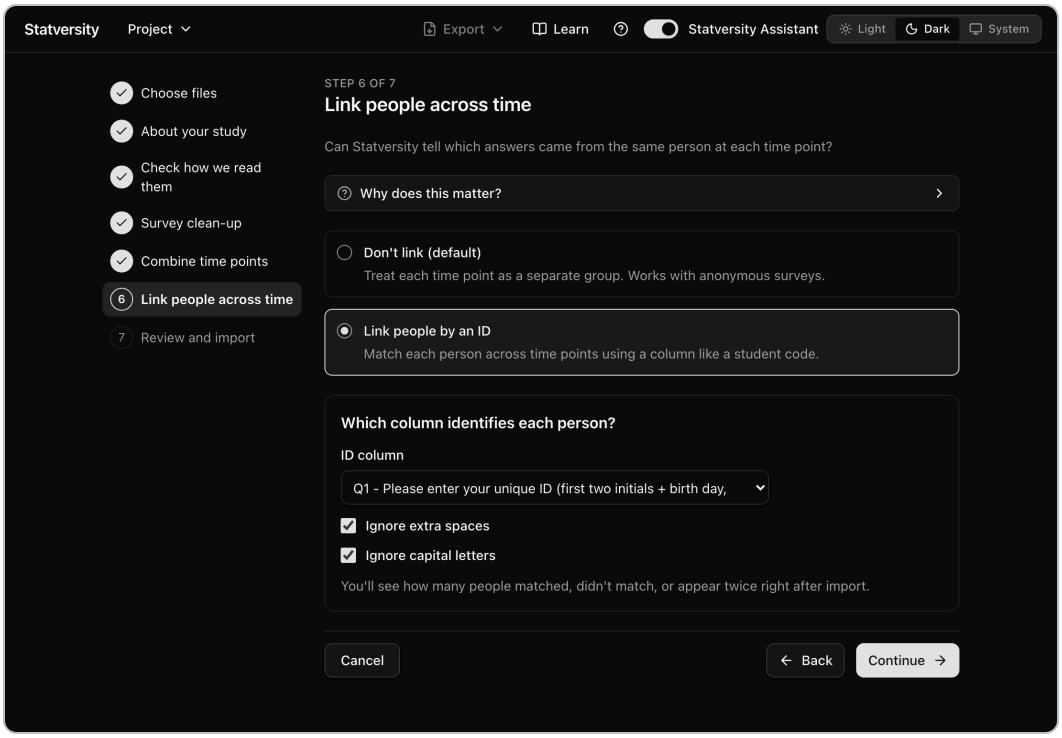


Figure 163. The ID column set to Q1, with both tidy-up boxes ticked.

> **IN PLAIN LANGUAGE**
>
> **Why leave those two boxes ticked?** The IDs were typed by hand. "pr23 " (lowercase, with a trailing space) and “PR23” are obviously the same student to you, but a computer sees them as different text. These two boxes tell Statversity to trim stray spaces and ignore upper and lower case before it compares IDs. They are on by default because they almost never hurt and often rescue several matches. The note on screen reminds you that “You’ll see how many people matched, didn’t match, or appear twice right after import.”

11. Click **Continue**. The screen shows **Review and import**. It lists two files, **161 rows** read, 19 columns kept, and the last line reads **Linking: by Q1**. Click **Import**.

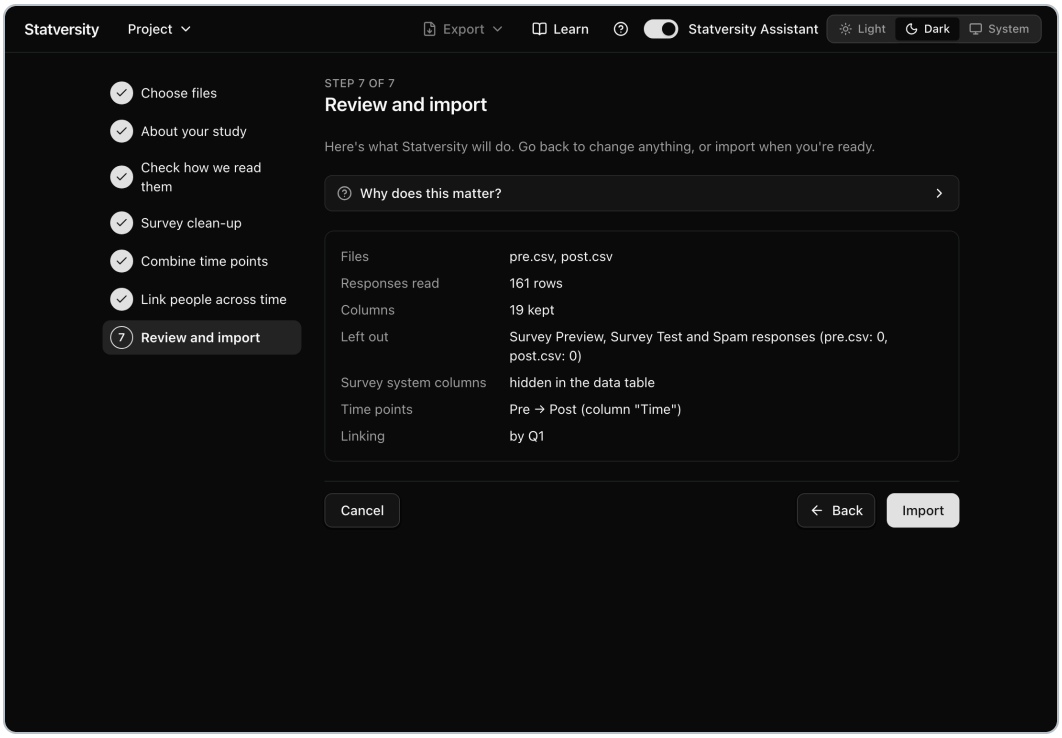


Figure 164. Review and import, showing Linking: by Q1.

Reading the import report

12. The screen shows **Import complete**. It says “Imported 161 responses and 20 variables.” (One more variable than the 19 columns, because Statversity added the **Time** column.) The part to read carefully is the box titled **How the IDs matched**.

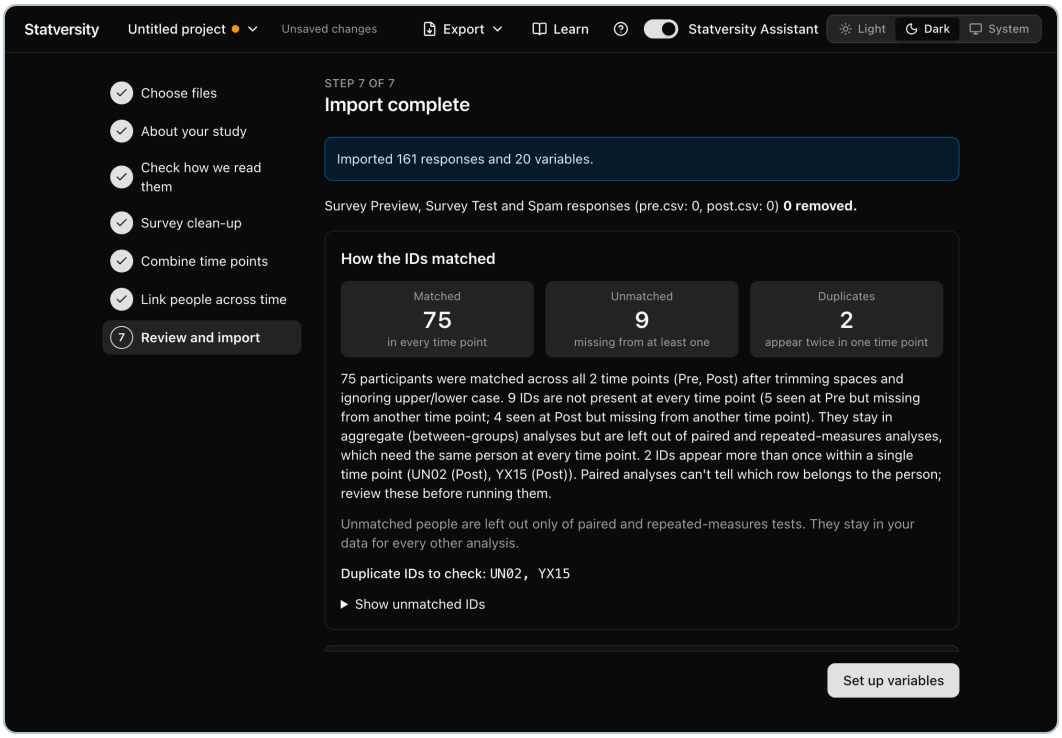


Figure 165. Import complete, with the How the IDs matched box showing Matched 75, Unmatched 9, Duplicates 2.

It has three tiles:

Tile	Number	What it says
Matched	75	in every time point
Unmatched	9	missing from at least one
Duplicates	2	appear twice in one time point

Below the tiles Statversity explains in words: “75 participants were matched across all 2 time points (Pre, Post) after trimming spaces and ignoring upper/lower case. 9 IDs are not present at every time point (5 seen at Pre but missing from another time point; 4 seen at Post but missing from another time point).” It also says these nine “stay in aggregate (between-groups) analyses but are left out of paired and repeated-measures analyses, which need the same person at every time point.”

Then comes the part that explains a number you will see later: “2 IDs appear more than once within a single time point (UN02 (Post), YX15 (Post)). Paired analyses can’t tell which row belongs to the person; review these before running them.” The duplicate IDs are listed again as **Duplicate IDs to check: UN02, YX15**, and a bar at the bottom says “This file has 2 repeated IDs.” with a **Review** button.

IN PLAIN LANGUAGE

Why does 75 turn into 73 later? Think of it as three groups of students.

- **Unmatched (9).** A student who took the Pre but not the Post (5 of them), or the Post but not the Pre (4 of them). A paired test needs both scores, so these nine cannot be used for pairing.
- **Matched (75).** Students found in both files by their ID.
- **Duplicates (2).** Two of the matched IDs, UN02 and YX15, appear twice in the Post file. Maybe one student submitted twice, or two students typed the same ID. Statversity cannot tell which Post row belongs to the Pre score, so it will not guess.

Take the two duplicates out of the 75 and 73 students remain whose Pre and Post scores Statversity can pair with confidence. That is exactly the number you will see in the result. Statversity is being cautious rather than quietly picking one of the two rows.

CAUTION

Check this report on your own data. If almost nobody matched, the problem is usually that you picked the wrong ID column or the IDs follow different formats in the two files. Click **Back** through the wizard and look at the ID column again. A few unmatched IDs, as here, are normal: students miss one session.

13. Click **Set up variables**.

Steps: telling Statversity what each question is

The screen shows **Welcome** for the Variable Interview (chapter 6 covers it). Notice the bar “This file has 2 repeated IDs” with the **Review** button. It follows you into the next screens so you cannot forget it. It also says “2 questions to check,” one for each of Q1 and Q4.

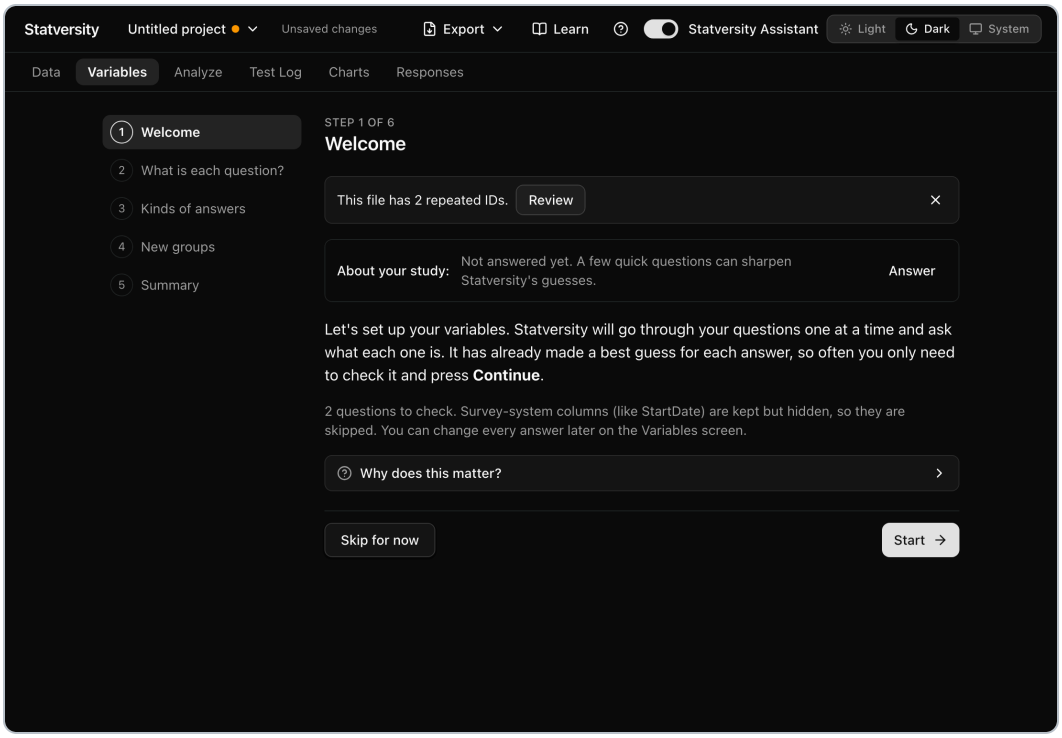


Figure 166. The Welcome step, with the repeated IDs reminder and 2 questions to check.

14. Click **Start**. The screen shows **What is each question?** for **Q1**. Statversity shows the question text (“Please enter your unique ID…”) and some example answers (PR23, TY23, BQ23, EW13…). It guessed **Identifier (ID)**: “A code that tells people apart, like a student ID.” That is correct. Click **Continue**.

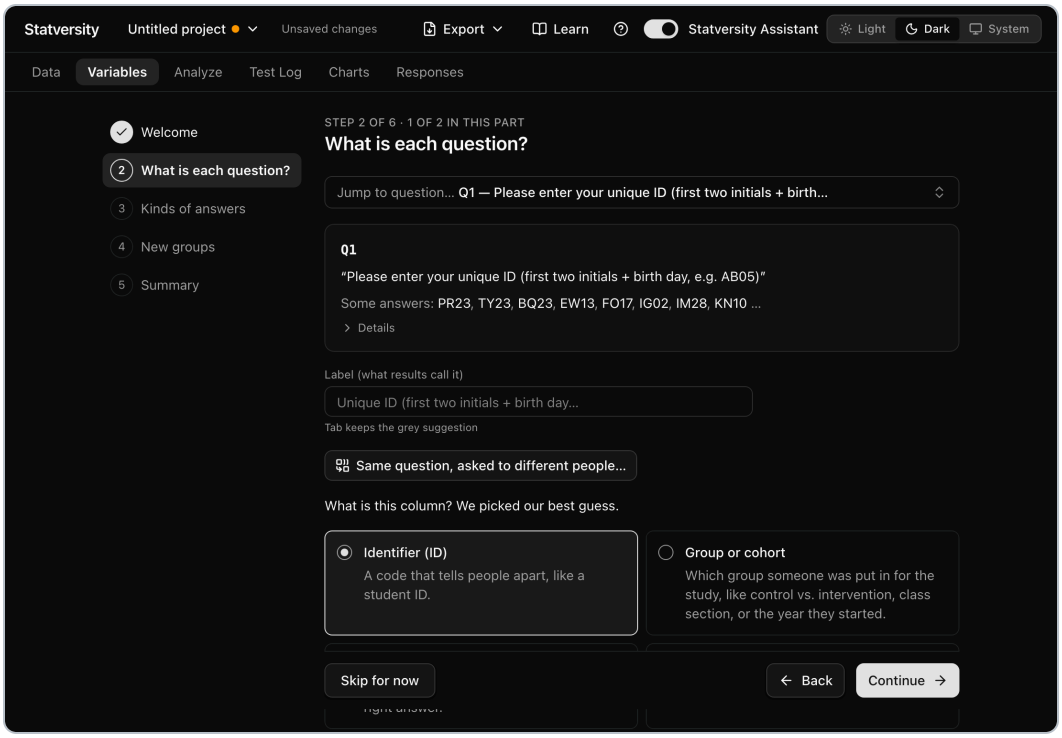


Figure 167. What is each question? for Q1, with Identifier (ID) selected.

15. The screen shows the same step for **Q4**, "Assessment score (0-100)," with sample answers (69, 82.8, 54.6, 77.7…). Statversity guessed **Test total score**: "A score that adds up a whole test." This is a good fit for a 0-to-100 assessment. Click **Continue**.

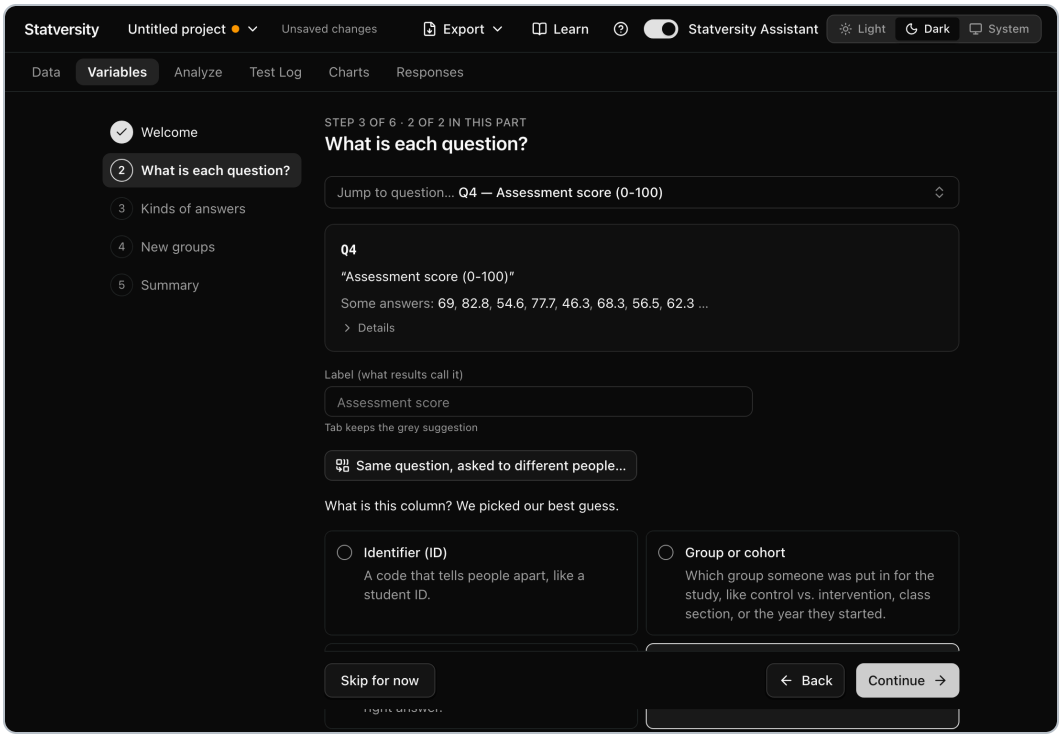


Figure 168. What is each question? for Q4, with Test total score selected.

16. The screen shows **Kinds of answers** for Q4. Statversity picked **Numbers (continuous)**: "Amounts where the steps are equal and averages make sense. Example: test score 0-100." Correct, because the gap between 60 and 70 means the same as the gap between 80 and 90. Click **Continue**.

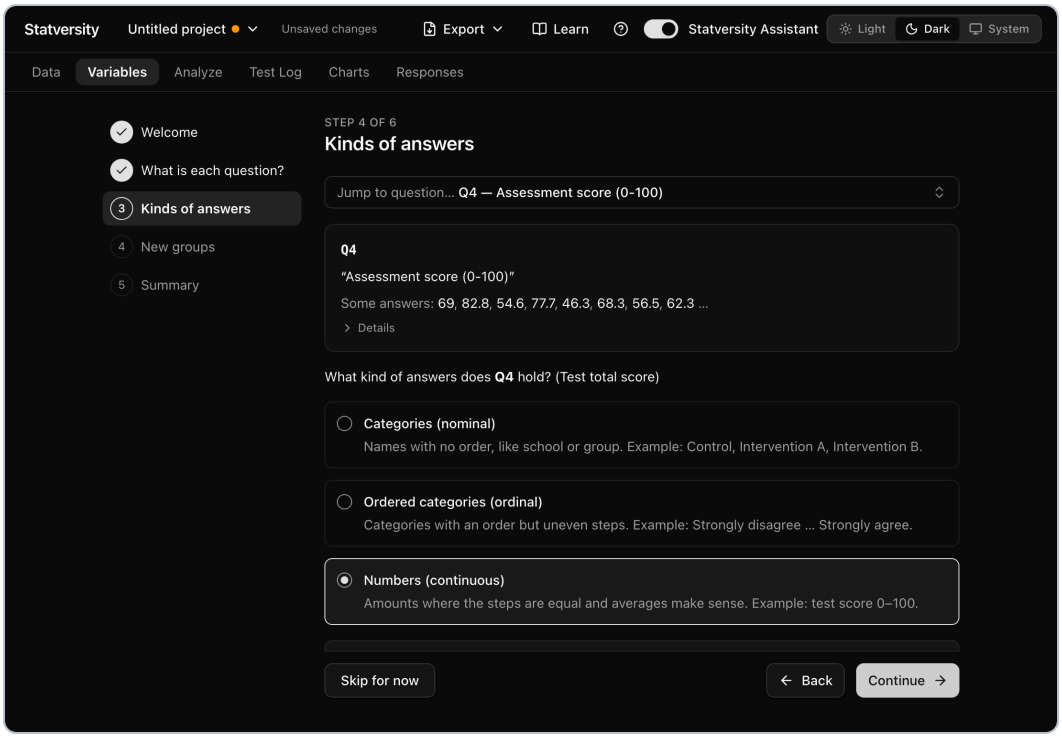


Figure 169. Kinds of answers for Q4, with Numbers (continuous) selected.

> **IN PLAIN LANGUAGE**
>
> **Why this matters for the test.** The kind of answer decides which tests make sense. Numbers where equal steps mean equal amounts allow averages, and averages are what a t test compares. A "strongly agree" rating would have been ordinal and handled with more care. Here there is no scale to build and no reverse-worded items, so the interview is much shorter than in chapter 11.

17. The screen shows **New groups**, an optional step. We do not need one. Click **Continue**.

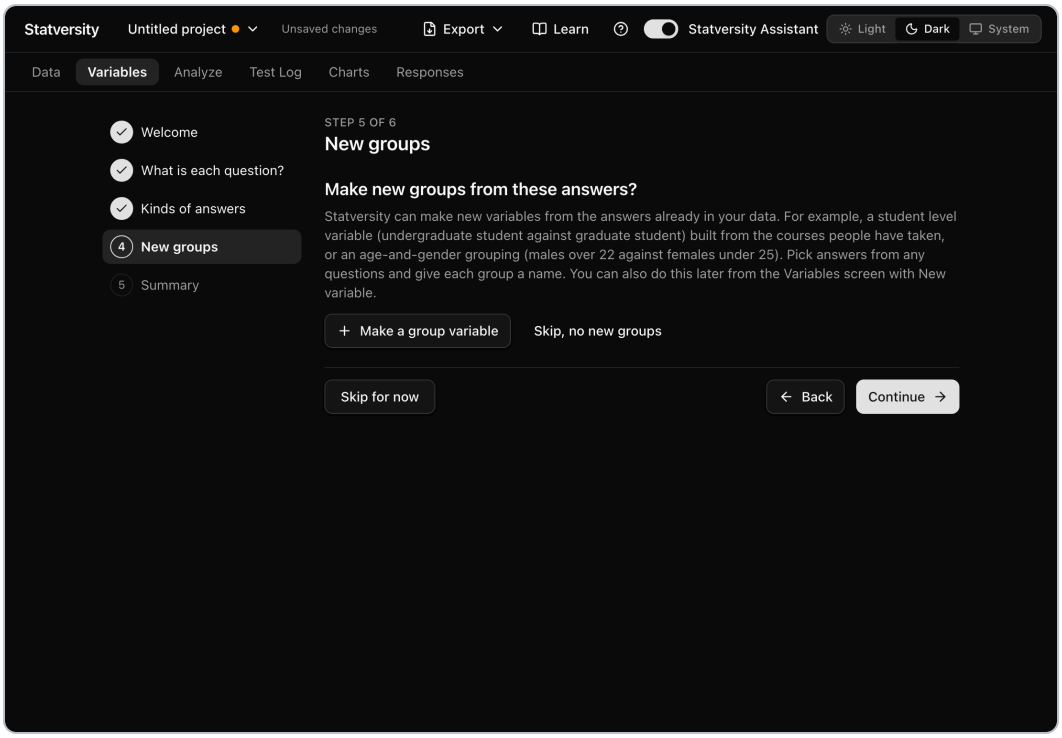


Figure 170. The New groups step.

18. The screen shows **Summary**. It lists Q1 as **Identifier (ID)** and nominal, and Q4 as **Test total score** and continuous. At the bottom: 1 variable to update and 0 scale scores to compute. Click **Finish**.

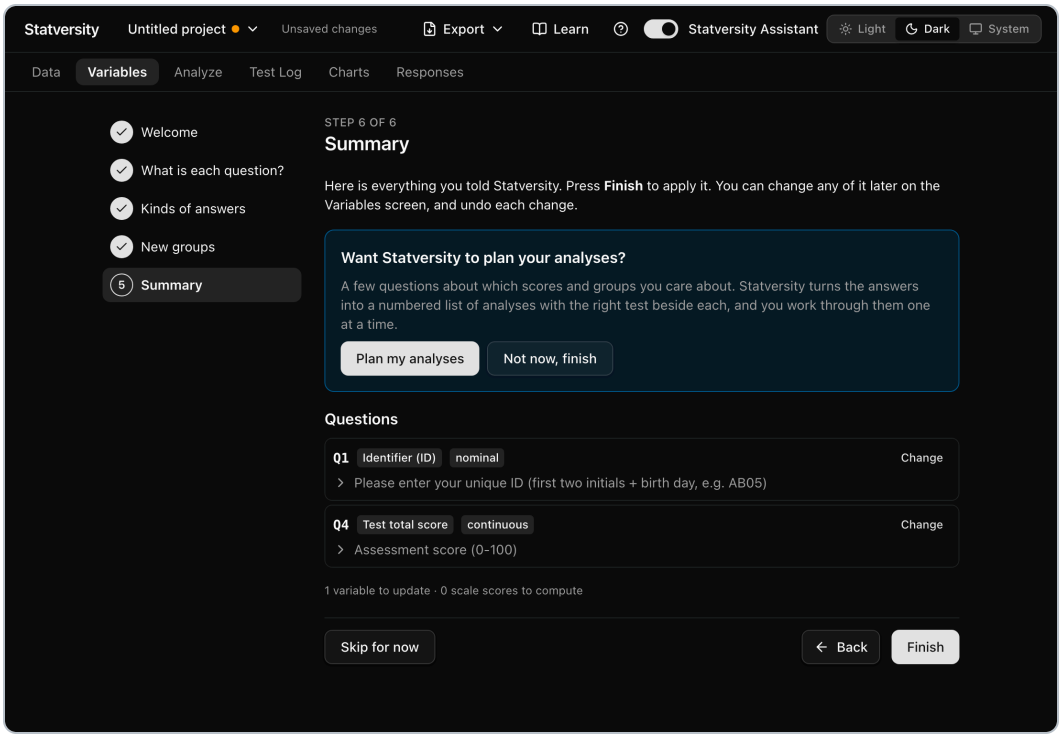


Figure 171. The Summary step.

19. The screen shows **Variables**, with 20 variables. The table has three real rows: **Time** (role **Time point**, two values), **Q1** (role **Identifier (ID)**), and **Q4** (role **Test total score**, level **continuous**). Look at the **Changes** list on the right. It records what you did, including “Linked people by Q1 (75 matched)” and “Set the linking column as Participant ID.” Statversity remembered your linking choice as part of the project.

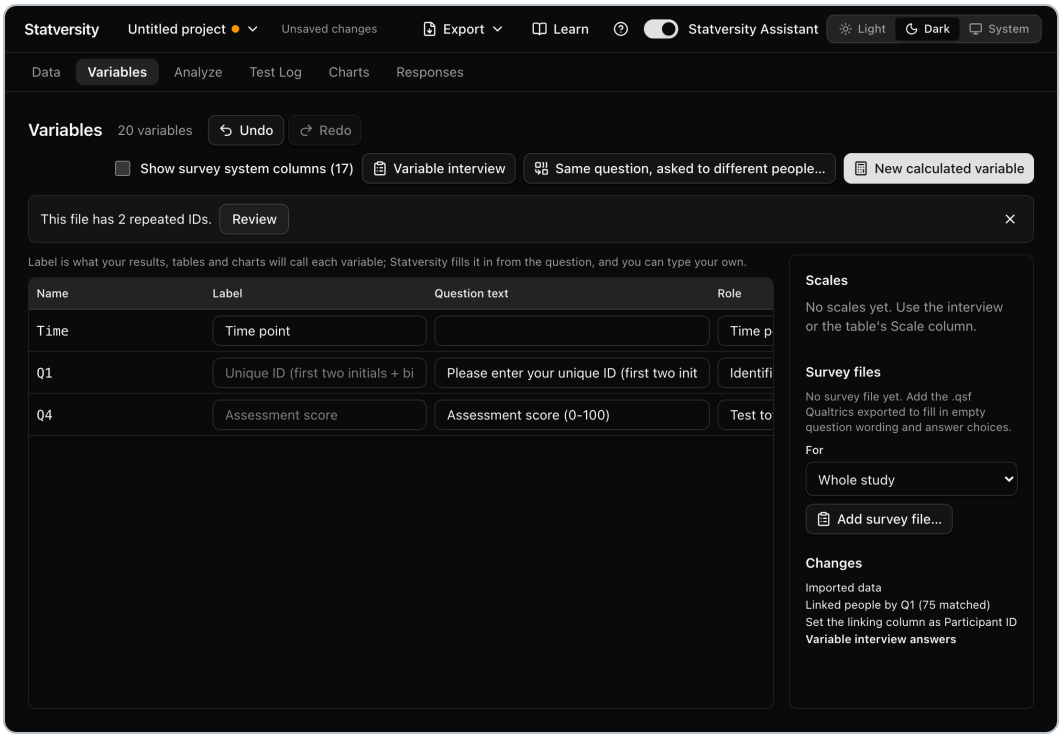


Figure 172. The Variables screen, with the Changes list showing Linked people by Q1 (75 matched).

Steps: asking Statversity which test to use

20. Click **Analyze**. The Test Advisor opens with “What do you want to know?”

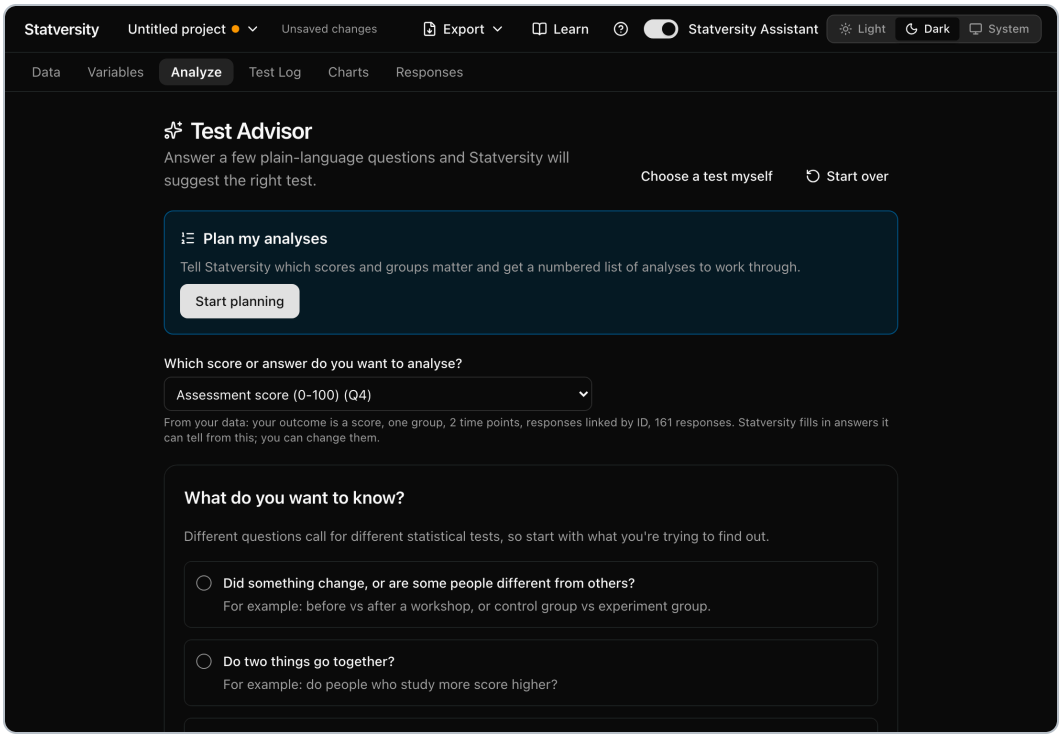


Figure 173. The Test Advisor at the first question.

21. In **Which score or answer do you want to analyse?**, choose **Assessment score (0-100) (Q4)**. Statversity's summary under it now reads: "your outcome is a score, one group, 2 time points, **responses linked by ID**, 161 responses." Compare that with chapter 11, where it said "responses not linked across time." That one phrase is the difference your linking choice made.

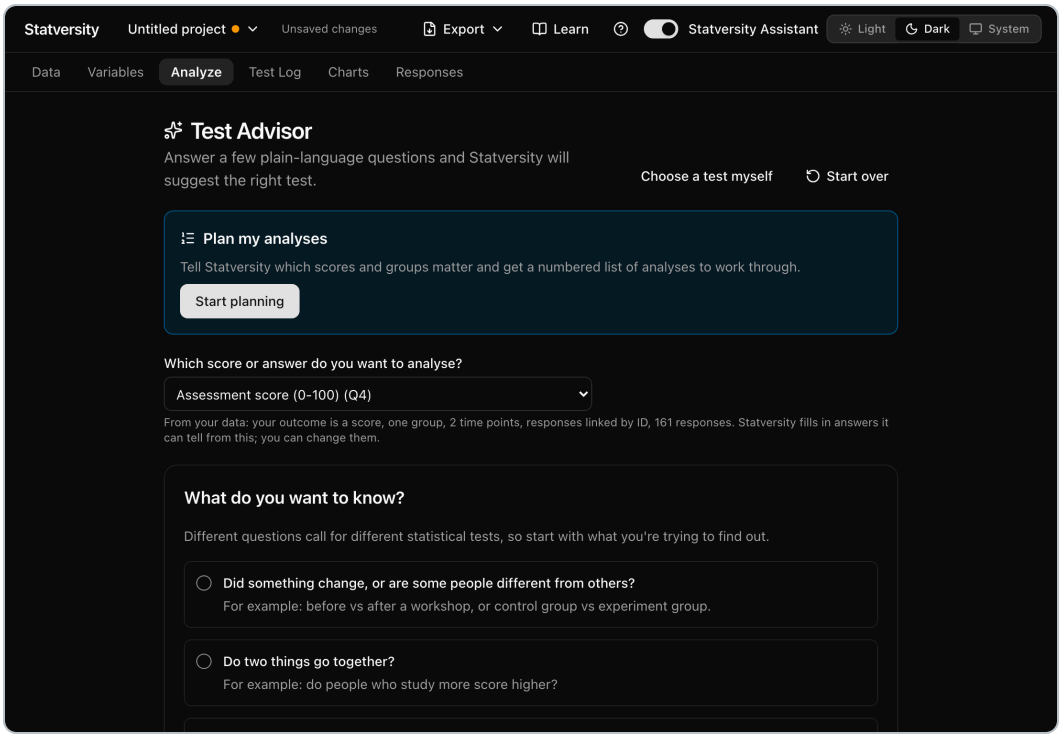


Figure 174. The outcome picker set to Assessment score (0-100) (Q4), with the summary saying responses linked by ID.

22. Under **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**. The next screen is **What are you comparing?**

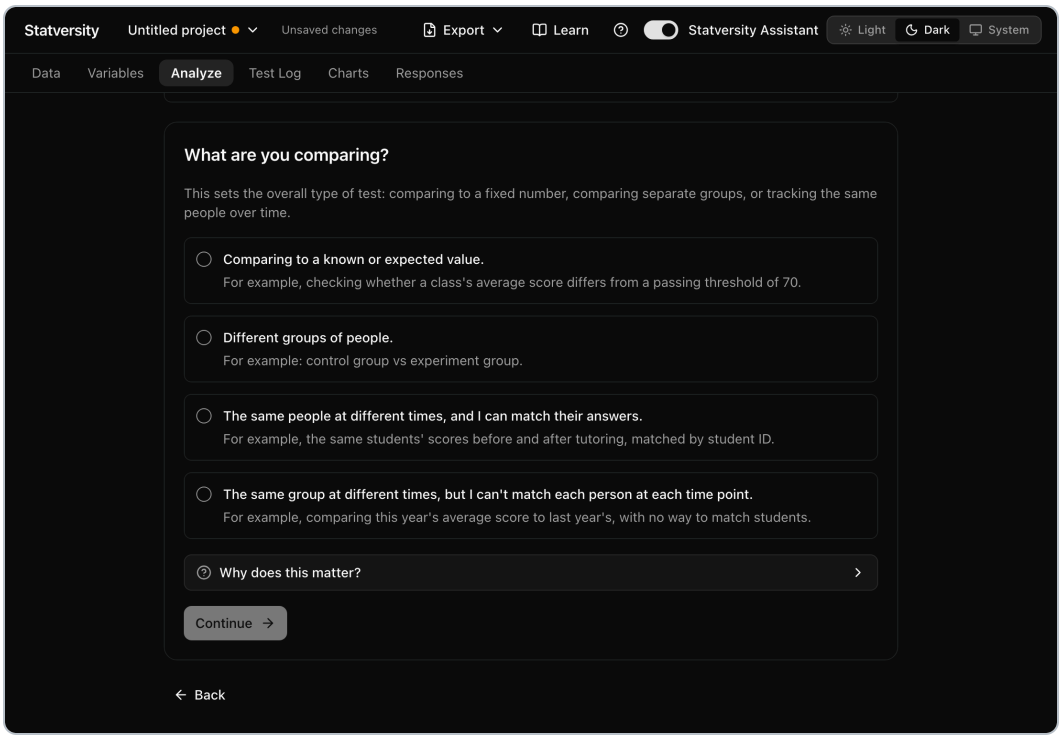


Figure 175. What are you comparing?, with the four options.

23. Choose **The same people at different times, and I can match their answers.** (“For example, the same students’ scores before and after tutoring, matched by student ID.”) Click **Continue**.

> **IN PLAIN LANGUAGE**
>
> **This is the answer chapter 11 could not give.** You are telling Statversity you have the same people at two times and a way to match them. That is true now, because the ID column links the rows. If you picked the other “same group at different times” answer, you would be telling Statversity it cannot match people, which would discard the benefit of everything you did in the import.

24. The Advisor shows **Statversity recommends: Paired-samples t test**. There is no orange warning this time. Statversity lists the backup test (**Wilcoxon signed-rank test**), the effect sizes it will report (**Cohen’s d (z)** and **Cohen’s d (av)**), and one assumption to check: **Normality of the differences**, meaning the before/after change scores roughly follow a bell-shaped pattern.

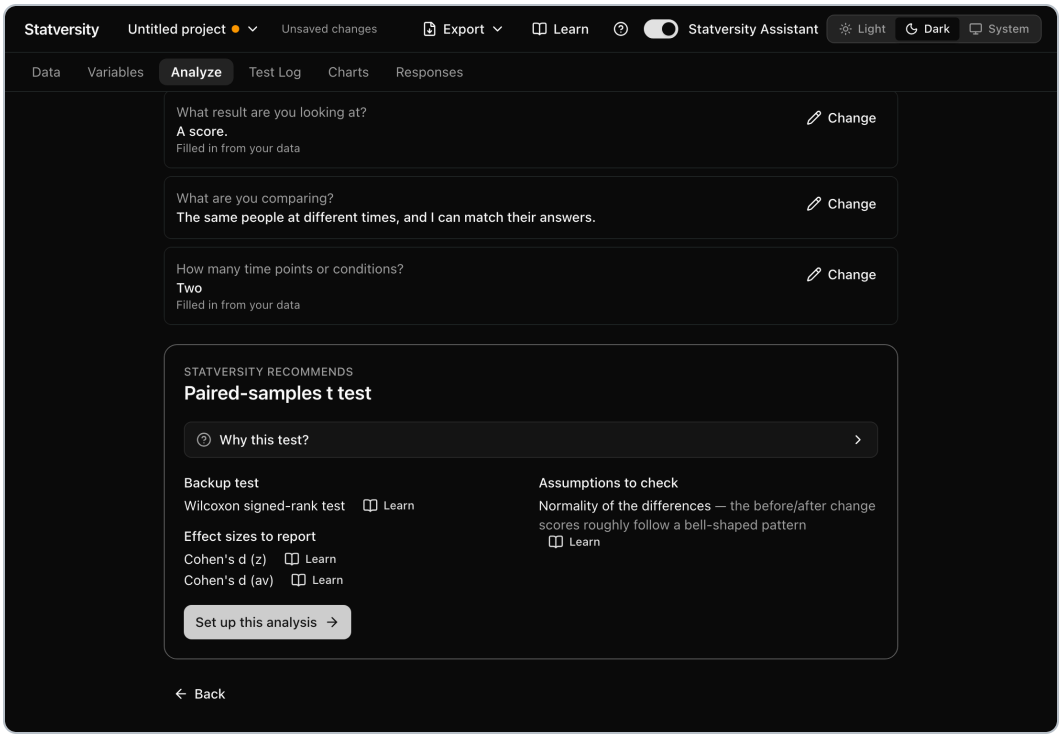


Figure 176. The recommendation, Paired-samples t test, with the backup test and the one assumption to check.

> **IN PLAIN LANGUAGE**
>
> **Paired-samples t test, in one paragraph.** For each matched student, Statversity subtracts their Post score from their Pre score to get one "change" number. A paired t test then asks whether the average of those change numbers is far enough from zero that luck alone would rarely produce it. Because each student is compared only with themselves, a student who always scores high and a student who always scores low contribute equal, clean information. That is why the paired test is usually more sensitive than the independent test in chapter 11.

25. Click **Set up this analysis**.

Steps: choosing variables and checking the assumption

26. The screen shows **Choose your variables**. Read each part.

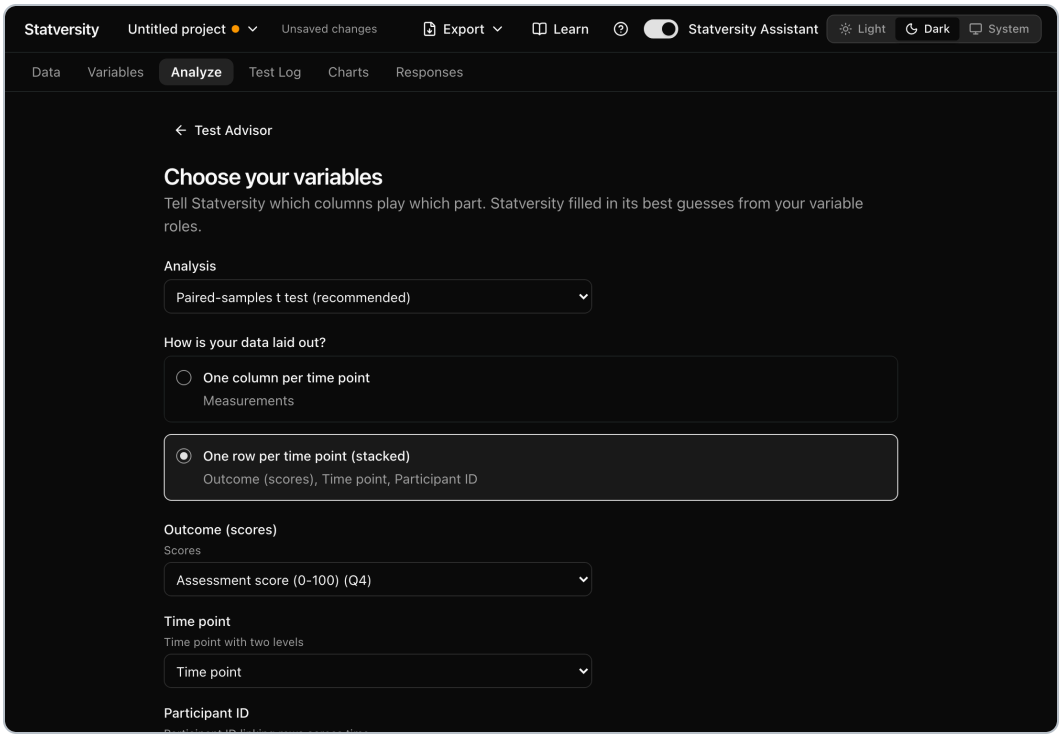


Figure 177. Choose your variables, with the stacked data layout, Time point, and Participant ID filled in.

- **Analysis** shows **Paired-samples t test (recommended)**.
- **How is your data laid out?** offers **One column per time point** and **One row per time point (stacked)**. Statversity picked the second, "Outcome (scores), Time point, Participant ID," because the Combine time points step stacked your files into rows with a Time column.
- **Outcome (scores)** shows **Assessment score (0-100) (Q4)**.
- **Time point** shows **Time point** ("Time point with two levels").

- **Participant ID** shows the Q1 question text. It says “Participant ID linking rows across time.” This is the ID column you chose in the import.

Statversity filled all of these in, so there is nothing to change.

27. Click **Check the assumptions**. There is just one check for this test.

Check 1 of 1: normality of the differences

The check is titled **Normality: Pre - Post**. Notice that Statversity checks the *differences* between each person’s two scores, not the Pre scores and Post scores separately. This is because the paired t test works from the change numbers. Statversity took the 73 differences, ran the Shapiro-Wilk test, and found W = 0.98, p = .533. The result is **Looks fine**: no clear sign that the differences are far from a bell-shaped curve.

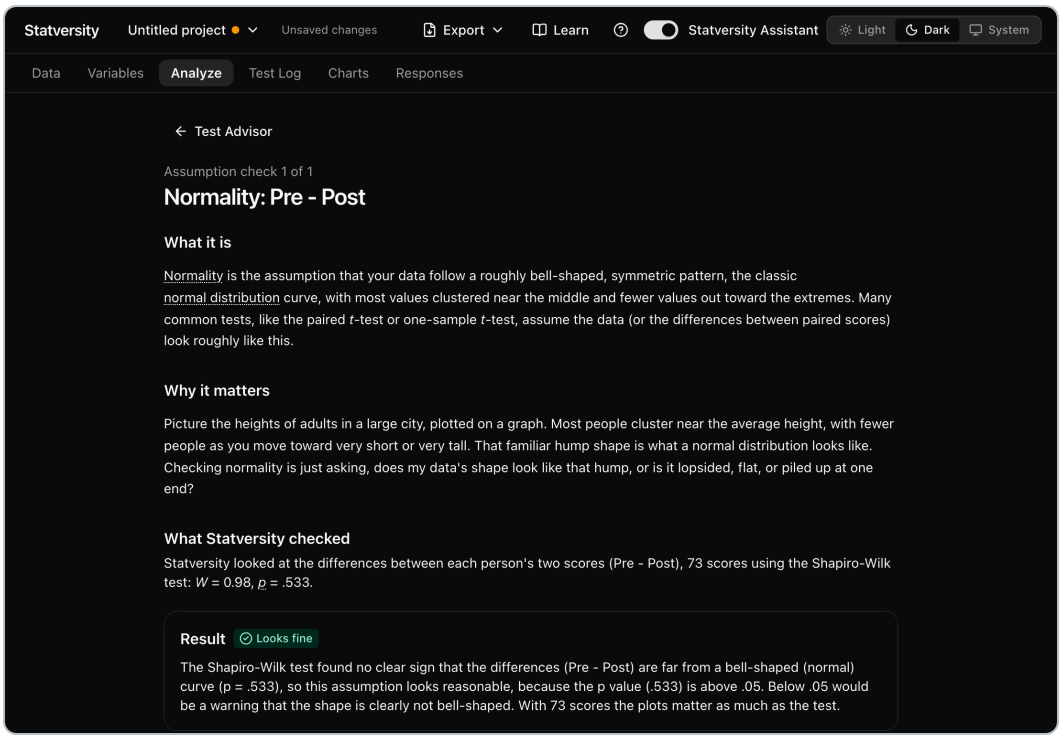


Figure 178. Normality of the Pre - Post differences, Looks fine with p = .533.

> IN PLAIN LANGUAGE
>
> **Why no “equal spread” check this time?** In chapter 11, Statversity checked that the two groups were spread out by similar amounts, because the independent test compares two separate piles. A paired test has only one pile, the change scores, so there is nothing to compare. The count of 73 here is also the first sign that Statversity used only the complete pairs: 75 matched IDs, less the two duplicate IDs.

28. Click **Decide which test to use**. The screen shows **Which test should you use?** with the one check (**Looks fine**) and the message “Statversity suggests: Paired-samples t test. Every check looks reasonable, so the recommended test is a good fit.” The other choice is the **Wilcoxon signed-rank test**, the rank-based alternative that does not need bell-shaped differences. This matches Statversity’s suggestion.

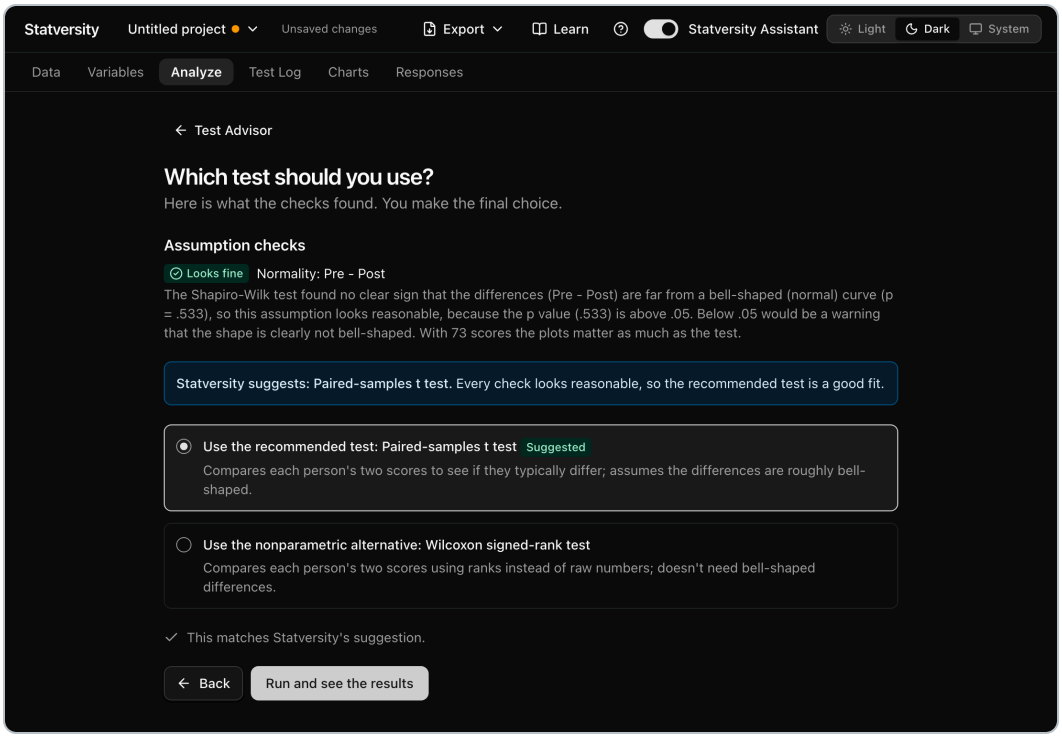


Figure 179. Which test should you use?, with the paired t test suggested.

29. Click **Run and see the results**.

Reading the results

The Results screen opens for the **Paired-samples t test**. Read the line under the title: “73 responses analysed, 15 left out (9 had a score at only one time point; 2 a duplicate ID at the same time point) (Pre: 73; Post: 73).”

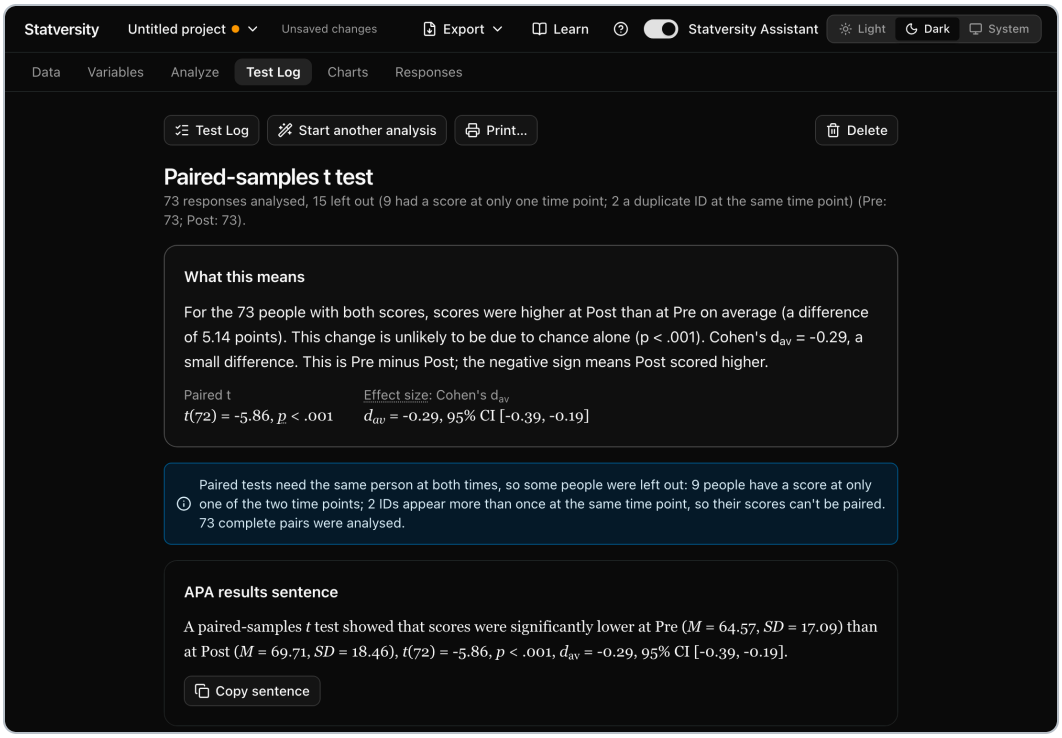


Figure 180. The top of the results, with the 15 left out explanation, What this means, and a blue notice.

Statversity states the exclusions openly instead of hiding them. Directly under the **What this means** box you will find a blue notice: “Paired tests need the same person at both times, so some people were left out: 9 people have a score at only one of the two time points; 2 IDs appear more than once at the same time point, so their scores can’t be paired. 73 complete pairs were analysed.”

> **NOTE**
>
> **Why “15 left out” and “2” both appear.** The count of 15 is a count of left-out responses (rows), while the import report counted IDs (people): 9 unmatched and 2 duplicated. The two counts describe the same exclusions in different units, which is why they are not the same number. What stays the same is the useful number: 73 complete pairs were analysed, 73 at Pre and 73 at Post.

30. Read **What this means** at the top of the result.

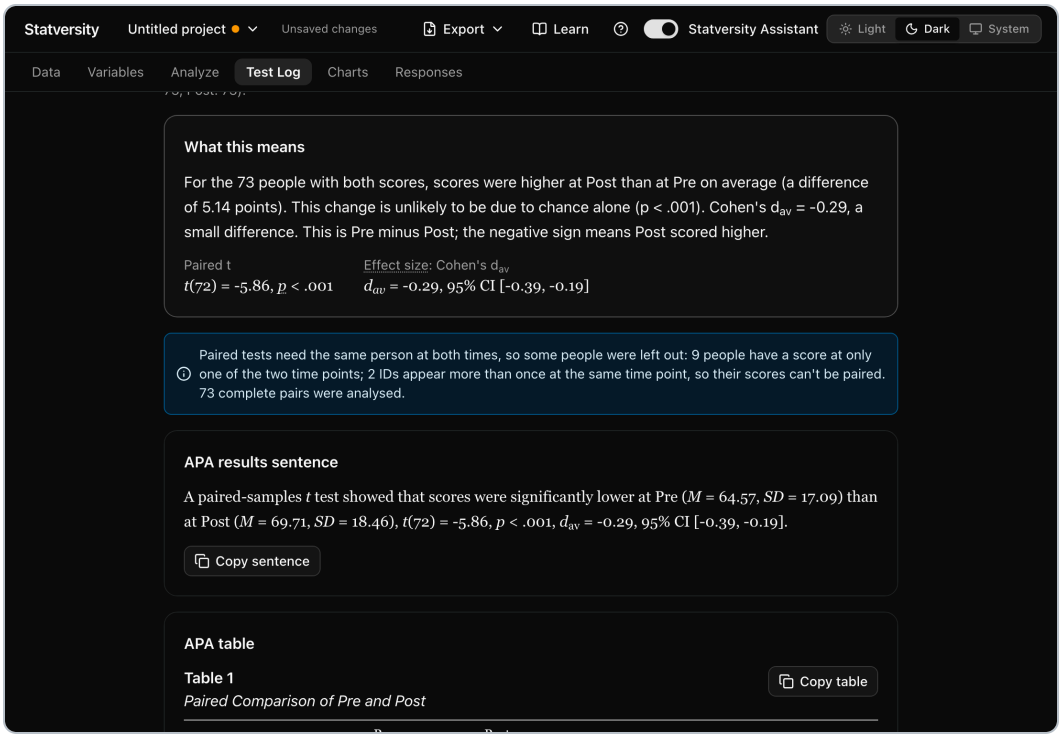


Figure 181. The What this means box with the paired t statistic and Cohen's d_av.

31. Scroll down to **APA results sentence** and **APA table**.

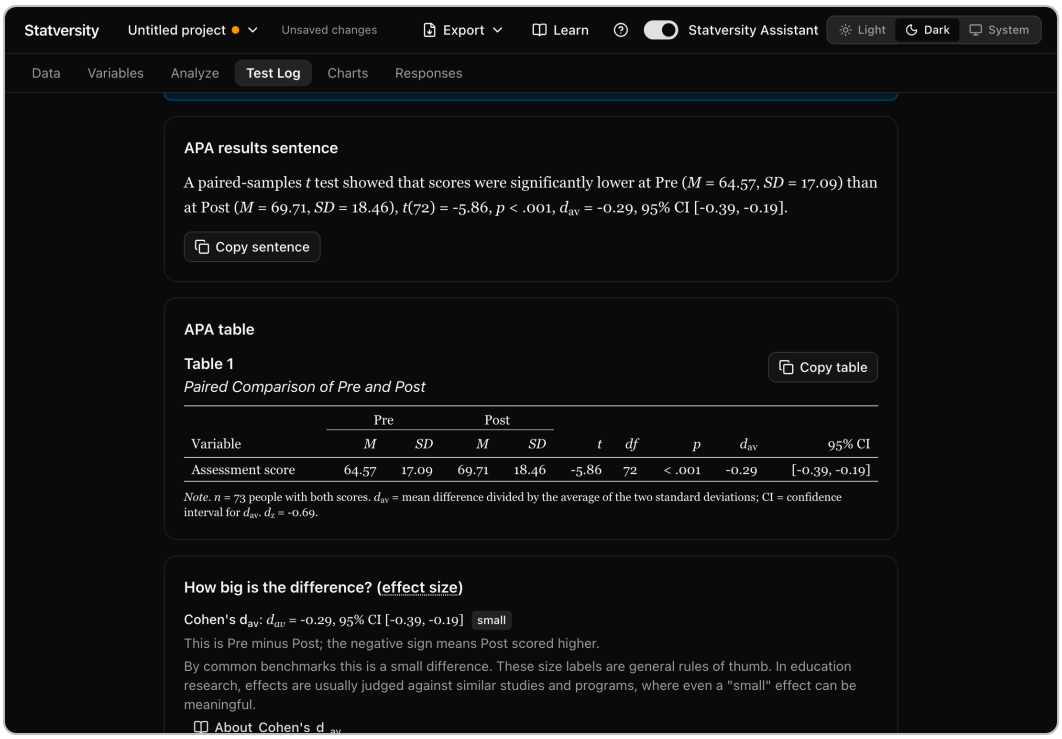


Figure 182. The APA results sentence and APA table.

Just below the table, Statversity shows **How big is the difference? (effect size)** with two measures, **Cohen's d(av)** and **Cohen's d(z)**. We come back to why there are two.

32. Scroll down to the chart. The first chart is a Q-Q plot of the Pre - Post differences, next to a histogram of the same numbers. Below them is **Descriptive statistics** for each time point.



Figure 183. The Q-Q plot and histogram of the differences, and the Descriptive statistics table.

33. Click **Test Log**. The entry reads **Paired-samples t test** for Q4, with n = 73 and p < .001.

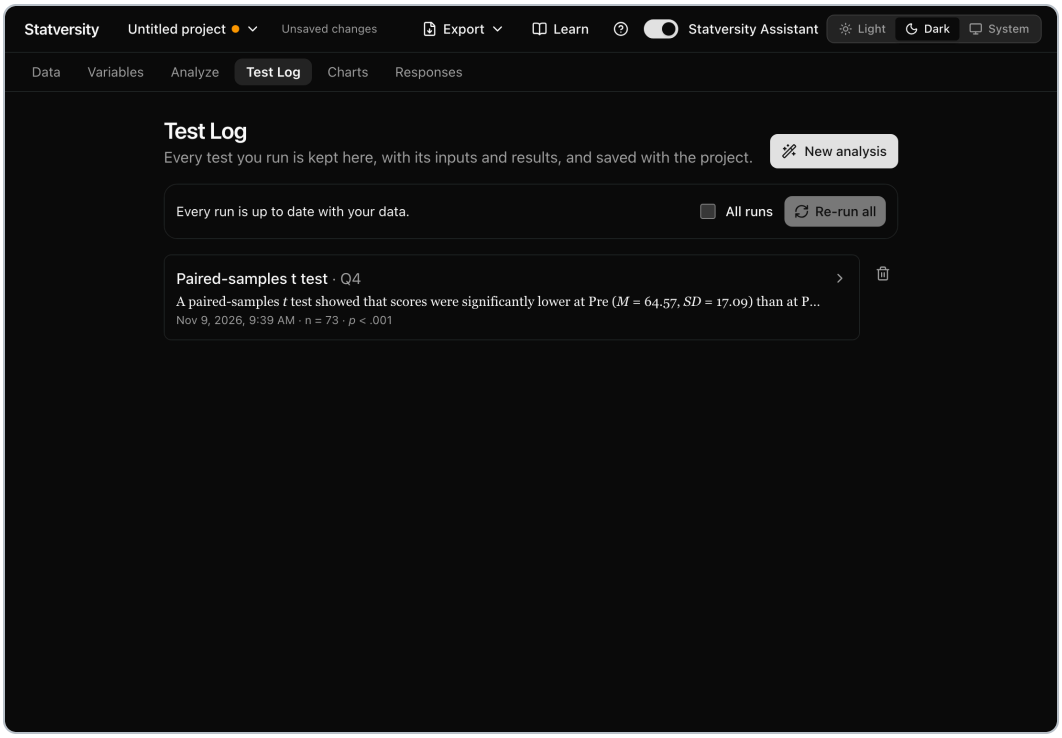


Figure 184. The Test Log with the paired analysis listed.

What Statversity found

	Pre	Post
People (n)	73	73
Mean (M)	64.57	69.71
Standard deviation (SD)	17.09	18.46
Median	64.10	68.10

Result	Value
Test	Paired-samples t test
Pairs analysed	73 (15 responses left out)
Average change	5.14 points higher at Post
Statistic	t(72) = -5.86
p-value	< .001
Effect size (headline)	Cohen’s d(av) = -0.29, 95% CI [-0.39, -0.19], small
Effect size (also shown)	Cohen’s d(z) = -0.69, 95% CI [-0.94, -0.43], medium

The degrees of freedom, 72, are the number of pairs minus one (73 - 1).

Two effect sizes, two labels

The two effect sizes disagree about the size label: small versus medium. They are not in conflict. They answer slightly different questions.

- **d(av)** divides the average change (5.14 points) by the *typical spread of the scores*, the average of the Pre SD (17.09) and the Post SD (18.46). It answers: "How big is the change compared with how different students are from one another?" This is the one the table and the APA sentence use, and it is the easiest to compare with studies that used independent groups.
- **d(z)** divides the average change by the spread of the *change scores themselves*. When students move up by similar amounts, that spread is small, so d(z) becomes larger. It answers: "How consistent is the change from student to student?"

Both are correct. Statversity's table note says dz = -0.69 for readers who want it. Report the one your field expects, say which one it is, and avoid picking the bigger label because you like it better.

What this means in plain English

For the 73 students for whom Statversity had both a Pre and a Post score, the average assessment score rose from 64.57 to 69.71, a gain of 5.14 points on a 0-to-100 scale. A gain this consistent would be very unlikely if the course made no difference ($p < .001$). Compared with how much students differ from one another, the gain is small ($d(av) = -0.29$). Compared with how much the students' own changes vary, it is a clear and fairly steady improvement ($d(z) = -0.69$).

Put simply: on average students did better after the course, the improvement is very unlikely to be a fluke, and it is a modest, not dramatic, number of points. Whether five points matters is a judgment about your course, not something the statistics can decide. Chapter 23 says more about "significant" versus "important."

> **CAUTION**
>
> **A word of care.** This result does not prove the course caused the gain. Students might have improved from practice or from simply retaking a similar assessment. A design with a comparison group (chapter 12) can speak to that.

What you would write in a paper

Statversity's sentence, ready to copy:

> A paired-samples *t* test showed that scores were significantly lower at Pre ($M = 64.57$, $SD = 17.09$) than at Post ($M = 69.71$, $SD = 18.46$), $t(72) = -5.86$, $p < .001$, $d_{av} = -0.29$, 95% CI [-0.39, -0.19].

To adapt it, replace "scores" with the name of your measure ("assessment scores") and add the sample facts a reader needs: "Of 80 students who took the pre-test, 73 had a single, matchable post-test and were included in the analysis; nine had a score at only one time point and two IDs appeared more than once and were excluded." Those numbers come straight from the **How the IDs matched** box and the results notice. Saying so openly is exactly what reviewers want to see.

What if

What if you had not linked the IDs? Statversity would have treated the Pre and Post responses as two independent groups and recommended an independent-samples t test, as in chapter 11. It would use every row, with no pairs thrown out, but it would be less powerful and could not talk about individual students.

What if almost nobody matched? The **How the IDs matched** box would show a small Matched number and a large Unmatched number. Go **Back** to **Link people across time** and check that you chose the right ID column. If the IDs really do follow different formats in the two files, fix the IDs in your original spreadsheet (Statversity never changes your original files) and import again.

What if the normality check had failed? Statversity would not show **Looks fine** for the differences, and you would choose the **Wilcoxon signed-rank test** on the **Which test should you use?** screen. It does the same job as the paired t test using ranks of the changes, and does not require bell-shaped differences.

What if the duplicate IDs had been errors you could fix? If you opened your Post file and found that the two UN02 rows were really UN02 and UN20, correcting the typo would let UN02 pair up again (and the UN20 row too, if that student has a Pre score). Fixing both duplicate IDs this way would bring the count back up to at least 75 pairs. Use the **Review** button on the duplicate notice to see the IDs Statversity flagged.

What if the gain had been tiny? With a mean change near zero, t would be small and p large, and the summary would say the change could easily be due to chance. You would report it the same way, and the confidence interval would include zero.

> **TRY IT YOURSELF**
>
> **Try it yourself.** Go back to the **Link people across time** step and untick **Ignore capital letters**, then import again. Does the **Matched** number on the **How the IDs matched** box go up, go down, or stay the same? Write a sentence explaining what that tells you about how the students typed their IDs.

CHAPTER 15

Does my survey measure what I think it does?

The question

“I wrote 15 survey questions meant to measure three different things about study skills. Do the answers really sort into those three topics? And separately, is my 20-question quiz reliable?”

These are questions about the **instrument** (the survey or test itself), not about the people who answered it. Researchers call this *validation*. Before you compare groups or track change with a survey, you want some evidence that the survey behaves the way you intended. This chapter runs two checks:

1. **Factor analysis** on the 15 survey items, asking whether they cluster into three topics.
2. **Reliability (KR-20)** on the 20 right/wrong quiz items, asking whether the quiz hangs together as one test.

> **NOTE**
> Chapters 5 to 8 explain the screens in detail (getting data in, variables, analyze, results). This chapter tells you what to click and why each answer is right for this study.

Two ideas in plain language

> IN PLAIN LANGUAGE
>
> **Factor analysis.** Imagine you give people 15 statements about study habits. People who agree with one statement about, say, planning ahead will also tend to agree with the other planning statements, because the same underlying habit drives all of those answers. Factor analysis looks at which questions tend to rise and fall *together* across all the respondents and groups them. Each group is called a **factor**: an underlying topic that you cannot measure directly but that shows up in clusters of answers. If you wrote five questions for each of three topics and factor analysis finds three tidy clusters of five, that is evidence that your questions are measuring what you meant them to.
>
> The word **exploratory** means you are asking the data to show you the clusters, without telling it in advance which question belongs where. (The alternative, *confirmatory* analysis, tests a structure you already specified.)
>
> **Reliability.** A bathroom scale is reliable if it gives you the same weight every time you step on it. A test is reliable if its questions agree with each other: students who know the material tend to get most of the questions right, and students who do not tend to miss most. We measure this with a number between 0 and 1, where higher means the questions agree more. For a test where every item is right or wrong, the standard number is called **KR-20** (Kuder-Richardson Formula 20). It is the right/wrong cousin of the better-known Cronbach's alpha.

The data

The practice dataset is one file, `survey.csv`, with **320 respondents**. Each person answered:

- **Q4**, a block of 15 agreement statements about study skills, rated 1 to 5. Five statements were written for each of three topics, and two of them (items 3 and 9) are worded in the opposite direction on purpose, a common trick in survey design.
- **Q5**, a 20-question quiz already marked right (1) or wrong (0) for each question.
- **SC0**, the quiz total, labelled "Total correct (0-20)".

To open it, go to Home, click **Open practice data**, and find the card titled **Does my survey measure what I think it does?**.

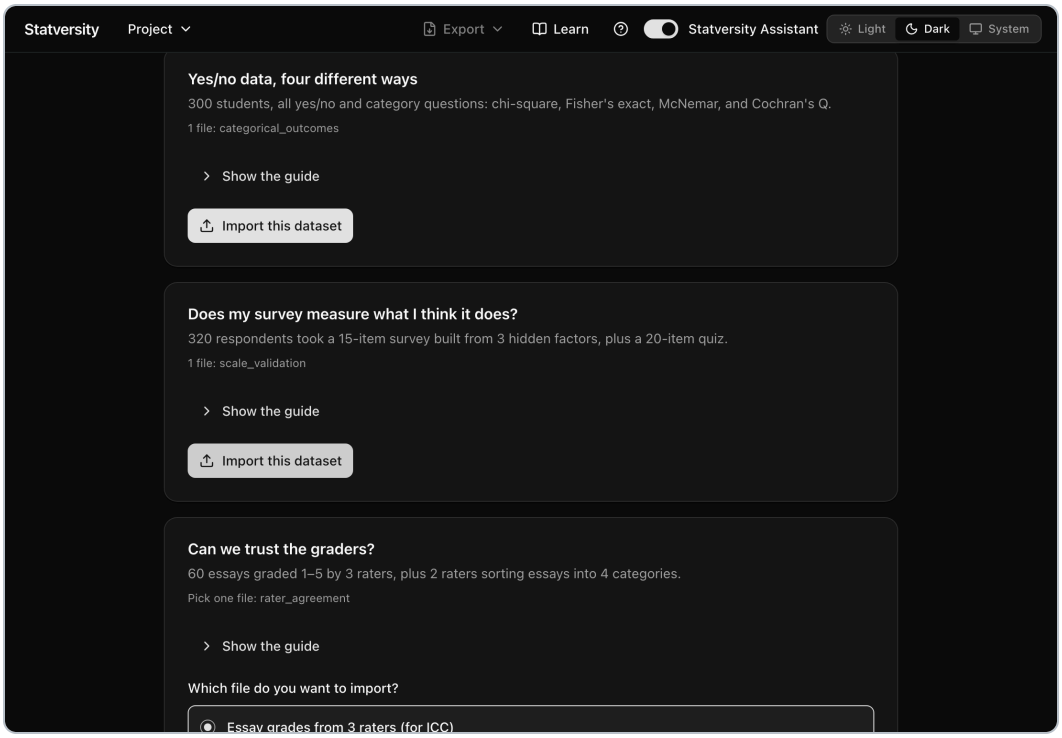


Figure 185. The Practice data screen with the card Does my survey measure what I think it does? in view. It describes 320 respondents who took a 15-item survey and a 20-item quiz.

Steps

Part 1: Import

1. On Home, click **Open practice data**. The **Practice data** screen lists the datasets.
2. On the card **Does my survey measure what I think it does?**, click **Import this dataset**. The **Choose files** screen opens with `survey.csv` already listed.

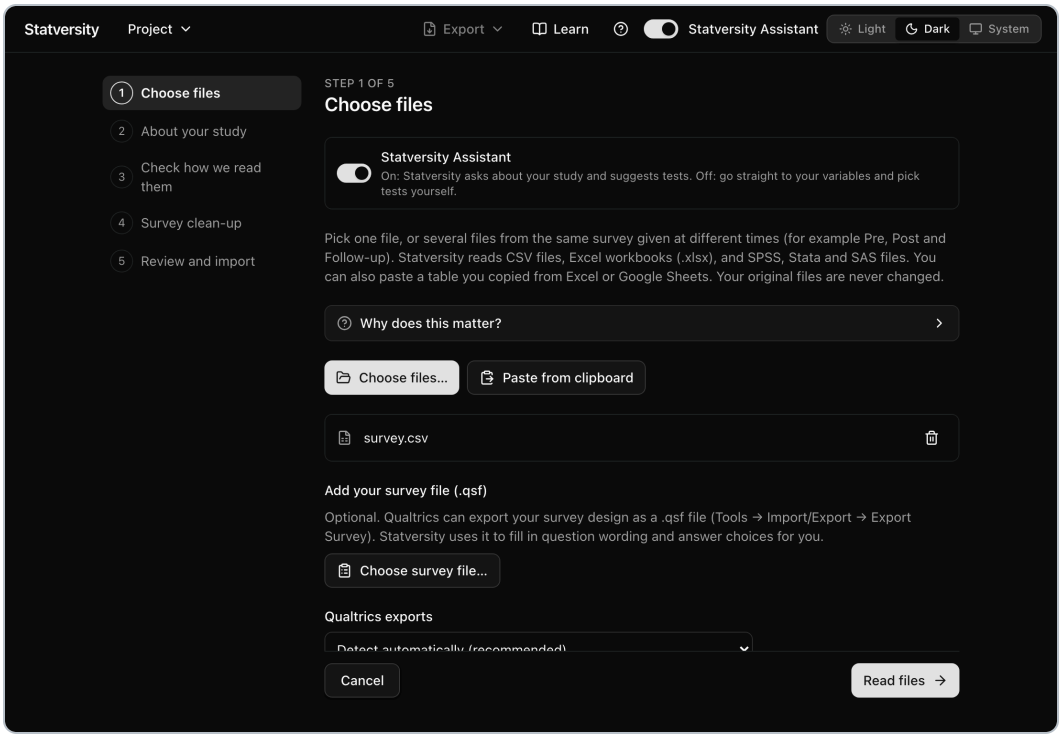


Figure 186. Choose files, with survey.csv listed and a Read files button.

3. Click **Read files**. The **About your study** screen appears. Click **Skip these questions**. (Those questions help Statversity guess better, but this walkthrough does not need them.)
4. The **Check how we read them** screen shows how Statversity read the file: `survey.csv` as a comma-separated CSV recognized as a **Qualtrics export**, 320 rows and 53 columns, with three header rows (short names, question text and internal IDs). Notice the blue notes. Statversity spotted that Q4_1 to Q4_15 look like one matrix question and will suggest treating them as a scale named Q4, and it reminds you that the Q4 answers are stored as numbers 1 to 5 and the Q5 answers as 0 and 1. Nothing needs changing. Click **Continue**.

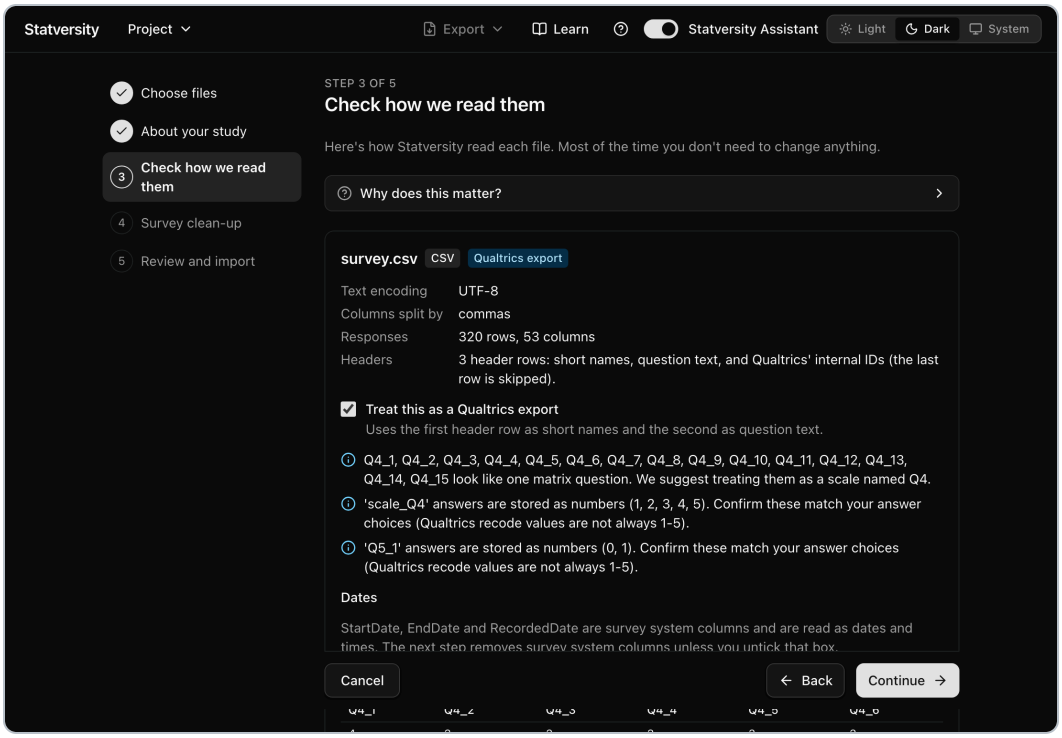


Figure 187. Check how we read them. The blue notes say Statversity will treat Q4_1 to Q4_15 as a scale named Q4.

5. On **Survey clean-up**, the defaults hide the Qualtrics system columns and leave out preview, test and spam responses. The badges show 0 rows would be removed, so the data are already clean. Click **Continue**.

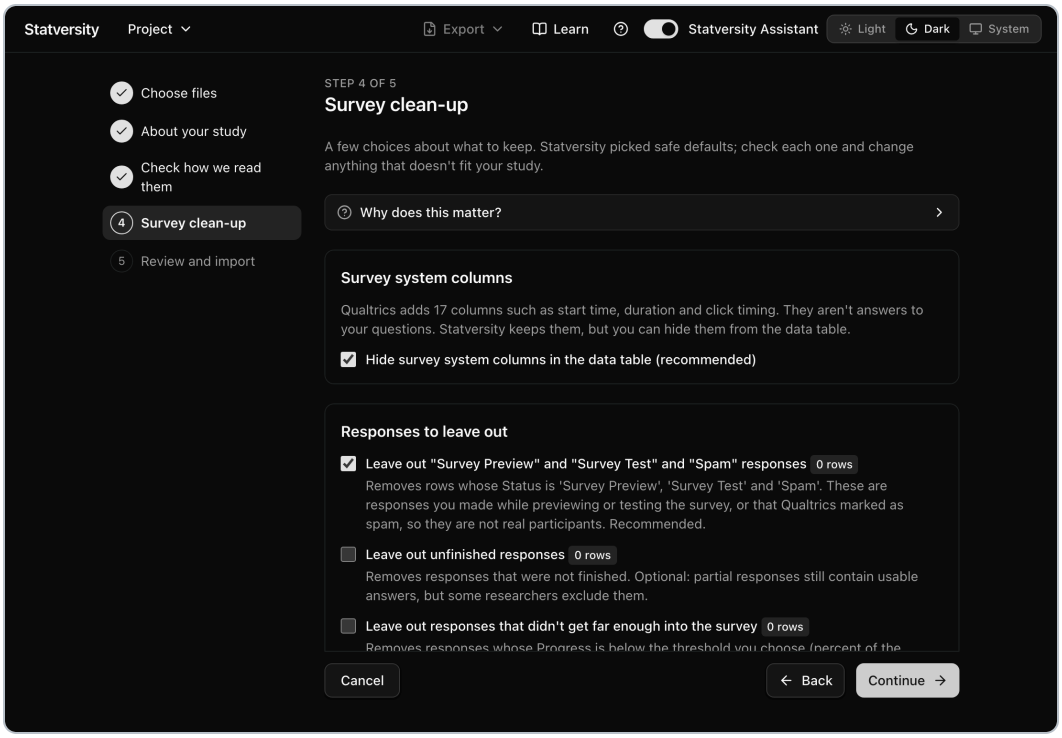


Figure 188. Survey clean-up. Every filter shows 0 rows affected.

6. **Review and import** lists: Files survey.csv; Responses read 320 rows; Columns 53 kept. Click **Import**.

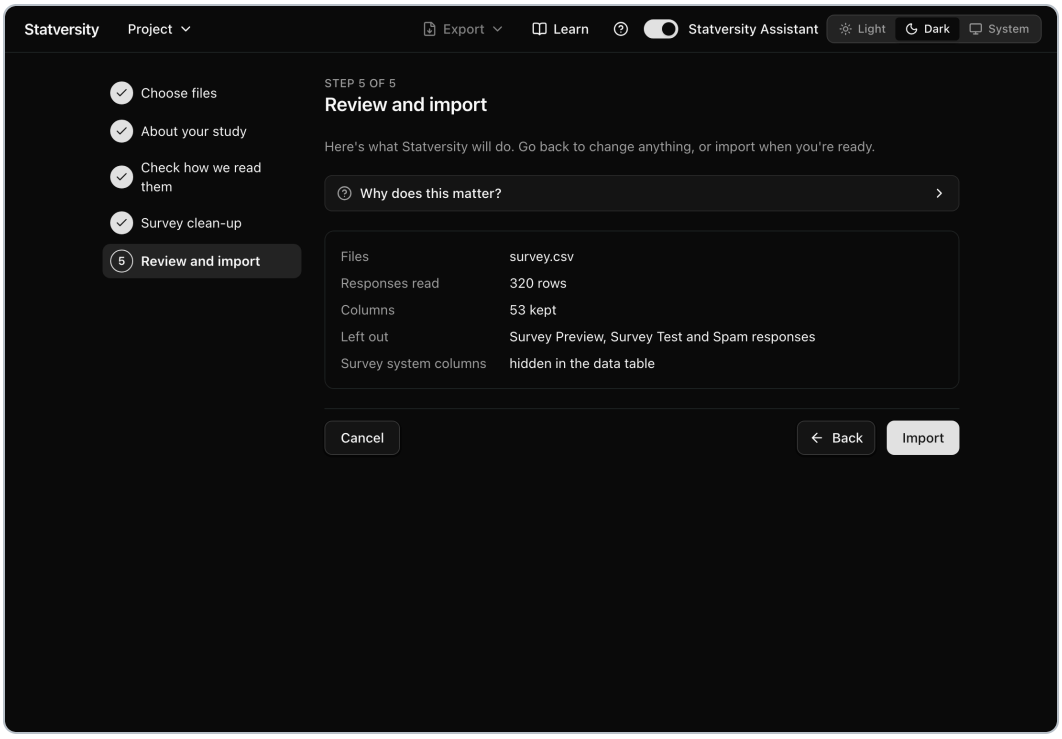


Figure 189. Review and import. 320 rows, 53 columns.

Part 2: The Variable Interview

For a survey study this is the most important part of the setup, because the answers you give here decide whether the factor analysis is trustworthy. Chapter 6 covers each control.

7. The **Welcome** step says “3 questions to check”. Click **Start**.

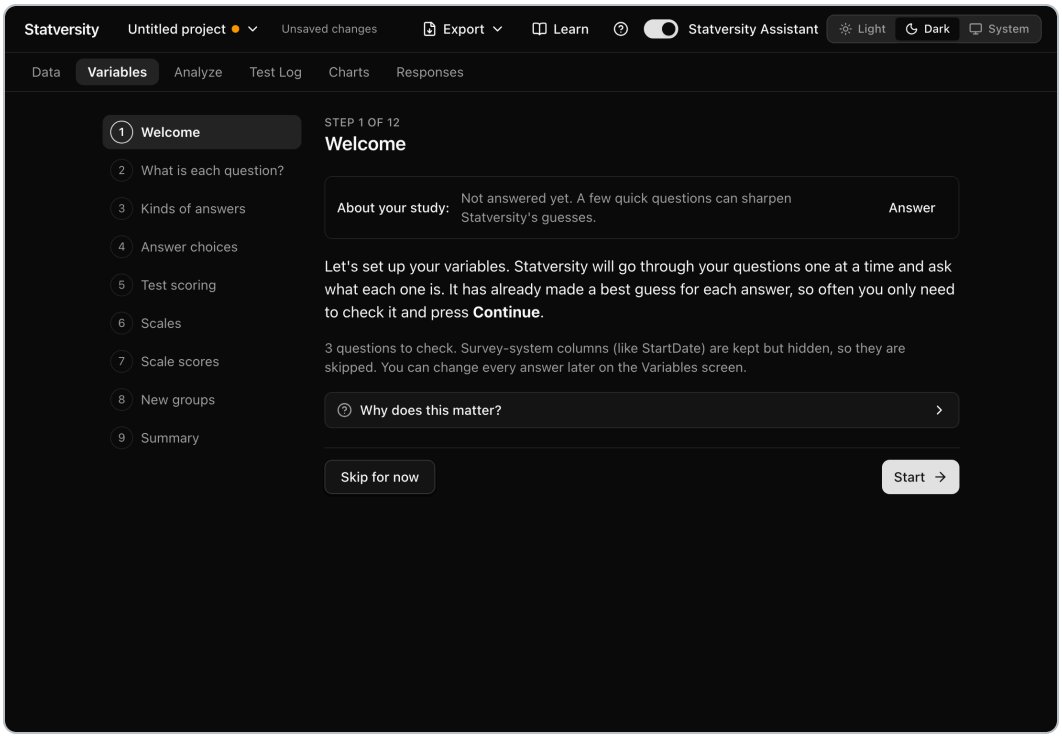


Figure 190. Welcome to the Variable Interview. Three questions to check.

8. The first **What is each question?** screen shows **Q4 (15 items)**, “Matrix item 1: statement about study skills”, with sample answers 4, 3, 5, 2, 1. Statversity chose **Survey (Likert) item**: “An agree/disagree or rating question, like 1 = Strongly disagree to 5 = Strongly agree.” That is correct. Click **Continue**.

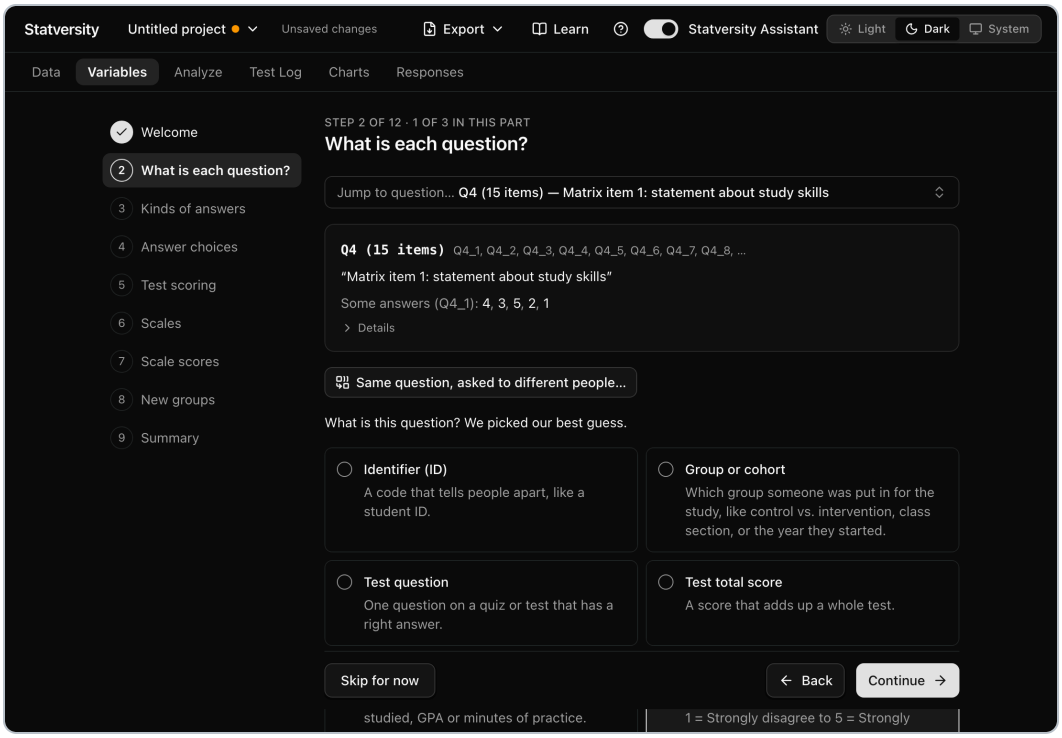


Figure 191. What is each question? for Q4. Survey (Likert) item is selected.

9. The second screen shows **Q5 (20 parts)**, "Quiz item 1 (correct=1, incorrect=0)", with answers 1 and 0. Statversity chose **Test question**, "One question on a quiz or test that has a right answer". Correct. Click **Continue**.

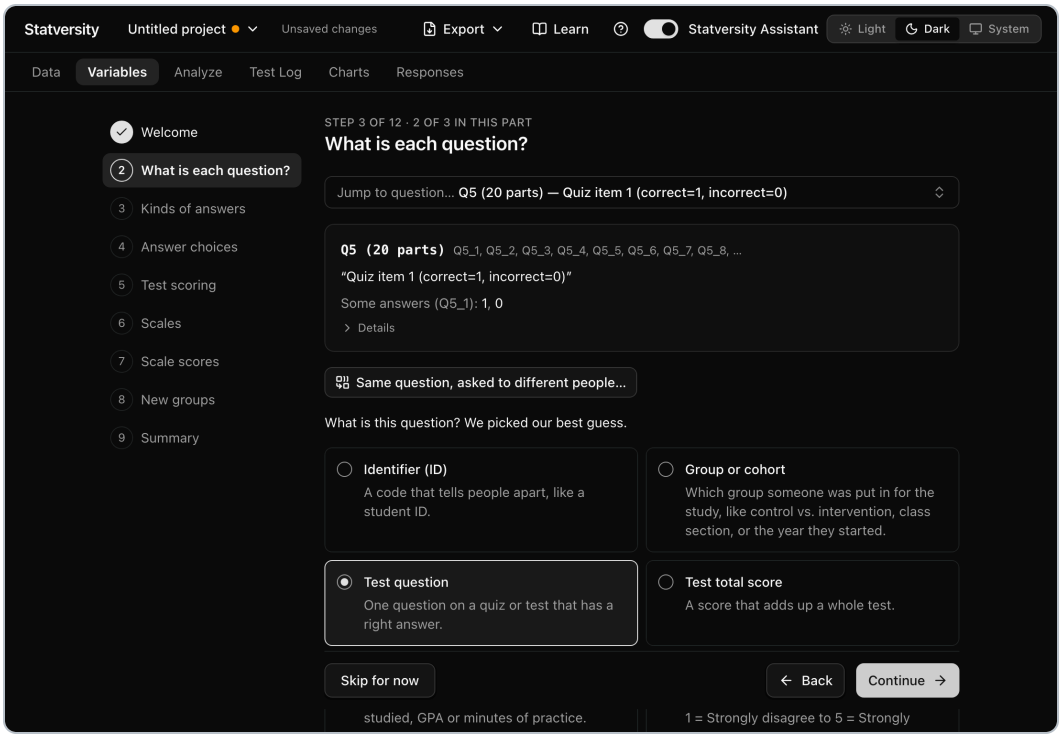


Figure 192. What is each question? for Q5. Test question is selected.

10. The third screen shows **SC0**, "Total correct (0-20)", with answers like 16, 7, 13, 6. Statversity chose **Test total score**. Click **Continue**.

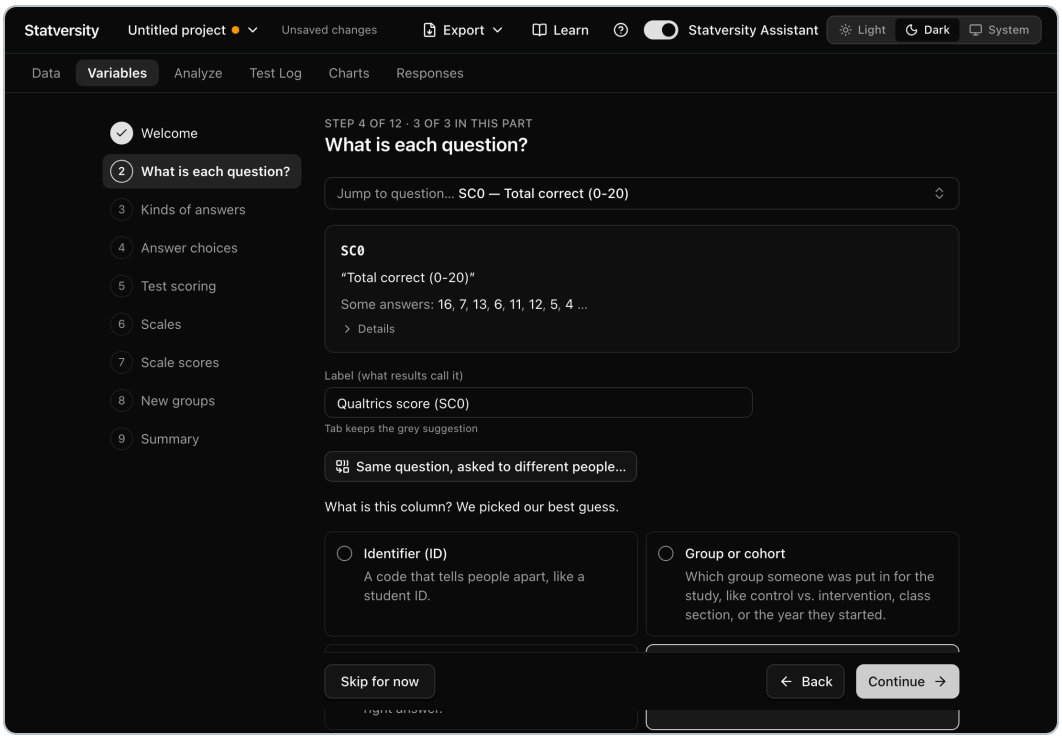


Figure 193. What is each question? for SC0. Test total score is selected.

11. **Kinds of answers** for **Q4** is set to **Ordered categories (ordinal)**. The blue note explains why: “A single agree/disagree question is ordered categories: ‘Agree’ is more than ‘Neutral’, but the gaps may not be equal. Once several questions are combined into a scale score, that score is usually treated as a number.” Click **Continue**.

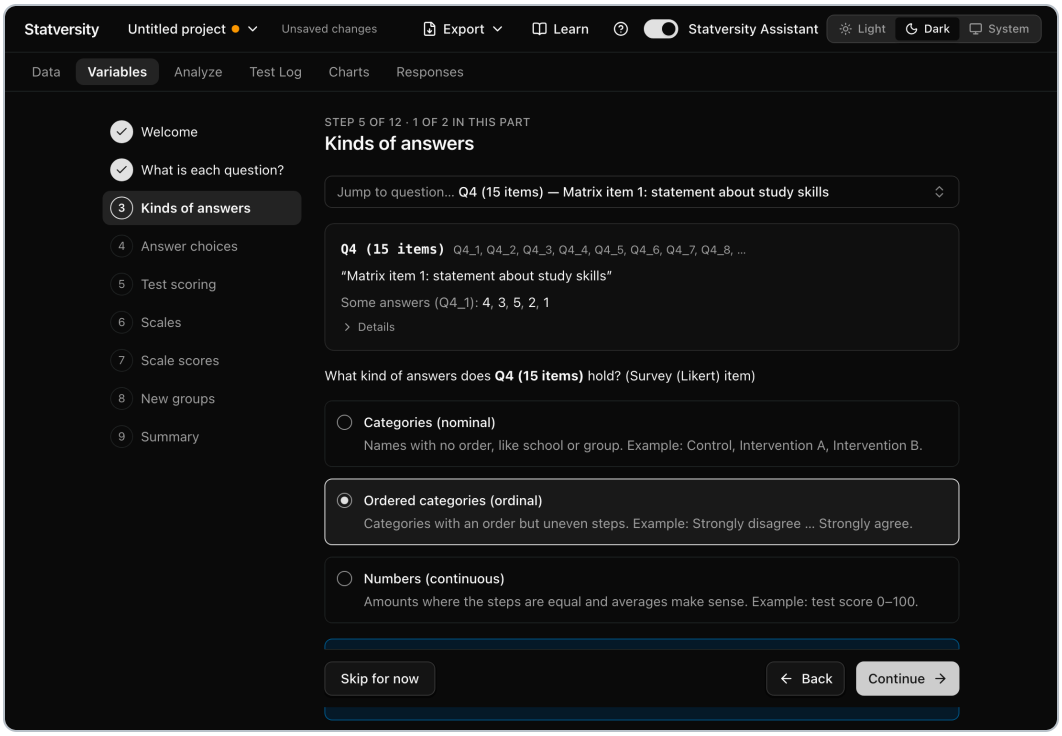


Figure 194. Kinds of answers for Q4. Ordered categories (ordinal) with an explanatory note.

12. For **SC0**, Statversity chose **Numbers (continuous)**. Click **Continue**.

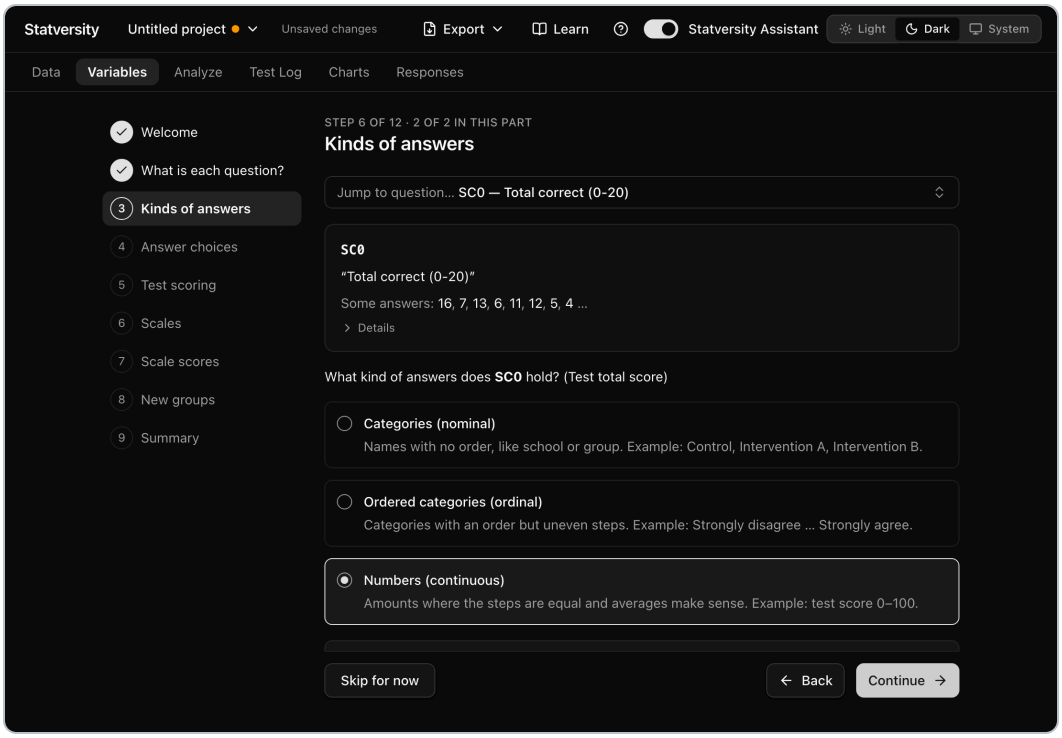


Figure 195. Kinds of answers for SC0. Numbers (continuous).

13. **Answer choices** shows the numbers 1 to 5 stored for Q4, with a row of shortcut buttons such as **Agreement** and **Keep the numbers**. The blue note says numbers still work for every test, so you can skip this. Click **Continue**.

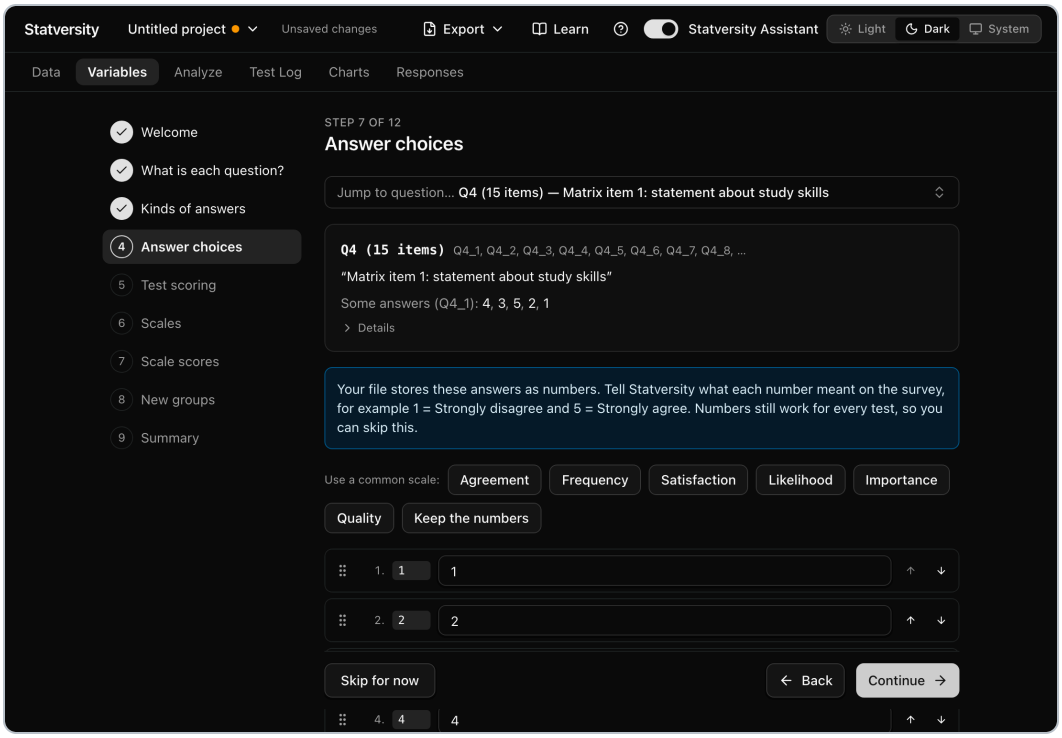


Figure 196. Answer choices for Q4, a numbered list 1 to 5.

14. **Test scoring** says "20 test questions are already scored as numbers. Should Statversity add them up into a total score?" **Yes, add them up (1 = correct)** is selected. We already have a total in the file, and the analysis we will run on the quiz uses the individual items, so either choice works. Leave it and click **Continue**.

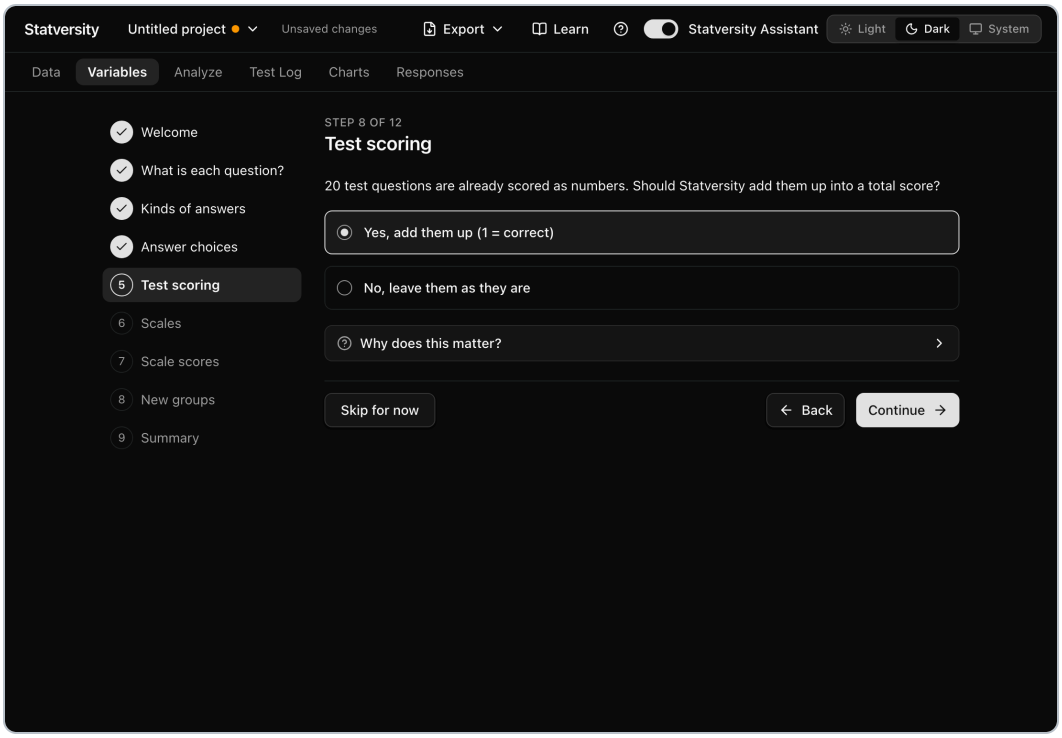


Figure 197. Test scoring with Yes, add them up selected.

15. **Scales** is the key screen. A **scale** is a set of questions that together measure one idea; Statversity averages each set into one score per person. It has grouped Q4_1 to Q4_15 into a single scale of 15 items. Each item has a box: **Worded the opposite way (Statversity will flip its score)**. Click **Continue** only after checking the next step.

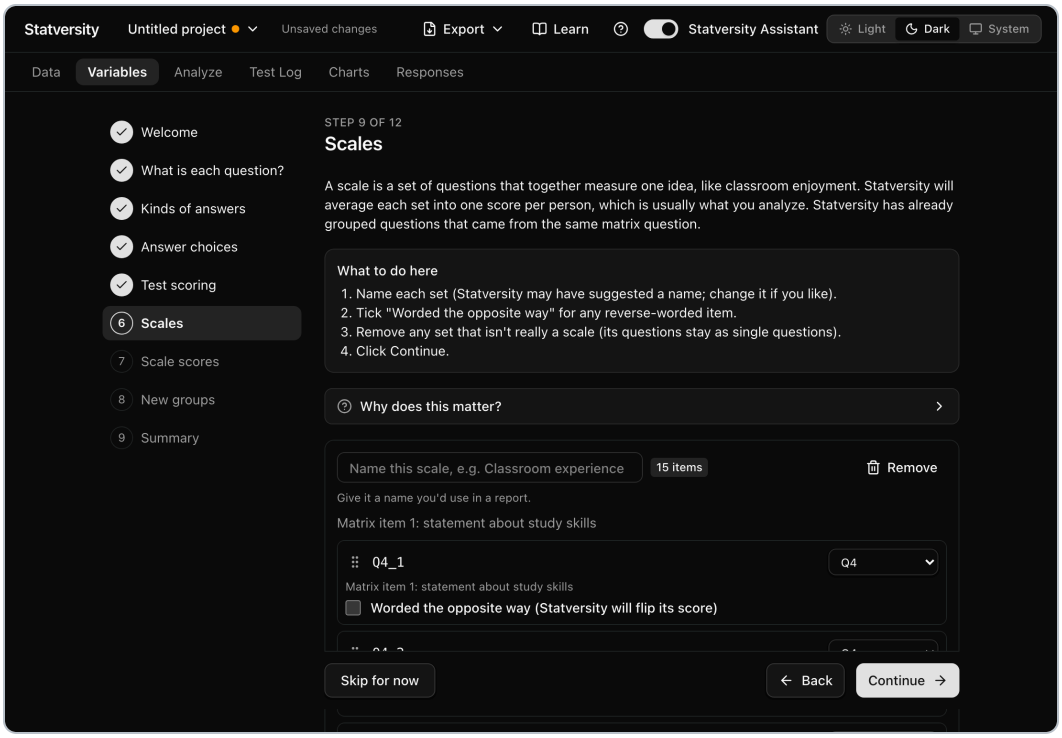


Figure 198. Scales, listing the 15 items with a Worded the opposite way checkbox on each.

16. Scroll to the reverse-worded items. **Q4_3** and **Q4_9** are already ticked, with the note “Ticked because the question text says it’s reverse-worded.” No other item is ticked. This matches the survey design. Click **Continue**.

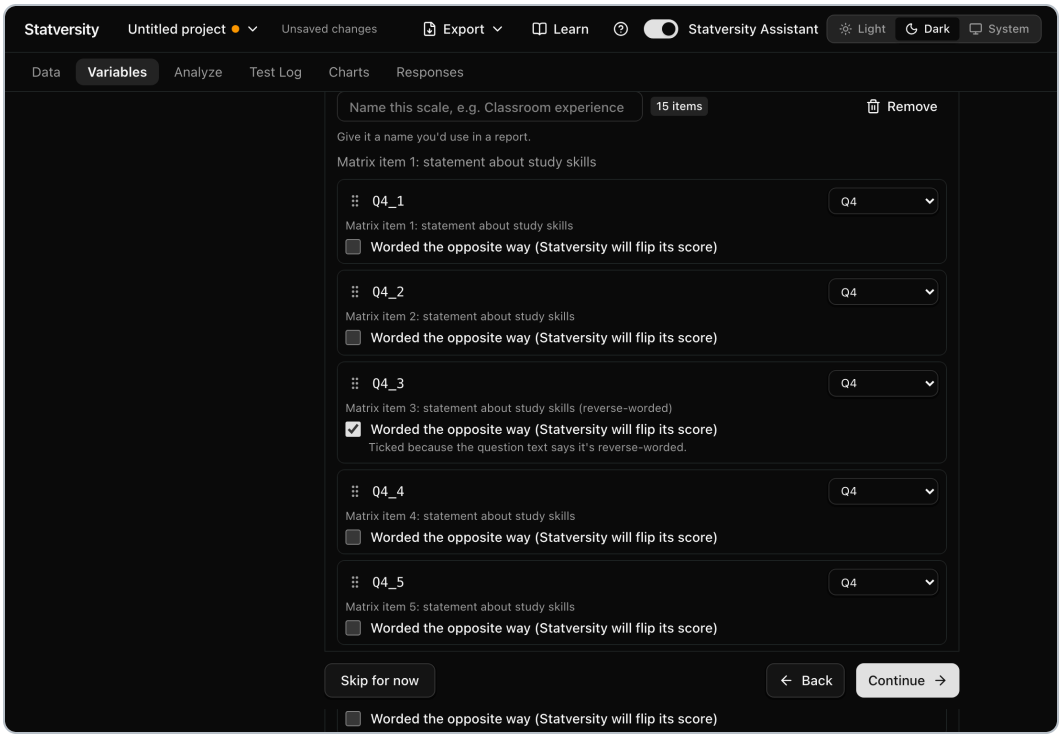


Figure 199. Scales scrolled to Q4_3, which is ticked as worded the opposite way.

IN PLAIN LANGUAGE

Why reverse-worded items matter. Suppose most items say things like “I plan my study time” (agree = good study skills) but one says “I often leave studying to the last minute” (agree = poor study skills). A person with excellent habits would answer 5 on the first and 1 on the second. If you left the second one alone, it would look like it disagrees with the rest and would pull the factor analysis off course. Ticking **Worded the opposite way** tells Statversity to flip that item’s answers (a 5 becomes a 1, a 4 becomes a 2) so all the items point the same way before it does any arithmetic. If you skipped this step for your own survey, the result would blame your questions for a problem the scoring caused.

17. **Scale scores** asks how to turn the scale into a score per person. The default, **Average of the answered questions (recommended)**, keeps the score on the same 1 to 5 scale. The box below says it will only score people who answered at least 8 of the 15 questions (at least half, rounded up), leaving others blank. Click **Continue**.

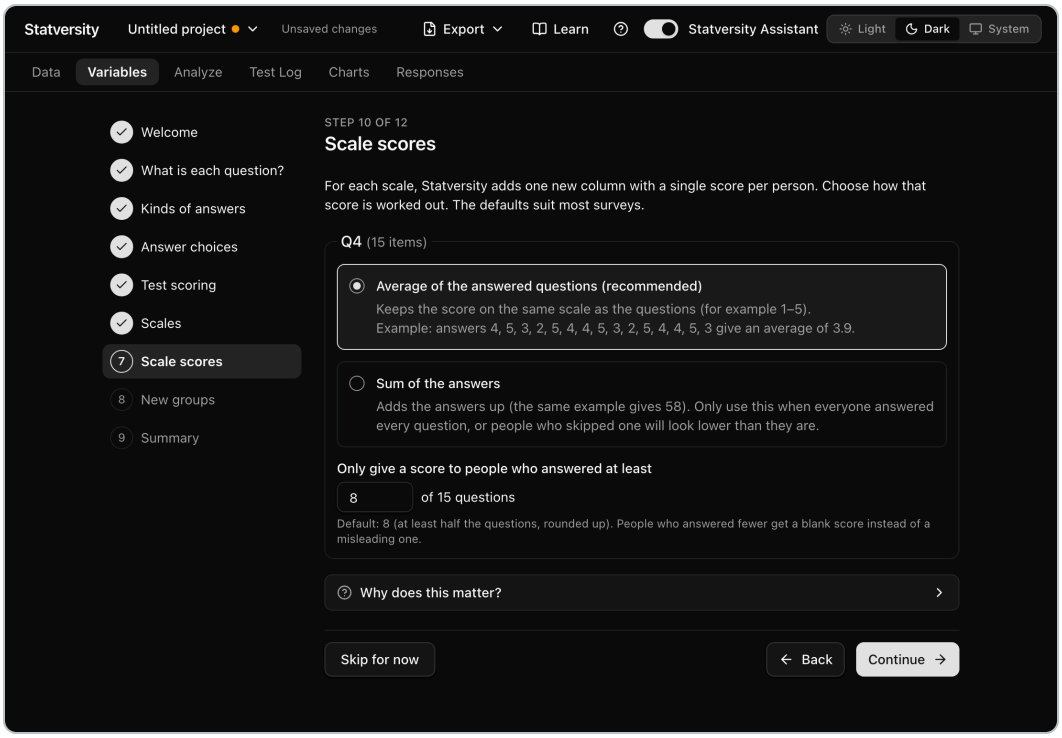


Figure 200. Scale scores. Average of the answered questions is selected, requiring at least 8 of 15.

18. **New groups** offers to build new grouping variables. We do not need any. Click **Continue**.

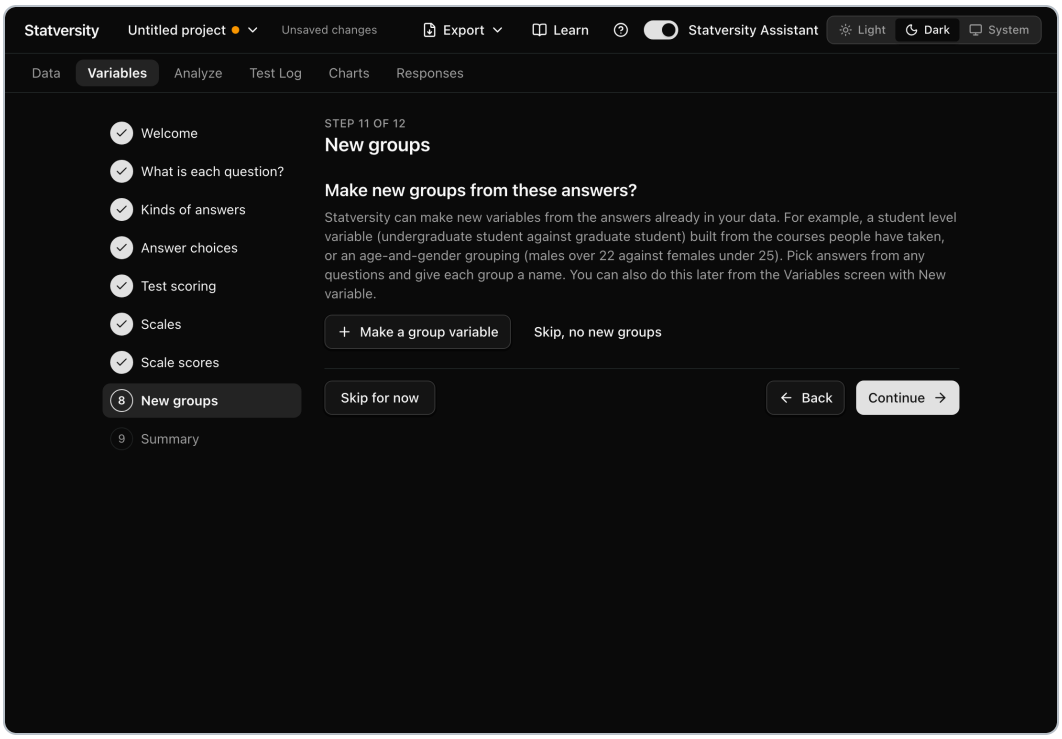


Figure 201. New groups. We made none.

19. The **Summary** lists what you told Statversity. Q4 is a **Survey (Likert) item** and **ordinal**. Q5 is a **Test question** and **nominal**. SC0 is a **Test total score** and **continuous**. Under **Scales** it says Q4 is the "average of Q4_1 … Q4_15 (reverse-scored: Q4_3, Q4_9); needs 8 answered." At the bottom: "35 variables to update · 1 scale score to compute". Click **Finish**.

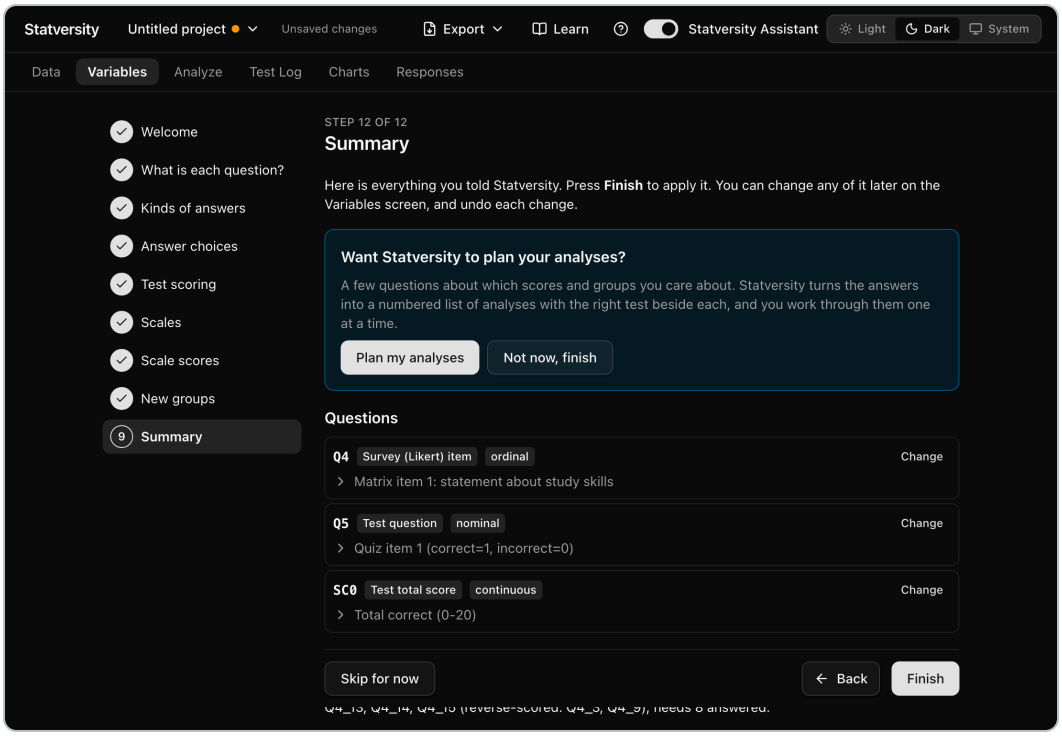


Figure 202. Summary, with Q4 listed as a scale reverse-scored on Q4_3 and Q4_9.

20. The **Variables** screen shows 54 variables. On the right, under **Scales**, you will see Q4: "Average of 15 items, needs 8 answered → Q4_score". Under **Changes** it says "Scored scale Q4".

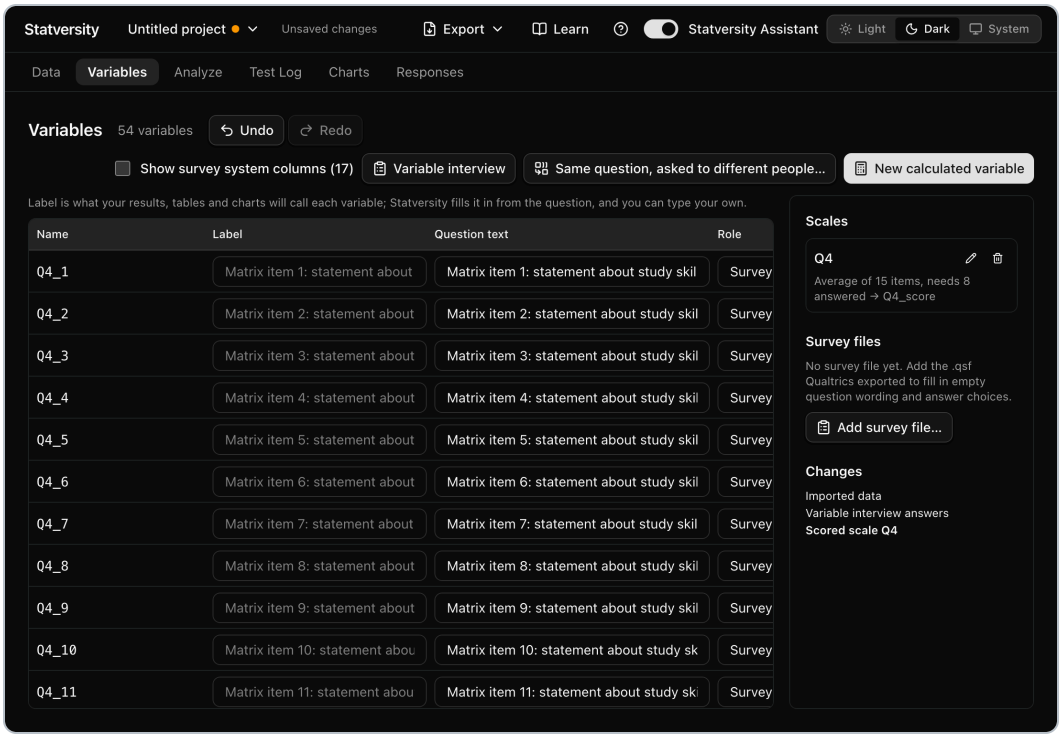


Figure 203. The Variables screen, with the Q4 scale shown at the right.

Part 3: Factor analysis on the 15 survey items

21. Click **Analyze**. The Test Advisor opens. In **Which score or answer do you want to analyse?**, choose **Q4 (average score)**. This tells Statversity you are working with the survey scale.

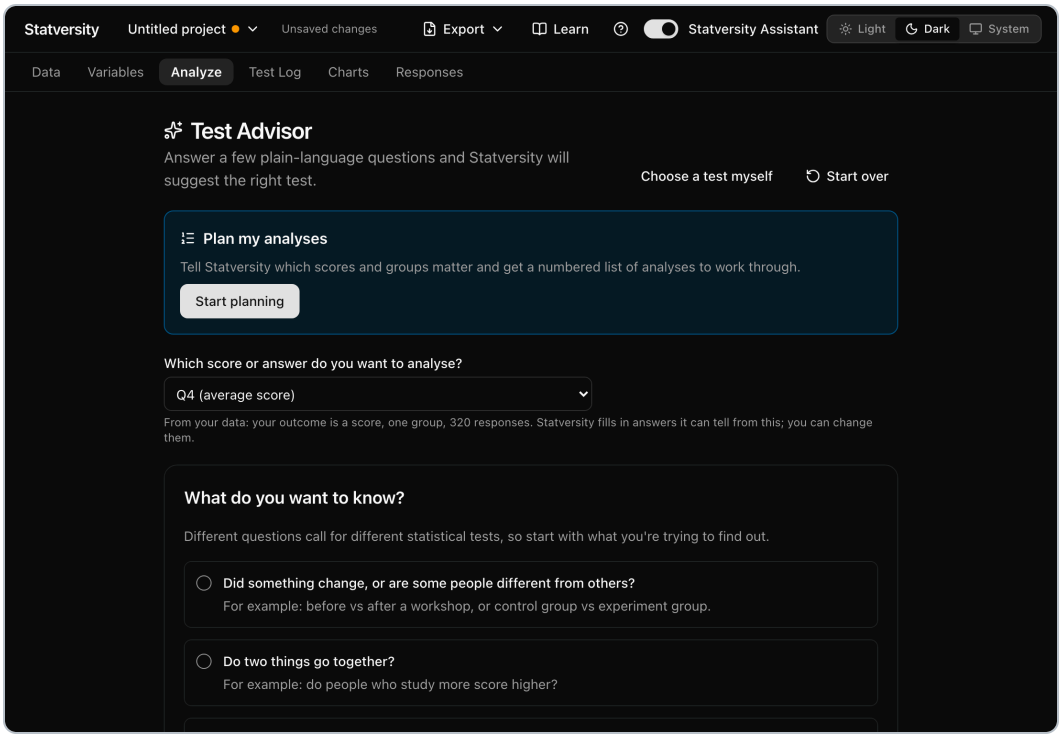


Figure 204. The Test Advisor with Q4 (average score) chosen.

22. Under **What do you want to know?**, choose **Do my questions sort into the topics I designed them around?** and click **Continue**.

> **IN PLAIN LANGUAGE**
>
> **What this question is really asking.** The other choices are about differences between groups, relationships between two things, and predictions. This one is about *structure*: do the individual questions group into the themes you built them around? That is exactly what factor analysis answers.

23. Statversity asks **Are you exploring the structure of your items for the first time, or confirming a structure you already expect?** Choose **Exploring for the first time** and click **Continue**.

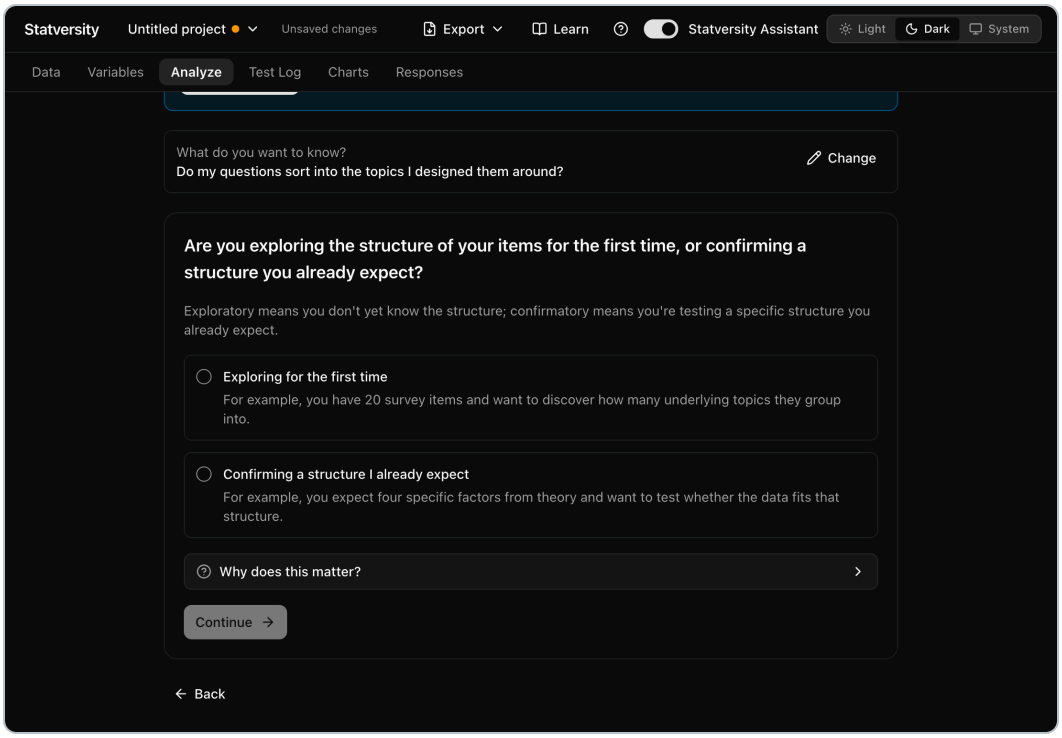


Figure 205. Exploring for the first time or confirming a structure you already expect?

> **IN PLAIN LANGUAGE**
>
> **Why exploring is the right answer here.** We want the data to show us the clusters without us pointing at them. Confirmatory analysis is for when a published theory already tells you exactly which items belong to which factor and you want to test whether your data fit it. For a new survey, you start by exploring.

24. Statversity recommends **Exploratory factor analysis**. An orange note warns that it usually needs at least 100 to 200 responses, which we have (320). On the right, **Assumptions to check** lists three things: **KMO adequacy** (your questions share enough common variance to be worth grouping into factors), **Bartlett sphericity** (your questions are correlated enough with each other to be worth grouping together) and **Enough participants**. Click **Set up this analysis**.

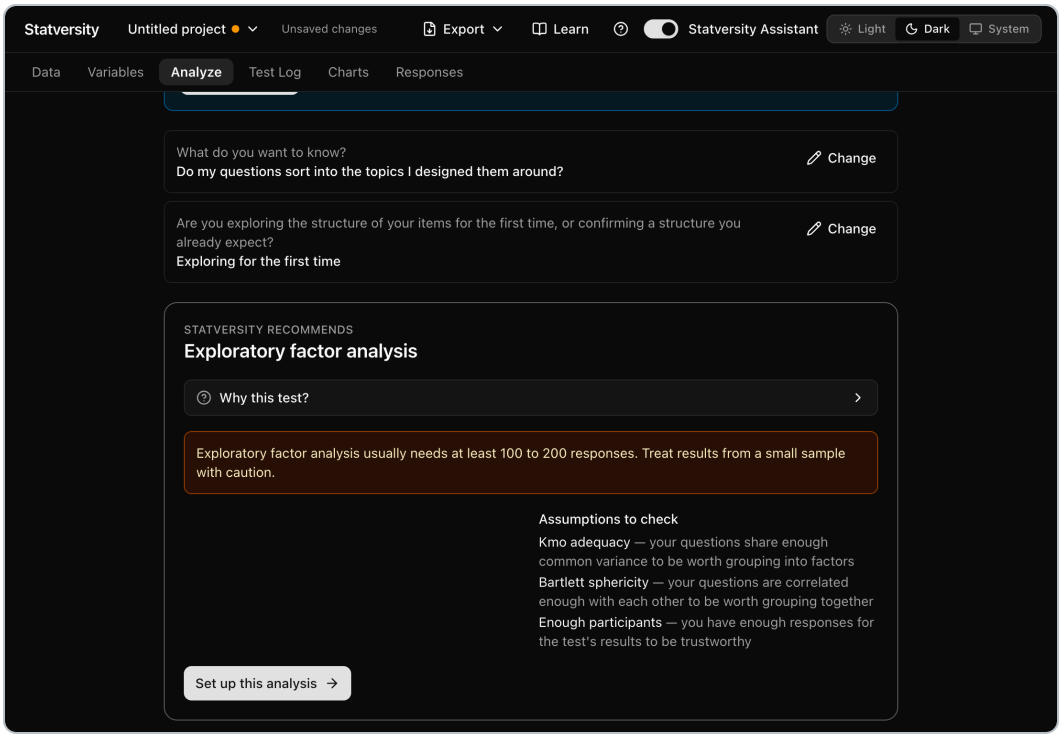


Figure 206. Statversity recommends Exploratory factor analysis, with three assumptions listed.

25. **Choose your variables** opens. Under **Items (3 or more)**, Statversity has already ticked all 15 items, from “Matrix item 1: statement about study skills (Q4_1)” to Q4_15. Notice that items 3 and 9 are labelled “(reverse-worded)”. No change is needed.

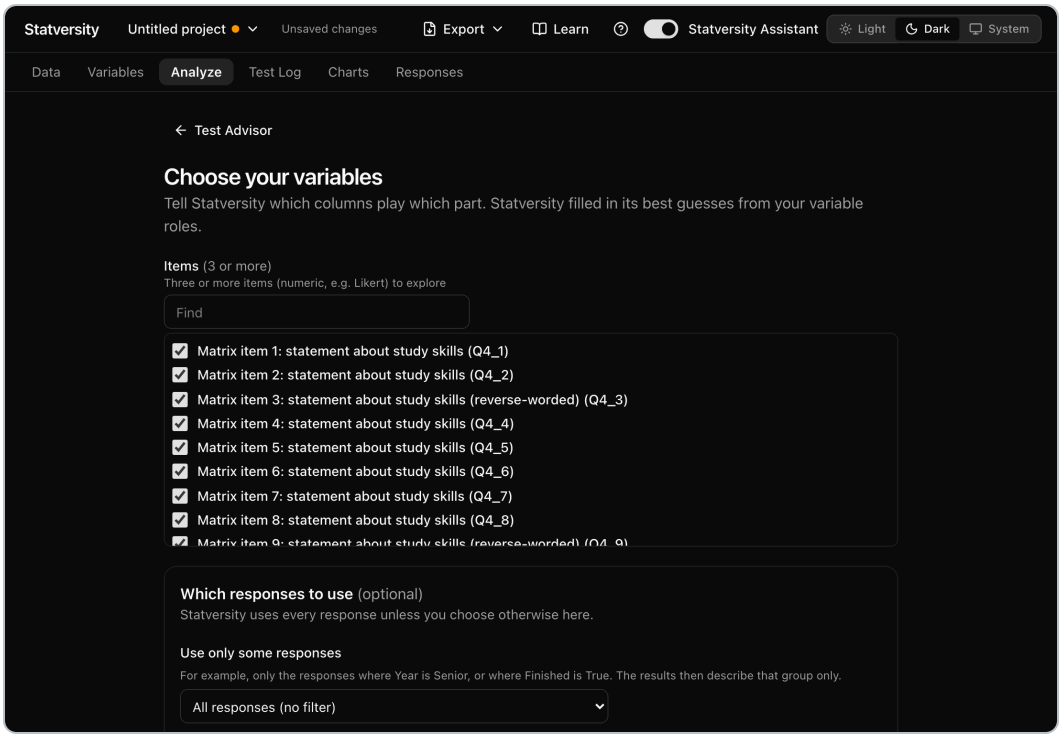


Figure 207. Choose your variables with all 15 matrix items ticked.

26. Click **Check the assumptions**.

> **NOTE**
>
> **This analysis skips the assumption pages.** For many tests, Statversity walks you through one screen per assumption (as in the Which intervention worked best? chapter). For factor analysis and for KR-20 it runs its checks as part of the analysis and shows them in the results, so **Check the assumptions** goes straight to the Results screen.

What Statversity found (factor analysis)

27. The Results screen shows **Exploratory factor analysis** and "320 responses analysed". Read the **What this means** box. Statversity says: "The answers to the items group into 3 factors (the number suggested by parallel analysis; parallel analysis suggested 3). Together they account for 43% of the variation in the items." It then lists which items belong together and says "The data were meritorious for factor analysis (KMO = .81; .60 or higher is usually considered adequate)."

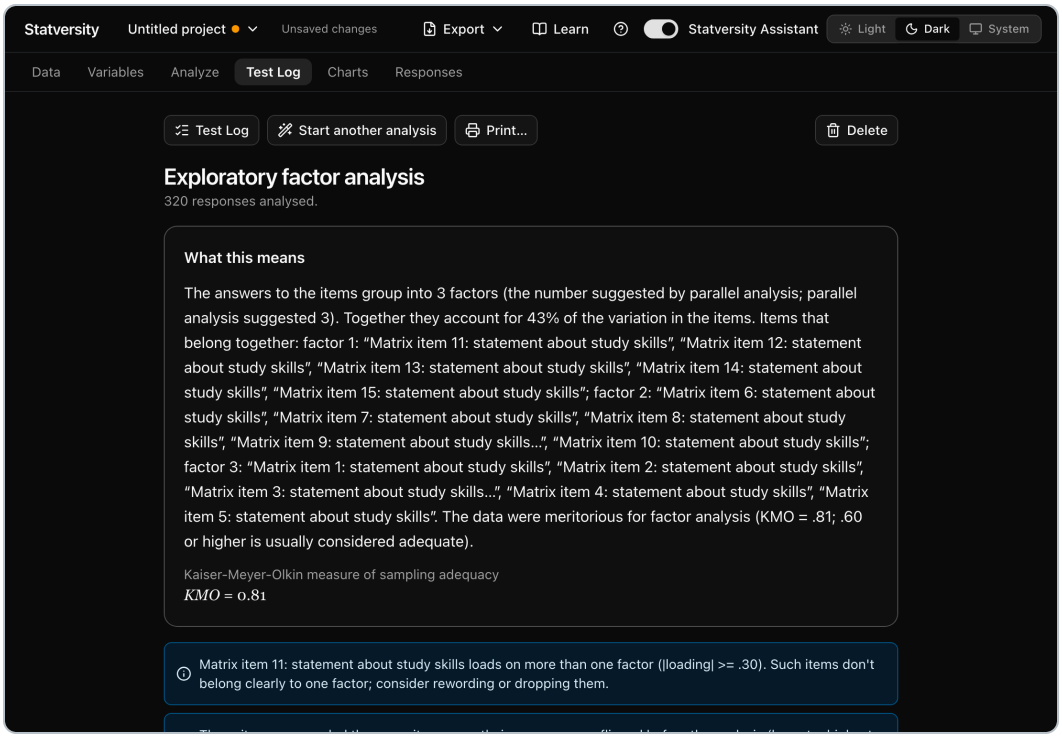


Figure 208. Exploratory factor analysis results, with the What this means box and KMO = 0.81.

Statversity names the three groups it found. Factor 1 is items 11 to 15. Factor 2 is items 6 to 10. Factor 3 is items 1 to 5. That is exactly the way the survey was written: three blocks of five.

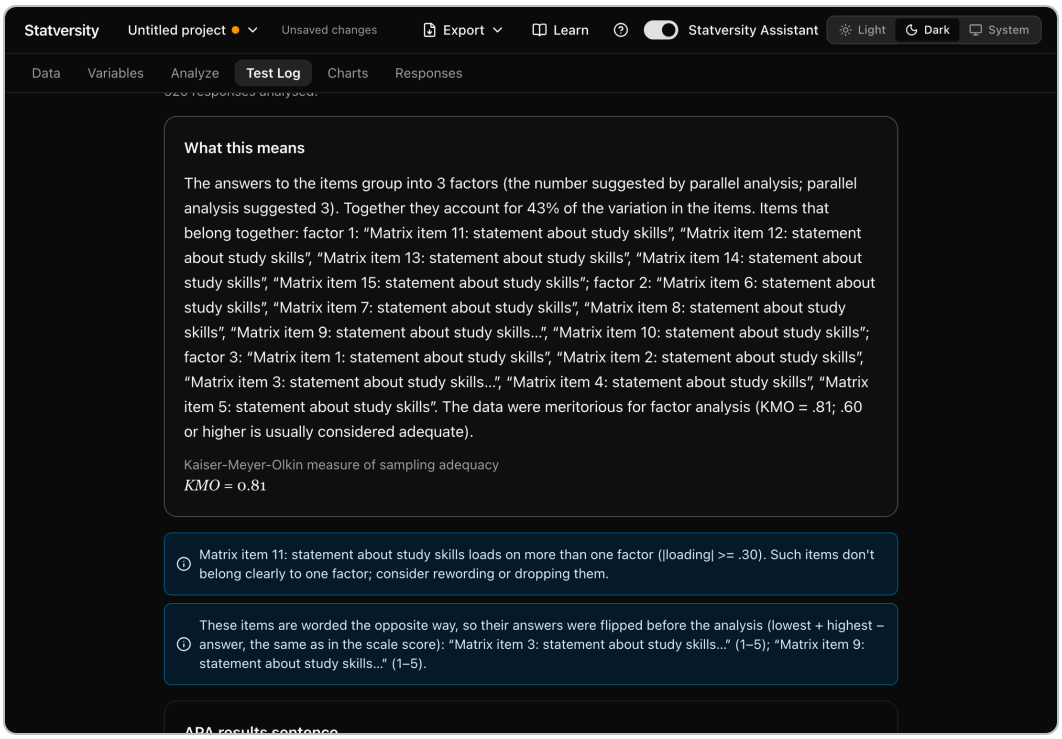


Figure 209. The What this means box scrolled into view, with two blue notices below it.

Two blue notices appear beneath the box:

- “Matrix item 11: statement about study skills loads on more than one factor (|loading| >= .30). Such items don’t belong clearly to one factor; consider rewording or dropping them.”
- “These items are worded the opposite way, so their answers were flipped before the analysis … ‘Matrix item 3’ (1-5); ‘Matrix item 9’ (1-5).” This confirms that your reverse-worded settings were applied.

> **IN PLAIN LANGUAGE**
>
> **Three words you just met.**
>
> *Parallel analysis* is how Statversity decided how many factors to keep. It compares your data with random data of the same size and keeps only the factors that explain more than chance would. It suggested 3.
>
> *KMO* (Kaiser-Meyer-Olkin) is a number from 0 to 1 that says whether your data are suitable for factor analysis in the first place. Higher is better. Statversity’s rule of thumb on screen is that .60 or higher is adequate, and .81 is labelled “meritorious”.
>
> *Loading* is how strongly an item is tied to a factor, like a correlation between the item and the underlying topic. A loading near .70 means the item is closely tied to the factor. Loadings below .30 are small enough that Statversity leaves them blank in the table.

28. Scroll down to **APA results sentence** and **APA table**. The sentence reads “An exploratory factor analysis (minimum residual extraction, oblimin rotation) of 15 items (n = 320) retained 3 factors, explaining 43.4% of the variance. Sampling adequacy was meritorious, KMO = .81, and Bartlett’s test was significant, $\chi^2(105) = 1266.03$, $p < .001$.” Below it is **Table 1: Factor Loadings for the Items**.

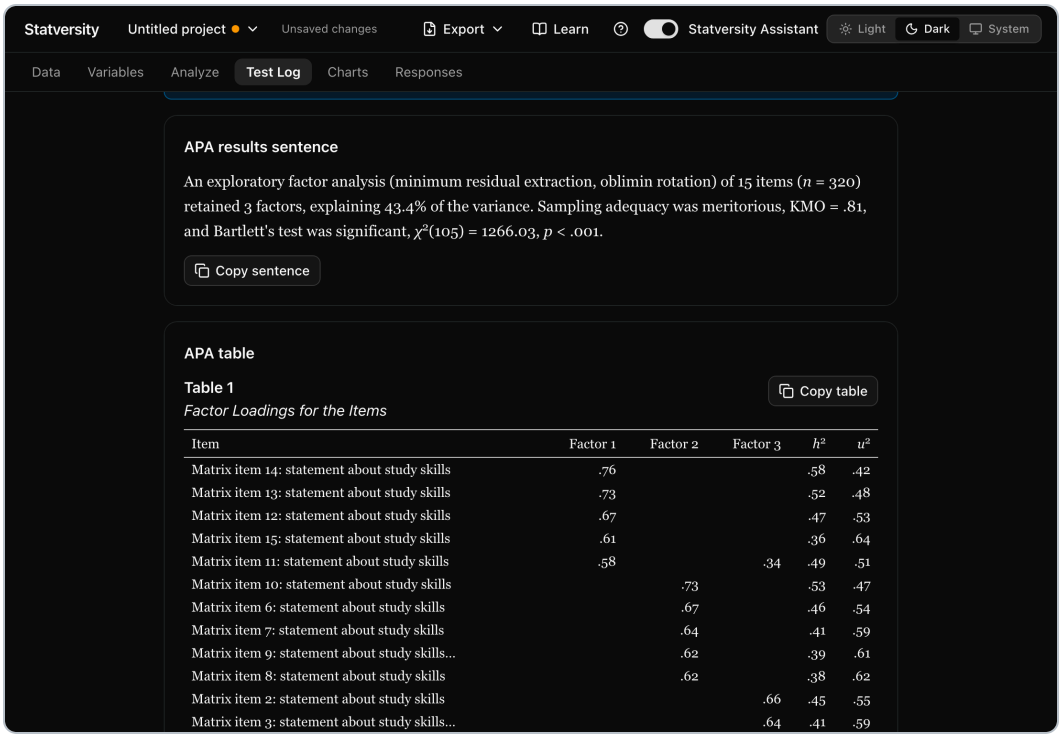


Figure 210. The APA sentence and the factor loading table.

Here is how to read the loading table. Each row is an item. Each factor column shows the loading only if it is .30 or higher.

Factor 1	Factor 2	Factor 3
Item 14: .76	Item 10: .73	Item 2: .66
Item 13: .73	Item 6: .67	Item 3: .64
Item 12: .67	Item 7: .64	Item 5: .63
Item 15: .61	Item 9: .62	Item 4: .58
Item 11: .58	Item 8: .62	Item 1: .57

The table also has two more columns, h² (communality: how much of the item's variation the three factors together explain) and u² (uniqueness: what is left over). The bottom row shows the share of variance each factor explains: 15.17%, 14.51% and 13.74%.

Every item has a clear main loading between .57 and .76 on exactly the factor it was written for. There is one exception worth knowing about. **Item 11** loads .58 on Factor 1 but also .34 on Factor 3. This is the **cross-loading** that Statversity flagged in the blue notice. It means item 11 partly measures the topic of the Factor 3 items as well as its own.

> **TIP**
>
> **What to do with a cross-loading item.** You have three choices: reword it so it speaks only to its own topic, drop it from the scale, or keep it and explain why. Dropping an item is a decision about your instrument, not a statistical requirement, and it is yours to make. A main loading of .58 is perfectly respectable, so many researchers would keep it while noting the cross-loading.

Part 4: Reliability of the 20-item quiz

Now to the second question. Statversity stays in the same project.

29. Click **Start another analysis**. The Test Advisor opens again. In **Which score or answer do you want to analyse?**, choose **Qualtrics score (SC0)**. We will pick the quiz items ourselves on the variables screen.

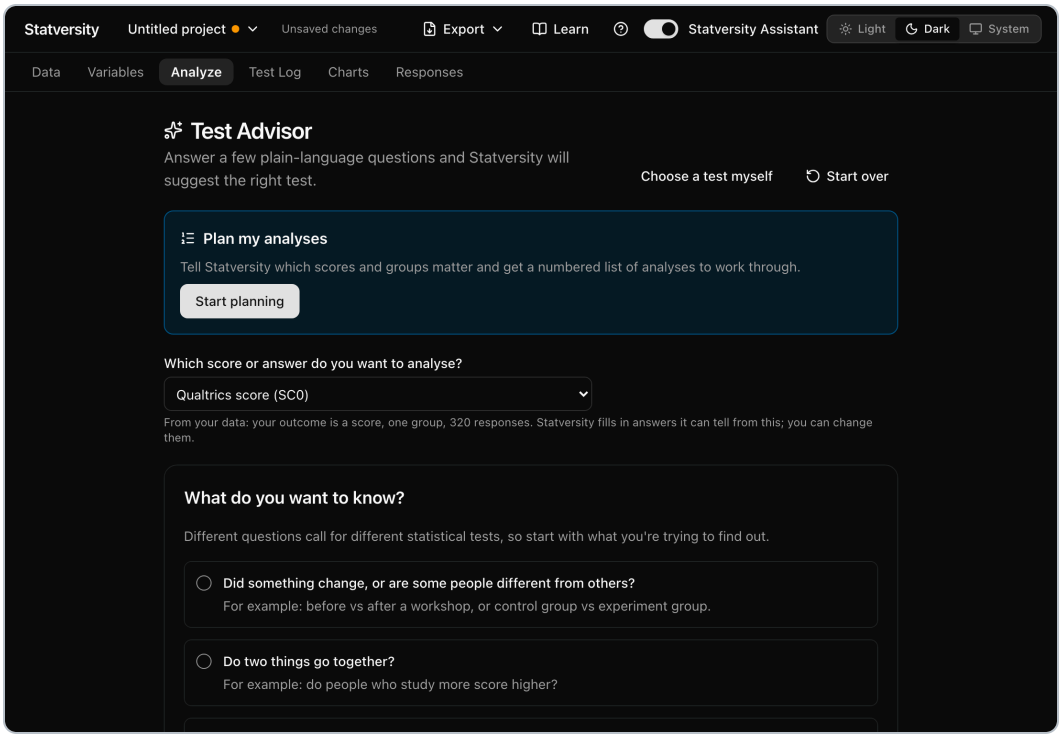


Figure 211. The Test Advisor with Qualtrics score (SC0) chosen.

30. Under **What do you want to know?**, choose **Do my rating questions consistently measure one idea?** and click **Continue**. Statversity asks **What kind of questions are they?**.

31. Choose **Test questions (can be correct or incorrect)** and click **Continue**.

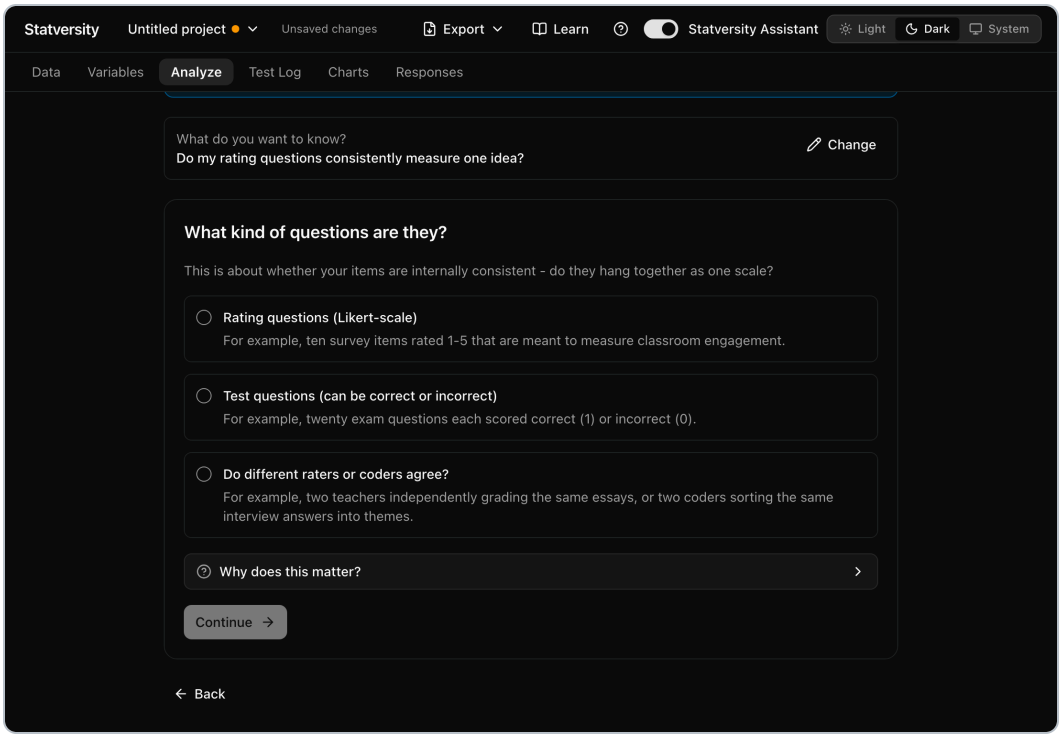


Figure 212. What kind of questions are they? The three choices are rating questions, test questions and rater agreement.

> **IN PLAIN LANGUAGE**
>
> **Why Test questions and not Rating questions?** Each quiz question is either right (1) or wrong (0), not a 1 to 5 rating. For right/wrong items, the correct reliability statistic is KR-20. For rating items you would use Cronbach's alpha. Statversity picks the correct one from your answer.

32. Statversity recommends **KR-20 (right/wrong items)**. Its one assumption is **Tau equivalence**: "every question measures the idea about equally strongly". Click **Set up this analysis**.

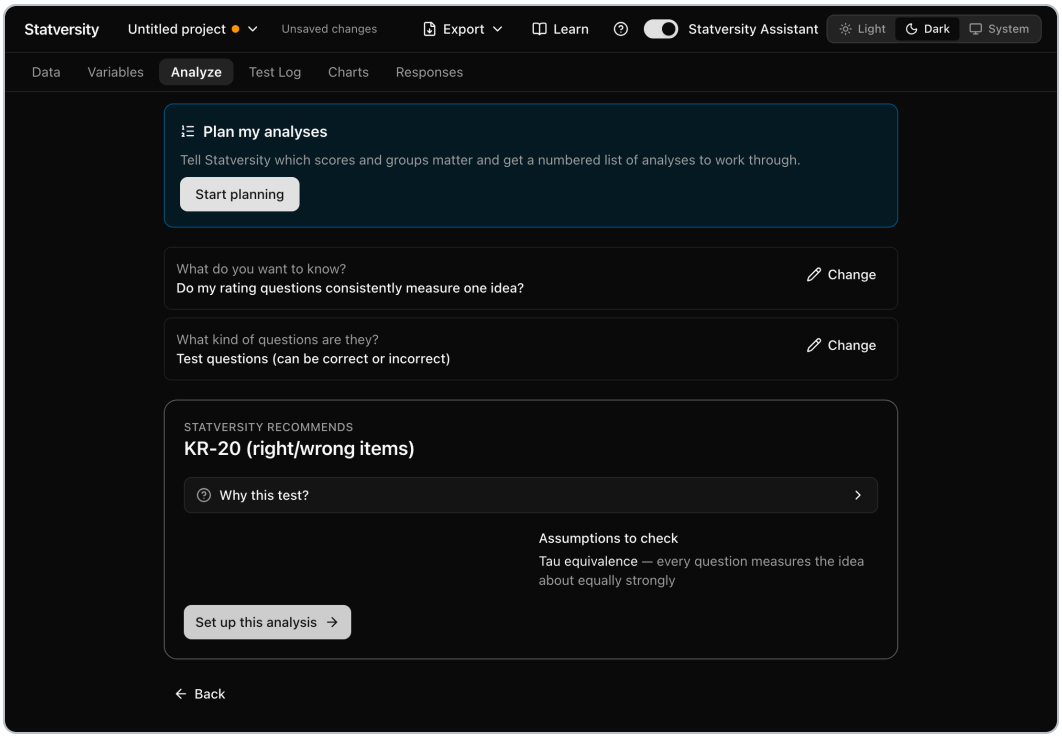


Figure 213. Statversity recommends KR-20 (right/wrong items).

33. **Choose your variables** opens, this time with every item unticked. Statversity could not guess which items are the quiz.

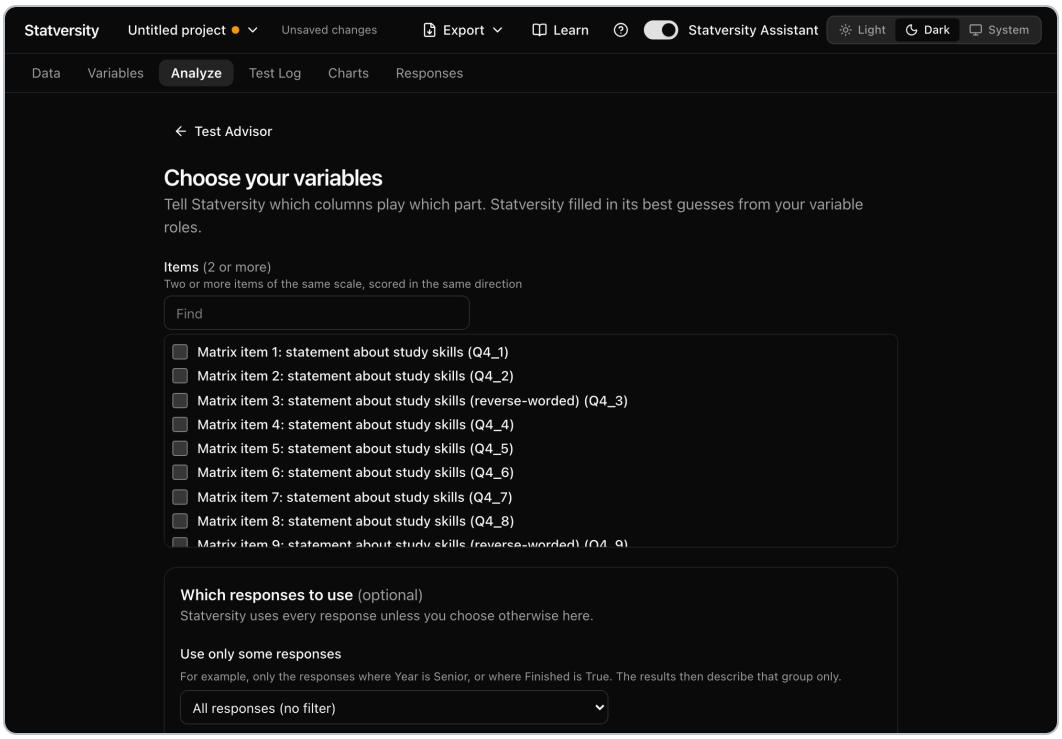


Figure 214. Choose your variables for KR-20, with the Q4 survey items unticked.

34. Under **Items (2 or more)**, tick **Quiz item 1** through **Quiz item 20** (Q5_1 to Q5_20) and make sure **Qualtrics score (SC0)**, at the bottom of the list, is unticked. You can use the **Find** box to type "Quiz" and narrow the list. The screen shows all 20 quiz items ticked and **Qualtrics score (SC0)** left unticked.

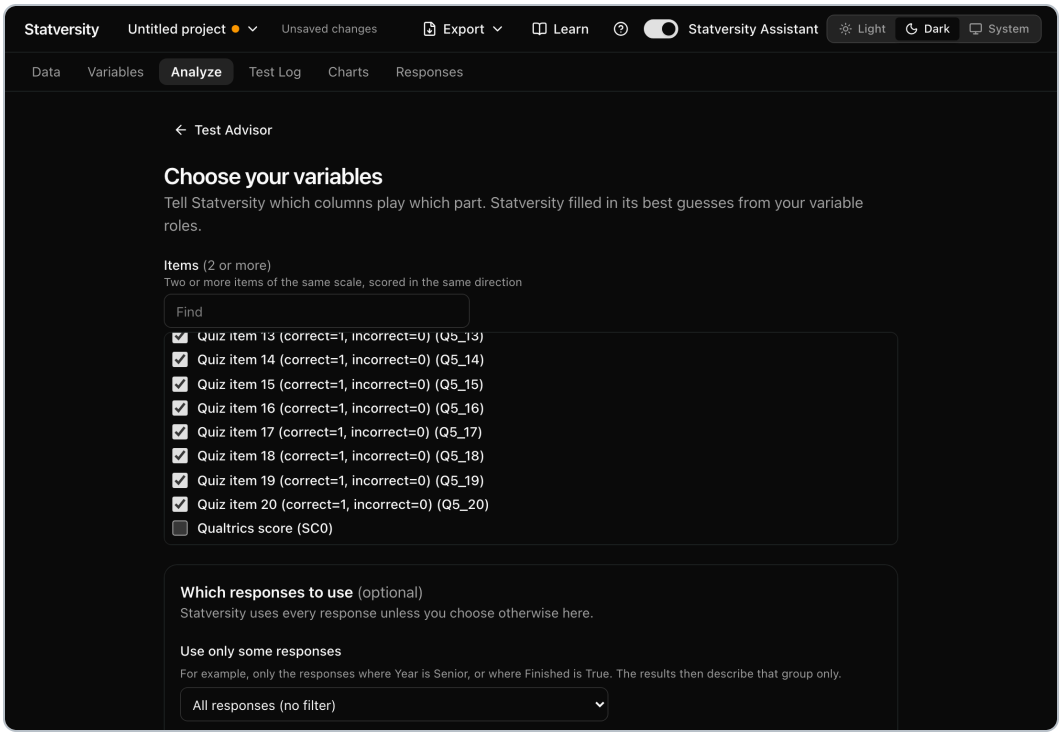


Figure 215. Choose your variables with Quiz items 13 to 20 ticked and the Qualtrics score unticked.

> **CAUTION**
>
> **Do not include the total score.** SC0 is the sum of the other 20 items. Putting it in the item list would make the quiz look perfectly consistent for a trivial reason: the total is built from the parts.

35. Click **Check the assumptions**. Again it goes straight to the Results screen.

What Statversity found (KR-20)

36. The Results screen shows **KR-20 (right/wrong items)** and "320 responses analysed". The **What this means** box says: "The items of the scale hang together well: KR-20 is .74, which is acceptable by common rules of thumb. These labels (George & Mallery, 2003) are rules of thumb: .70 or higher is usually acceptable for research, while decisions about individual students call for .90 or higher. The item that fits least well is 'Quiz item 10' (item-total r = -.32); removing it would give .78." Below it, an orange warning reads "Quiz item 10 goes against the rest of the test (negative item-total correlation). Review this item for a scoring or wording problem, or consider dropping it."

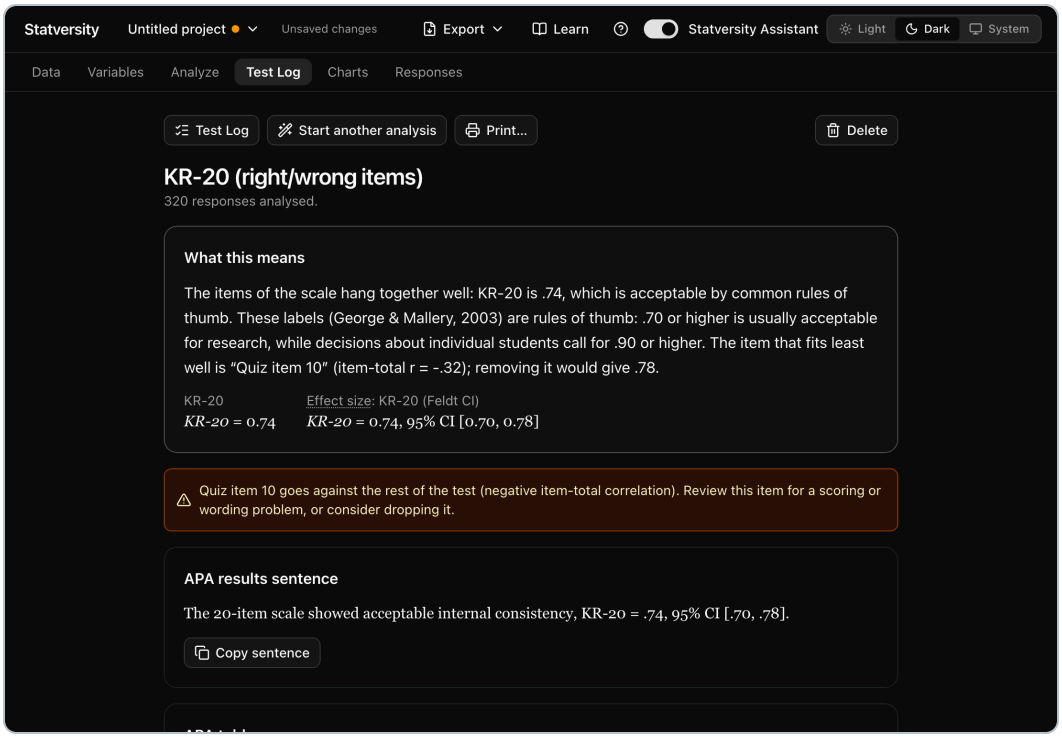


Figure 216. KR-20 results. The What this means box and a warning about Quiz item 10.

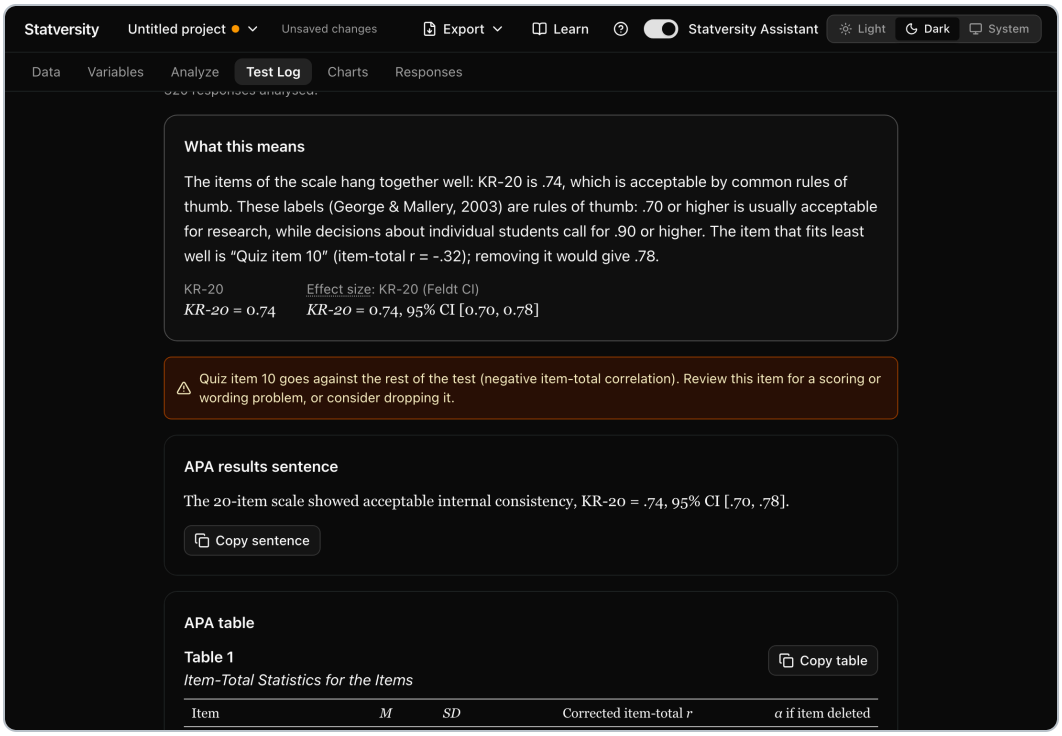


Figure 217. The What this means box, the warning about Quiz item 10, and the APA sentence.

The headline numbers are KR-20 = .74 and the 95% confidence interval [.70, .78].

37. Scroll to the **APA table**. The title is **Item-Total Statistics for the Items**. For each item it shows M (the proportion who got it right), SD, the **corrected item-total r** (how closely that item's right/wrong pattern tracks the rest of the quiz), and **α if item deleted** (what the reliability would be without that item).

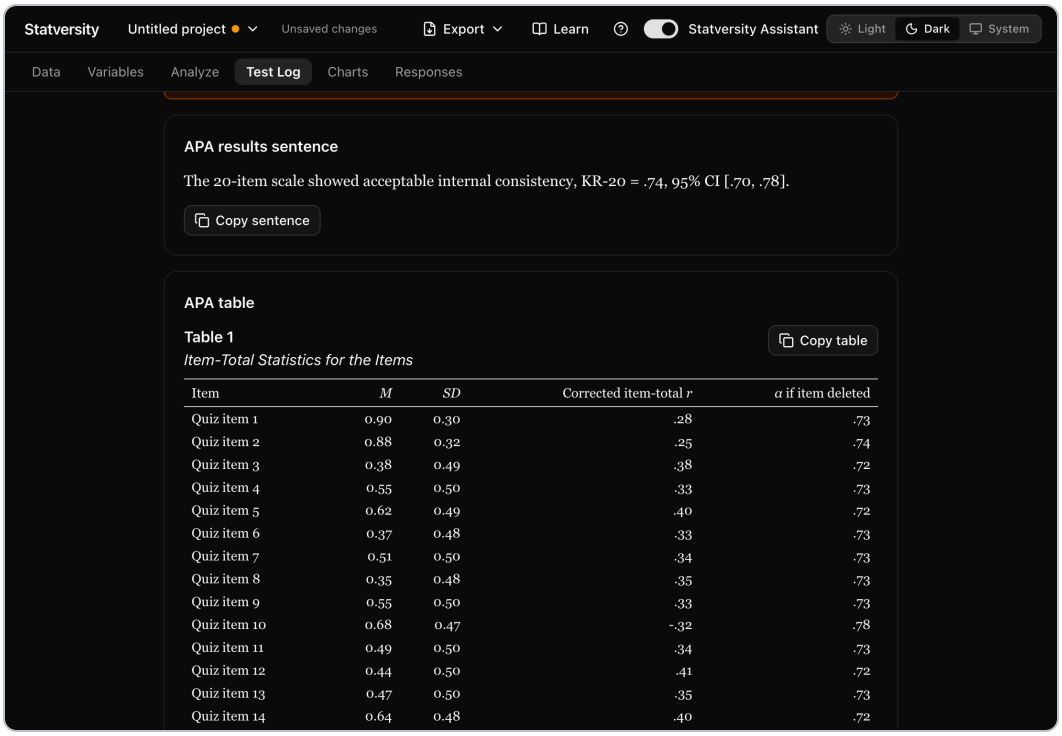


Figure 218. The APA sentence and the item-total table, listing every quiz item.

Most items have an item-total correlation between .25 and .42 and an "if deleted" value of .72 to .74, which is what a healthy item looks like. Four rows stand out:

Item	Proportion correct (M)	Item-total r	If deleted
Quiz item 1	.90	.28	.73
Quiz item 2	.88	.25	.74
Quiz item 19	.09	.14	.74
Quiz item 20	.09	.20	.74
Quiz item 10	**.68**	**-.32**	**.78**

Items 1 and 2 are very easy (about 9 in 10 people got them right) and items 19 and 20 are very hard (about 1 in 10). Their correlations are low but positive, so they are doing their job a little, just less than the middle items. **Item 10 is different**: its item-total correlation is *negative*, meaning students who did well on the rest of the quiz were *less* likely to get item 10 right. Removing it would raise reliability from .74 to .78.

38. Scroll to **Descriptive statistics**. This repeats the M and SD for each item (for right/wrong items, M is simply the proportion correct). This result has no chart.

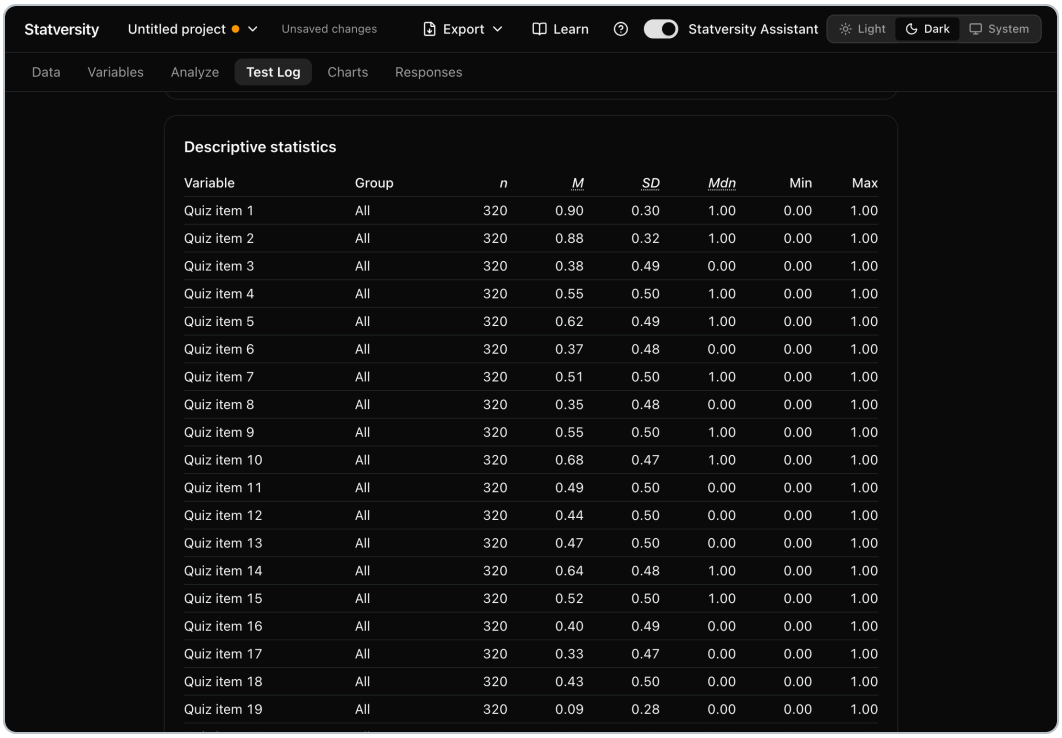


Figure 219. Descriptive statistics for the 20 quiz items.

39. Click **Test Log**. Both analyses are saved: **KR-20 (right/wrong items)** with its item list and **Exploratory factor analysis** with Q4_1 to Q4_15, each with its sentence, date and n = 320.

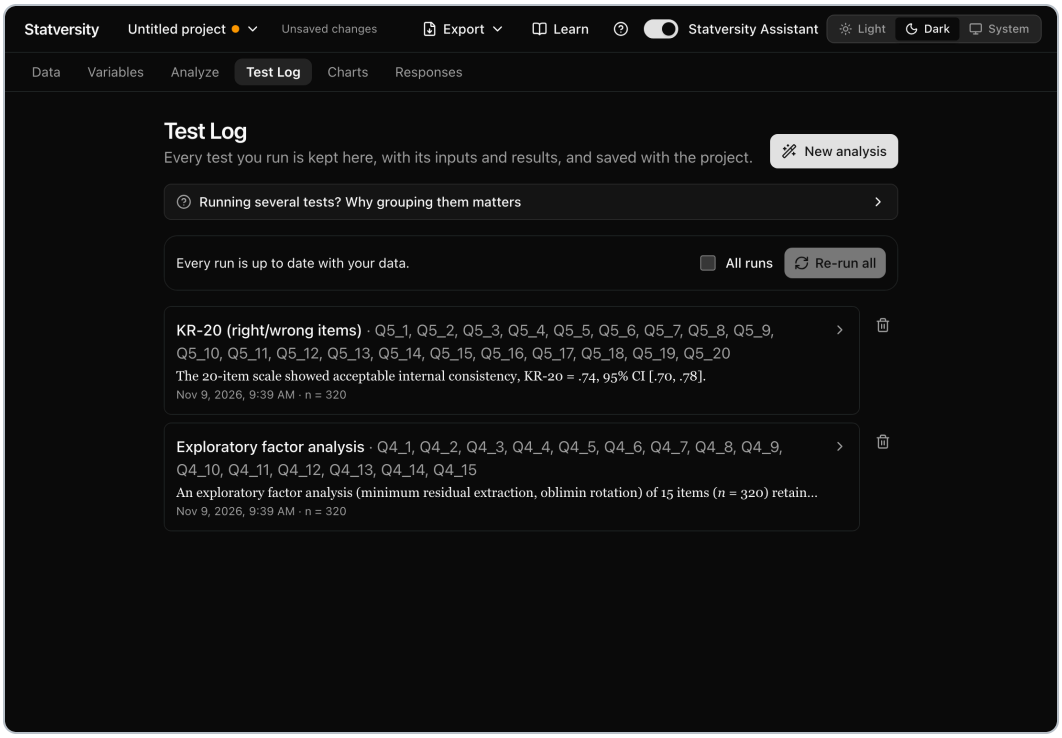


Figure 220. The Test Log showing both analyses.

Numbers at a glance

Analysis	Result
Factor analysis, sample	n = 320, 15 items
Factors retained	3 (parallel analysis suggested 3)
Variance explained	43.4% total (15.17%, 14.51%, 13.74%)
KMO	.81 (meritorious)
Bartlett's test	$\chi^2(105) = 1266.03$, $p < .001$
Cross-loading item	Matrix item 11 (.58 on Factor 1, .34 on Factor 3)
KR-20	.74, 95% CI [.70, .78] (acceptable)
Worst item	Quiz item 10, item-total r = -.32; KR-20 would be .78 without it

What this means in plain English

The survey. The 15 items sort into three groups of five, exactly the blocks they were written for. That is good evidence that your survey measures three distinct topics. KMO of .81 says the data were well suited to this kind of analysis, and Bartlett's test says the items are correlated with one another rather than unrelated, which factor analysis needs. The three factors together explain about 43% of the variation in

the items. That number sounds modest, but in survey work it is typical, because people's answers to single questions are noisy. One item, number 11, is tied to its own factor but is also partly tied to a second. That does not break the survey; it is a hint that this question mixes two ideas and might be worded more cleanly.

The quiz. A KR-20 of .74 says the 20 questions hang together acceptably well, which is enough for research purposes, though not for high-stakes decisions about individual students (the app notes that decisions about individual students call for .90 or higher). Item 10 is the one question that works against the rest of the test. Look at it before you reuse the quiz. Perhaps the key is wrong, or the wording tricks strong students.

> **CAUTION**
> A reliable test is not automatically a valid one. Reliability says the questions agree with each other. It does not say they measure what you want. Likewise, finding three factors tells you the items cluster, not what each cluster *means*. Reading the items in each cluster and naming the topic is your job.

What you would write in a paper

For the factor analysis, Statversity gives you this sentence:

> An exploratory factor analysis (minimum residual extraction, oblimin rotation) of 15 items (n = 320) retained 3 factors, explaining 43.4% of the variance. Sampling adequacy was meritorious, KMO = .81, and Bartlett's test was significant, $\chi^2(105) = 1266.03$, $p < .001$.

For the reliability:

> The 20-item scale showed acceptable internal consistency, KR-20 = .74, 95% CI [.70, .78].

To adapt them, name the instrument ("the 15-item study skills survey", "the 20-item knowledge quiz"), say how you decided on the number of factors (parallel analysis), report the loading table as your Table 1, and mention the cross-loading item and what you did about it. Reverse-scored items should be named in your methods section.

What if

What if you forgot to mark Q4_3 and Q4_9 as reverse-worded? Those two items would be treated as if they pointed in the same direction as the others. Their loadings would likely appear with the wrong sign or be weak, and the factor structure could look messier than it really is. Statversity shows a blue notice on the results listing which items were flipped, so check that list against your survey.

What if the KMO had been below .60? Statversity would label sampling adequacy “miserable” (.50 to .59) or “unacceptable” (below .50), and the “What this means” box would repeat that .60 or higher is usually considered adequate. A KMO that low means the items are not correlated enough for factor analysis to be meaningful. Check for items that are very different from each other, or for too little variation in answers.

What if parallel analysis had suggested a different number of factors? You would be told, and the groups of items shown would change accordingly. Treat the number as a suggestion to check against theory. If you expected three and the data said two, look at which items merged.

What if you had a theory-driven structure already? Choose **Confirming a structure I already expect** at the Advisor’s second question. That is a different analysis (confirmatory factor analysis) with different setup.

What if KR-20 had been .55? That is below the .70 rule of thumb. The usual remedies are removing items with low or negative item-total correlations (check the table), or adding more good items, because reliability tends to rise as a test gets longer.

What if your quiz questions were rated 1 to 5 instead of right/wrong? On the **What kind of questions are they?** screen you would choose **Rating questions (Likert-scale)**, and Statversity would recommend a reliability statistic for rating items instead of KR-20.

> **TRY IT YOURSELF**
>
> **Try it yourself.** Rerun the KR-20 analysis with **Quiz item 10** unticked, leaving the other 19 items ticked. Predict the result before you click: the app told you it would be .78 without the item. Does it match? Then compare the new “What this means” box with the original. Which item is now the weakest, and is its item-total correlation still positive?

CHAPTER 16

Yes/no data, four different ways

The question

Many survey and school questions have only a few possible answers: pass or fail, yes or no, Education or Psychology or Business. You cannot average those answers, so the tests in the earlier chapters (which compare averages) do not apply. Instead you count how many people gave each answer and ask whether the counts show a pattern that is more than chance.

This chapter runs four tests for that situation on one dataset, and it is really one question asked four ways, depending on how your yes/no data was collected:

1. **Do pass rates differ between programs?** Two category questions, each person in one group. Test: **chi-square test of independence**.
2. **Is using extra credit linked to passing?** The same kind of question, but one answer is rare, so there are too few people in some cells to trust chi-square. Test: **Fisher's exact test**.
3. **Did more students use office hours after the midterm than before it?** The same people asked the same yes/no question twice. Test: **McNemar test**.
4. **Did use of study group, tutor and writing center differ?** The same people asked three yes/no questions. Test: **Cochran's Q test**.

The skill to take away is choosing the right test from how the data was collected. Statversity's Test Advisor does that for you, and in the second analysis you will see it change its own mind part-way through.

The data

The practice dataset is **Yes/no data, four different ways**. It is a single file, `survey.csv`, with **300 students** and a handful of yes/no and category questions:

- **Q2**: What is your program? (Education, Psychology or Business)
- **Q3**: Did you pass the course? (Fail or Pass)
- **Q4**: Did you use an extra-credit opportunity? (No or Yes)
- **Q5_pre** and **Q5_post**: Did you use office hours before the midterm? After the midterm?
- **Q6_1**, **Q6_2**, **Q6_3**: Did you use a study group this term? A tutor? The writing center?

To open it, click **Open practice data** on the Home screen and find the card titled **Yes/no data, four different ways**.

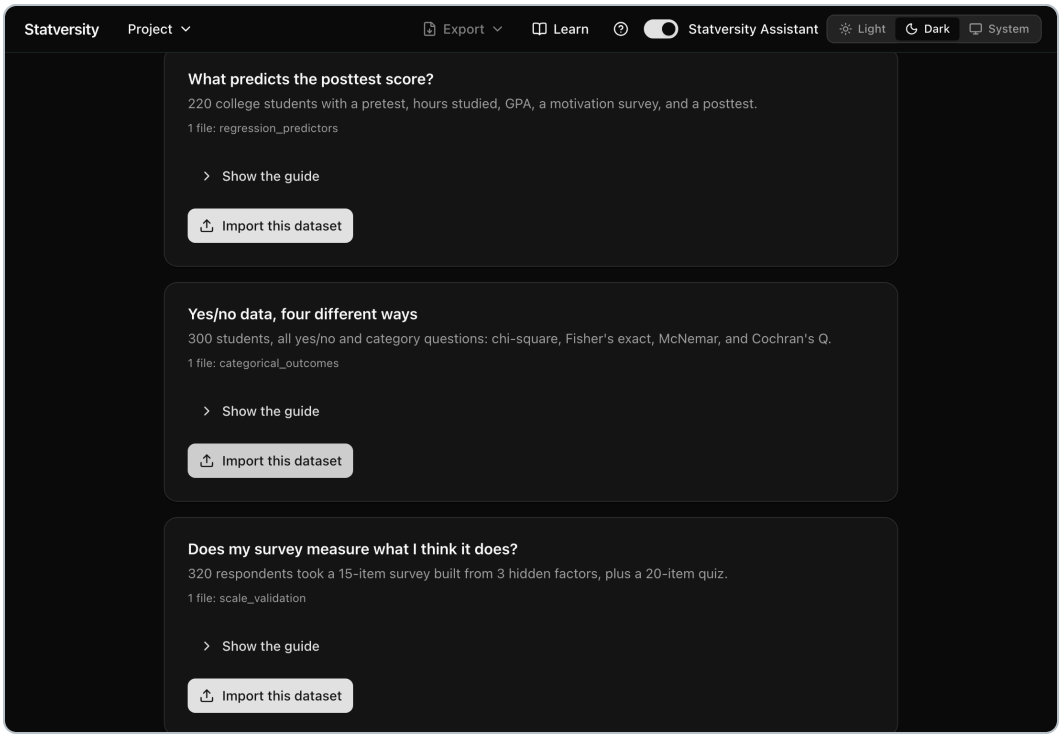


Figure 221. The Practice data screen with the card Yes/no data, four different ways: 300 students, all yes/no and category questions.

Steps: getting the data in

This part is quick, because the file is clean. Chapters 5 and 6 explain every screen in detail.

1. On the card, click **Import this dataset**. The screen shows **Choose files**.
2. Click **Read files**. The screen shows **About your study**.
3. Click **Skip these questions**. The screen shows **Check how we read them**.
4. Click **Continue**. The screen shows **Survey clean-up**. There is nothing to choose here for this dataset, because there are no personal-information columns or test responses to deal with.
5. Click **Continue**. The screen shows **Review and import**.

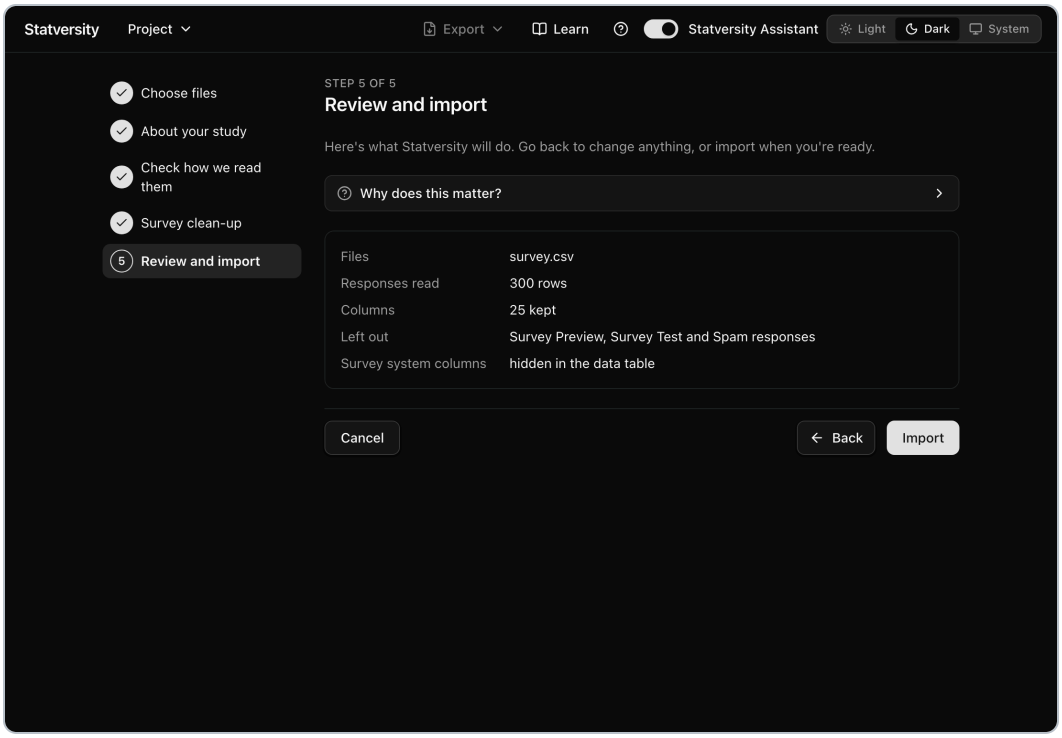


Figure 222. Review and import showing survey.csv, 300 rows read, 25 columns kept.

The review lists the file **survey.csv**, **300 rows** read and **25 columns** kept.

6. Click **Import**, then click **Start** on the **Welcome** screen. The Variable Interview begins with **What is each question?**

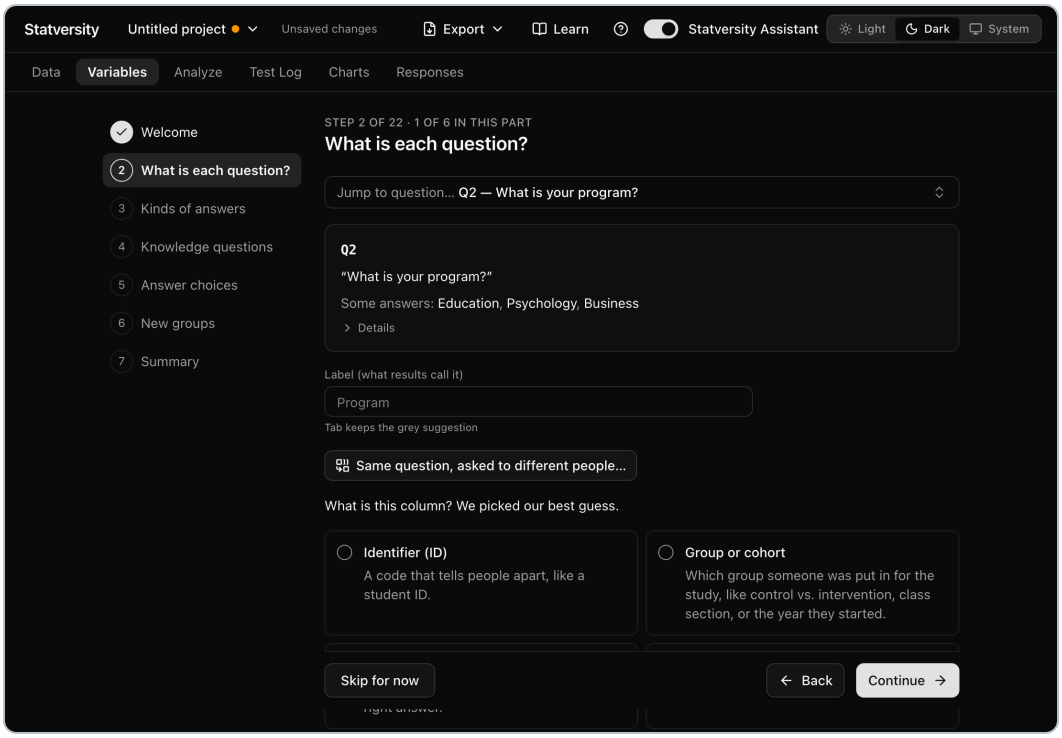


Figure 223. What is each question? showing Q2, What is your program?, with the answers Education, Psychology, Business, a suggested label Program, and a menu of kinds of question.

For each question you see its wording, a sample of its answers, a suggested label (here **Program**), and Statversity's best guess about what kind of column it is. When in doubt, accept Statversity's guess.

7. Click **Continue** through the interview, checking each screen. There are six **What is each question?** screens, six **Kinds of answers** screens, **Knowledge questions**, six **Answer choices** screens, **New groups**, and a **Summary**. Nothing needs changing. On **Knowledge questions**, none of these are quiz questions.
8. On the **Summary** screen, click **Finish**. The screen shows **Variables**.

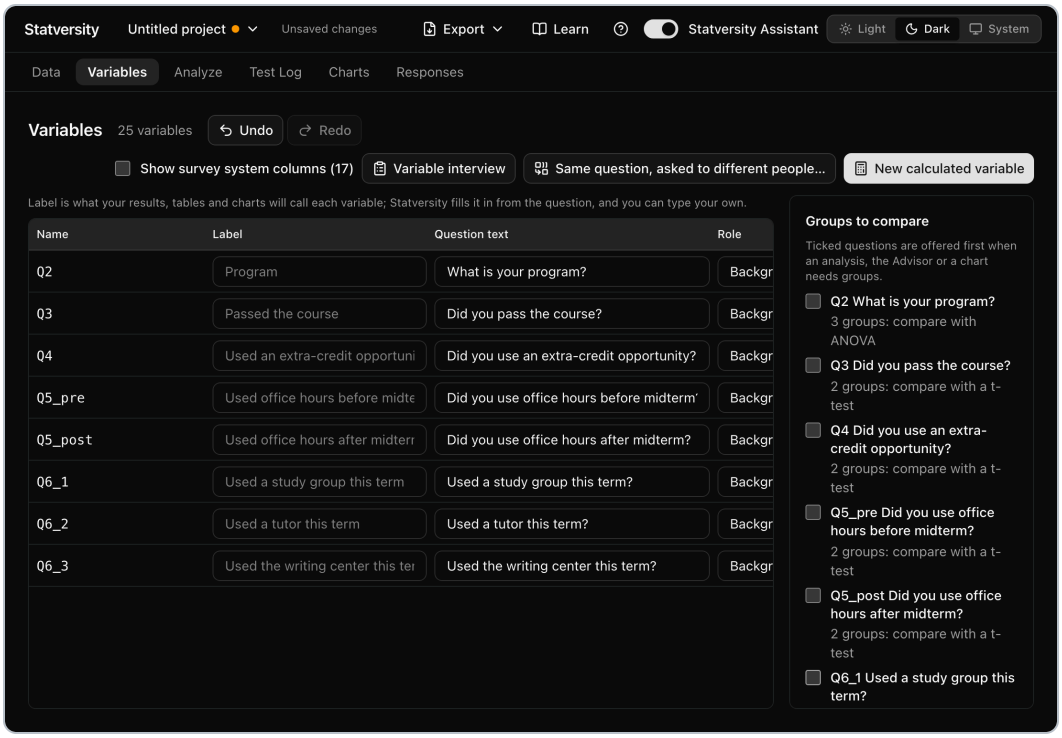


Figure 224. The Variables table with Q2 to Q6_3 as Background / demographic and nominal, with their labels such as Program, Passed the course and Used a tutor this term.

The Variables screen shows 25 variables. Each question you will use is listed as **nominal**, meaning its answers are categories with no natural order or size (a program is not "more" than another). The **Label** column gives each its short name (**Program**, **Passed the course**, **Used an extra-credit opportunity**, and so on). The small badge in **Value labels** shows how many answers each has: 3 for Program, 2 for each yes/no.

Analysis 1: Do pass rates differ between programs? (chi-square)

Steps

1. Click **Analyze**. The Test Advisor asks **What do you want to know?**
2. In **Which score or answer do you want to analyse?**, choose **Did you pass the course? (Q3)**.
3. Under **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**.

IN PLAIN LANGUAGE

Despite its wording, this option covers any comparison. “Are some people different from others?” includes “do students in one program pass more often than students in another?” The alternatives ask about relationships between two measured quantities, predicting one thing from others, or checking a scale, none of which is what you are doing here.

4. Under **What do your yes/no or pick-one answers look like?**, choose **Do the answers differ between groups?** and click **Continue**.

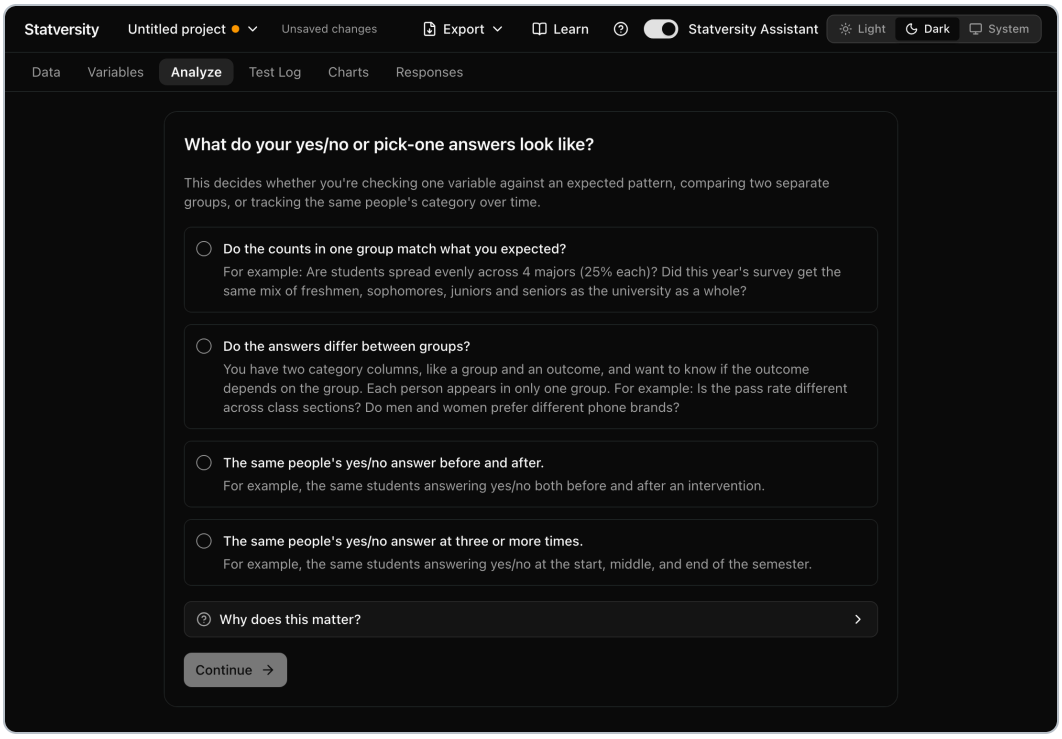


Figure 225. The Advisor question What do your yes/no or pick-one answers look like? with four options. The choice Do the answers differ between groups? explains that you have two category columns and each person appears in only one group.

IN PLAIN LANGUAGE

This question is really asking: **who are the people in each group, and are they the same people?** The four options are four different study designs:

- *Do the counts in one group match what you expected?* One category column, compared with a pattern you had in mind (for example, "is the class spread evenly over four majors?").
- *Do the answers differ between groups?* Two category columns, such as a group and an outcome. **Each person appears in only one group.** That is your situation: each student is in one program and either passed or failed.
- *The same people's yes/no answer before and after.* The same people measured twice (you will use this in Analysis 3).
- *The same people's yes/no answer at three or more times.* The same people measured three or more times (Analysis 4).

The right choice is about the design, not the data values.

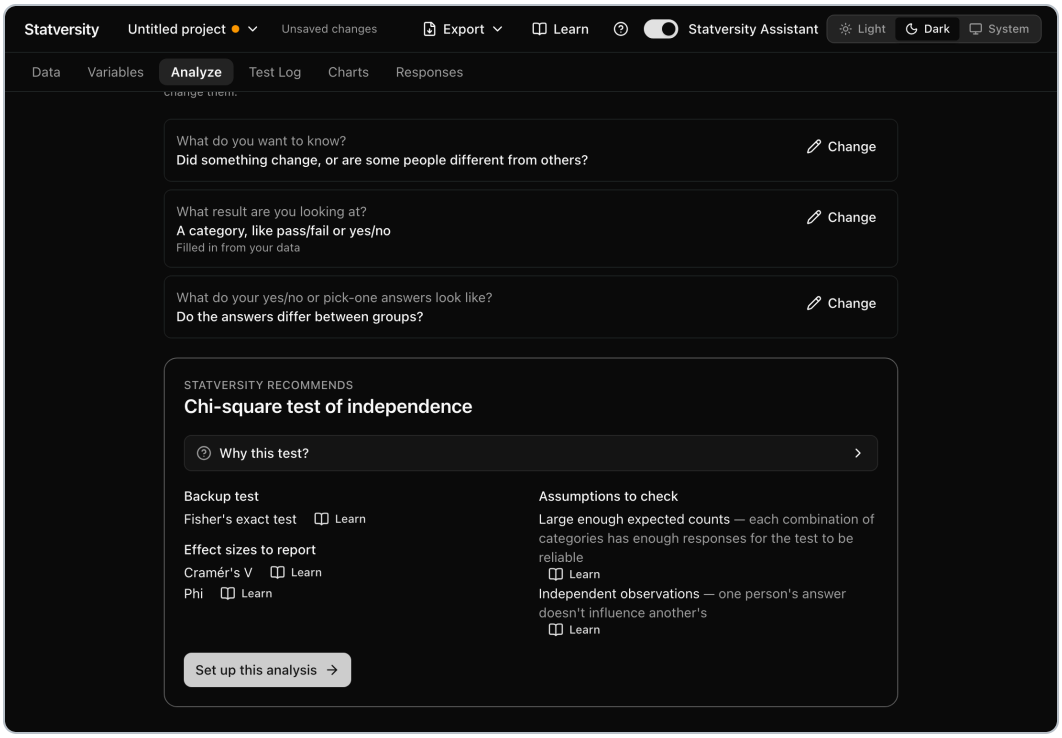


Figure 226. The Advisor recommending the Chi-square test of independence, with Fisher's exact test as a backup test, Cramér's V and Phi as effect sizes, and expected counts and independent observations as assumptions.

Statversity recommends the **chi-square test of independence**. The card also names a **Backup test** (Fisher's exact test), the **Effect sizes to report** (Cramér's V and Phi), and **Assumptions to check**: large enough expected counts and independent observations. Independence here means one student's answer does not influence another's, which is true by design in a survey where each student answers alone.

5. Click **Set up this analysis**. The screen shows **Choose your variables**.

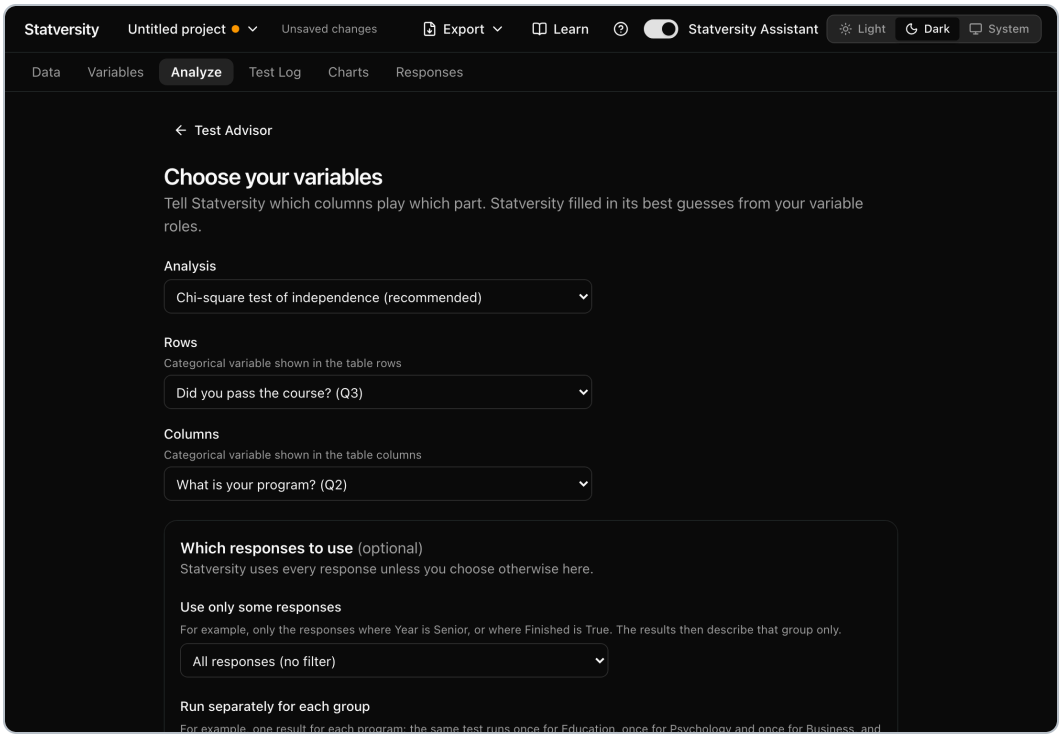


Figure 227. Choose your variables with Chi-square test of independence selected, Rows set to Did you pass the course? (Q3), and Columns set to What is your program? (Q2).

6. Check that **Rows** already shows **Did you pass the course? (Q3)**. Statversity filled it in.
7. Check that **Columns** already shows **What is your program? (Q2)**. No change needed.

The rows and columns choice only decides how the table is laid out. The test gives the same answer either way.

The expected count check

8. Click **Check the assumptions**. A check titled **Expected cell counts: Passed the course × Program** appears.

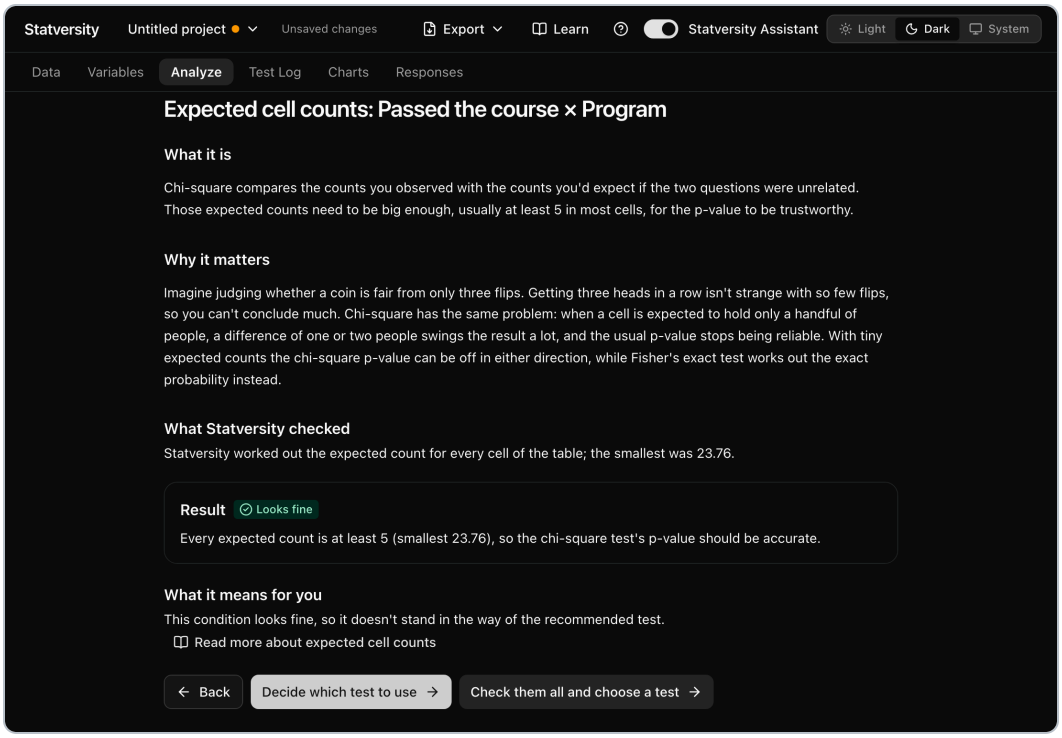


Figure 228. The expected cell counts check, with a Looks fine verdict: every expected count is at least 5 (smallest 23.76), so the chi-square p-value should be accurate.

This check introduces an idea you need for the rest of this chapter, so it is worth slowing down.

> IN PLAIN LANGUAGE
>
> **What is an "expected count"?** Chi-square works by asking what the table would look like if the two questions were completely unrelated, so that passing had nothing to do with program. In this data, 99 of 300 students failed, which is 33%. If program made no difference, you would expect about 33% of each program to fail. The Business program has 72 students, so you would expect about 33% of 72, or 23.76, to fail. That number, 23.76, is the **expected count** for the "Business and Fail" cell. It is not a count of real people (you cannot have 0.76 of a student); it is the count that "no relationship" would predict.
>
> Chi-square then compares each cell's real count with its expected count. The bigger the gaps overall, the stronger the evidence of a relationship.
>
> **Why do expected counts need to be big enough?** The screen uses a coin analogy: three heads in three flips is not unusual, so you cannot conclude the coin is unfair from three flips. In the same way, when a cell is only expected to hold a handful of people, one or two people more or fewer swings the result a lot, and the usual p-value stops being reliable. A common rule of thumb is that the expected count should be at least 5 in most cells.

Here the **smallest expected count is 23.76**, comfortably above 5, so the verdict is **Looks fine**: “Every expected count is at least 5 (smallest 23.76), so the chi-square test’s p-value should be accurate.”

9. Click **Decide which test to use**. **Which test should you use?** appears.

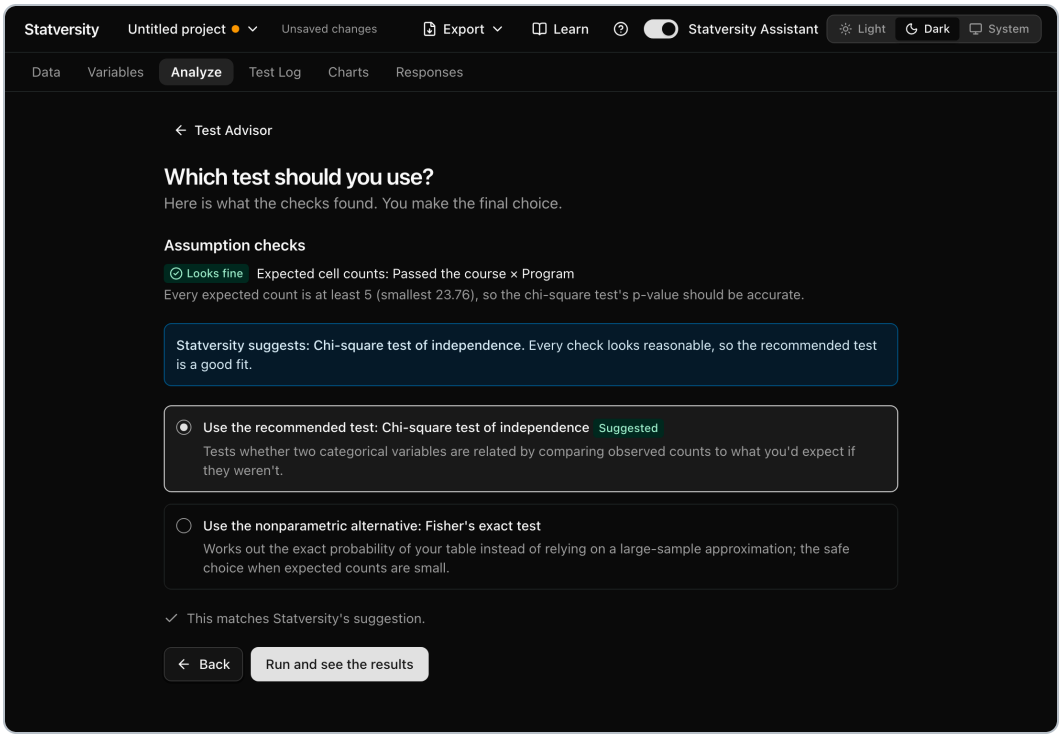


Figure 229. Which test should you use? with Chi-square test of independence selected and labelled Suggested, and Fisher's exact test as the alternative.

Statversity suggests the chi-square test: “Every check looks reasonable, so the recommended test is a good fit.” The alternative, **Fisher’s exact test**, is described as “the safe choice when expected counts are small”. You make the final choice; the recommended test is already selected.

10. Click **Run and see the results**. The Results screen for **Chi-square test of independence** appears.

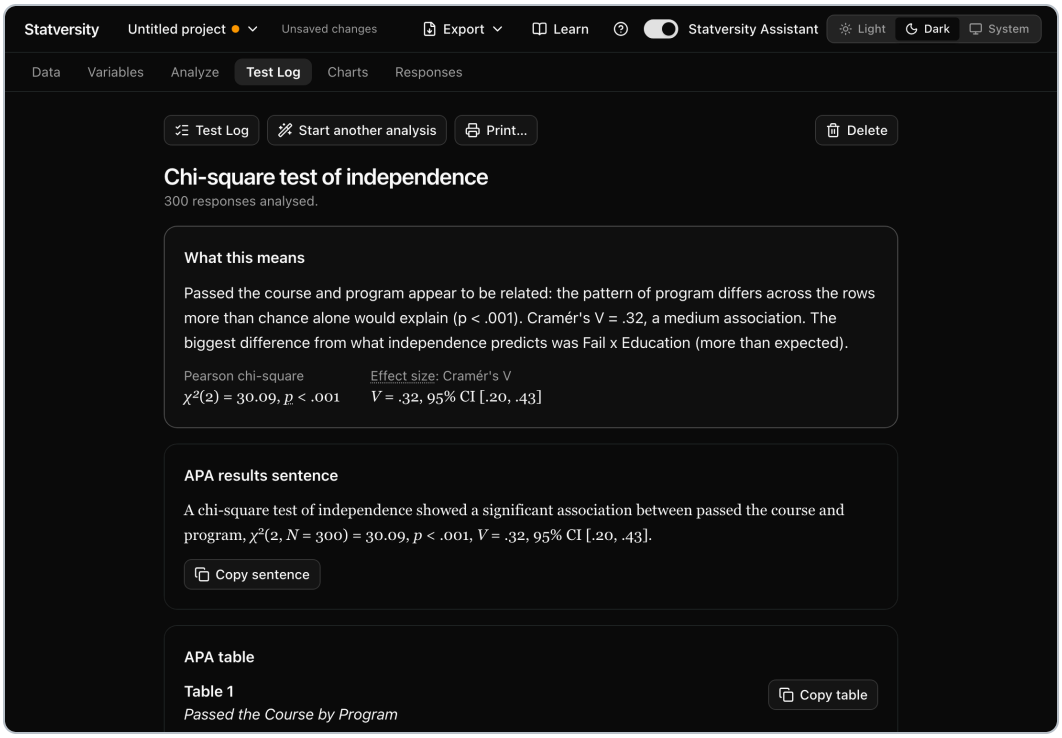


Figure 230. The Chi-square test of independence result for 300 responses, with the What this means box: chi-square(2) = 30.09, p < .001, Cramér's V = .32, a medium association.

What Statversity found

Test	Chi-square test of independence
Responses analysed	300
Statistic	$\chi^2(2) = 30.09$
p	< .001
Effect size	Cramér's V = .32, 95% CI [.20, .43], medium
Biggest difference from independence	Fail in Education (more than expected)

The APA table lays out the real counts.

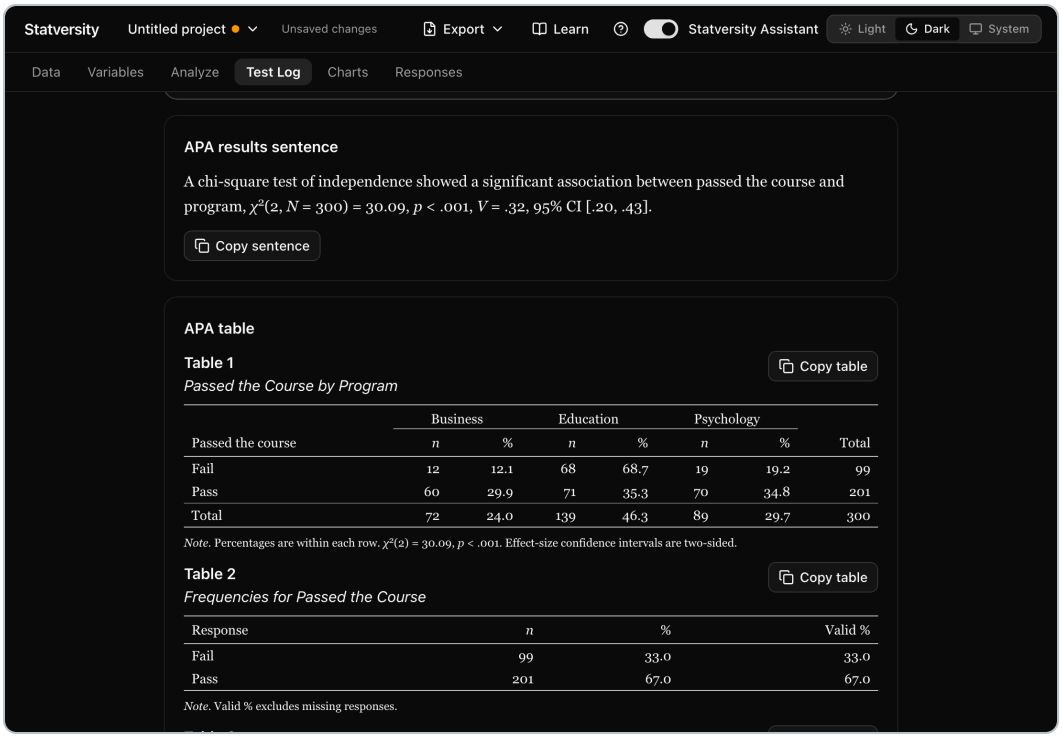


Figure 231. The APA sentence and Table 1, Passed the Course by Program: Business 12 Fail and 60 Pass, Education 68 Fail and 71 Pass, Psychology 19 Fail and 70 Pass.

Program	Fail	Pass	Total
Business	12	60	72
Education	68	71	139
Psychology	19	70	89
All	99	201	300

In Statversity's table the rows are the outcome (Fail and Pass, as you chose under **Rows**) and the percentages are within each of those rows, so the 68.7% beside Education's 68 means that 68.7% of all failures were in Education. To compare pass rates, divide each program's Pass count by its Total. Doing that arithmetic on the table gives roughly 83% for Business (60 of 72), 79% for Psychology (70 of 89), and 51% for Education (71 of 139).

The descriptive statistics, and the expected cell counts check repeated on the result, are further down the page.

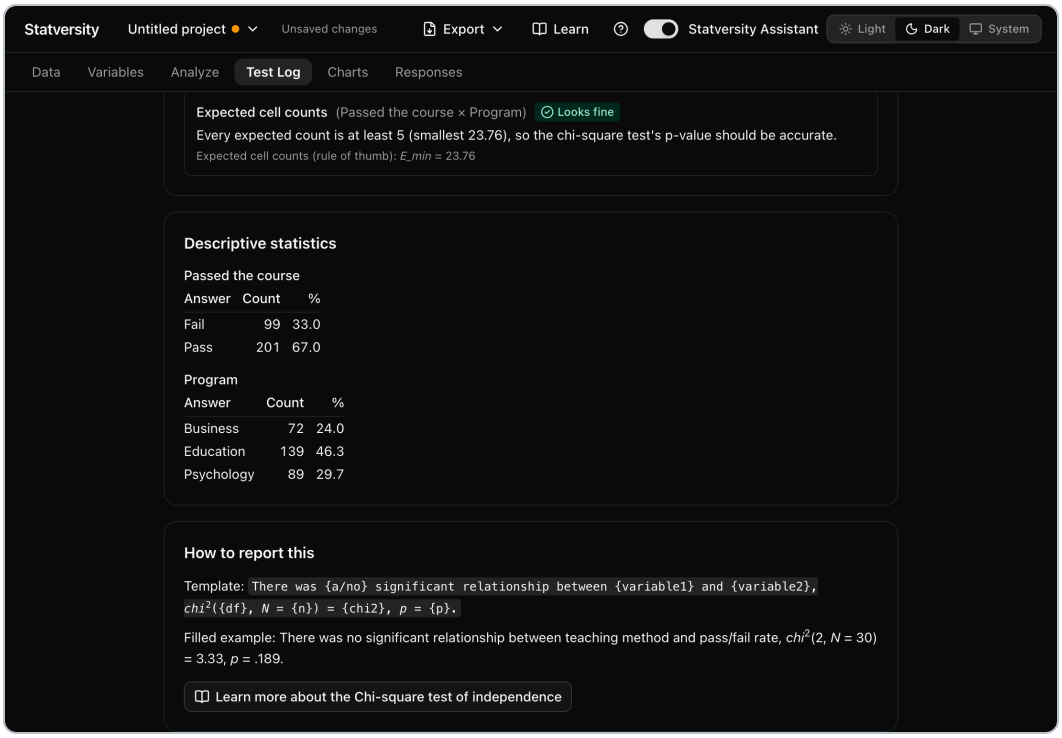


Figure 232. The result's Assumption checks showing Looks fine for expected cell counts, and Descriptive statistics tables for Passed the course (Fail 99, 33.0%; Pass 201, 67.0%) and Program (Business 72, Education 139, Psychology 89).

What this means in plain English

Passing the course and program are related. The three programs do not have the same pass rate: the differences between programs are larger than chance alone would produce ($p < .001$). Cramér's V is .32, which Statversity labels a medium association. (Cramér's V is the effect size for category-by-category tables: 0 means no relationship and 1 means a perfect one, and it tells you how strong the pattern is, where the p-value tells you only whether a pattern this strong could easily arise by chance.) The pattern is driven mostly by Education: far more Education students failed than the "no relationship" model predicts.

The statistic $\chi^2(2) = 30.09$ has **2 degrees of freedom**, which comes from the table's shape: three programs and two outcomes give $(3 - 1) \times (2 - 1) = 2$. You do not need to compute that; just report it.

What you would write in a paper

> A chi-square test of independence showed a significant association between passed the course and program, $\chi^2(2, N = 300) = 30.09$, $p < .001$, $V = .32$, 95% CI [.20, .43].

To adapt it, replace the variable names with your own wording (for example "course outcome and program of study") and add a sentence giving the pass rates for each program, since the test itself does not say which program differs.

Analysis 2: Is using extra credit linked to passing? (Fisher's exact)

Steps

1. Click **Start another analysis**. The Test Advisor asks **What do you want to know?**
2. In **Which score or answer do you want to analyse?**, choose **Did you pass the course? (Q3)**.
3. Under **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**.
4. Under **What do your yes/no or pick-one answers look like?**, choose **Do the answers differ between groups?** and click **Continue**.

This is the same set of answers as the first analysis, because the design is the same: two category columns, each person in one group.

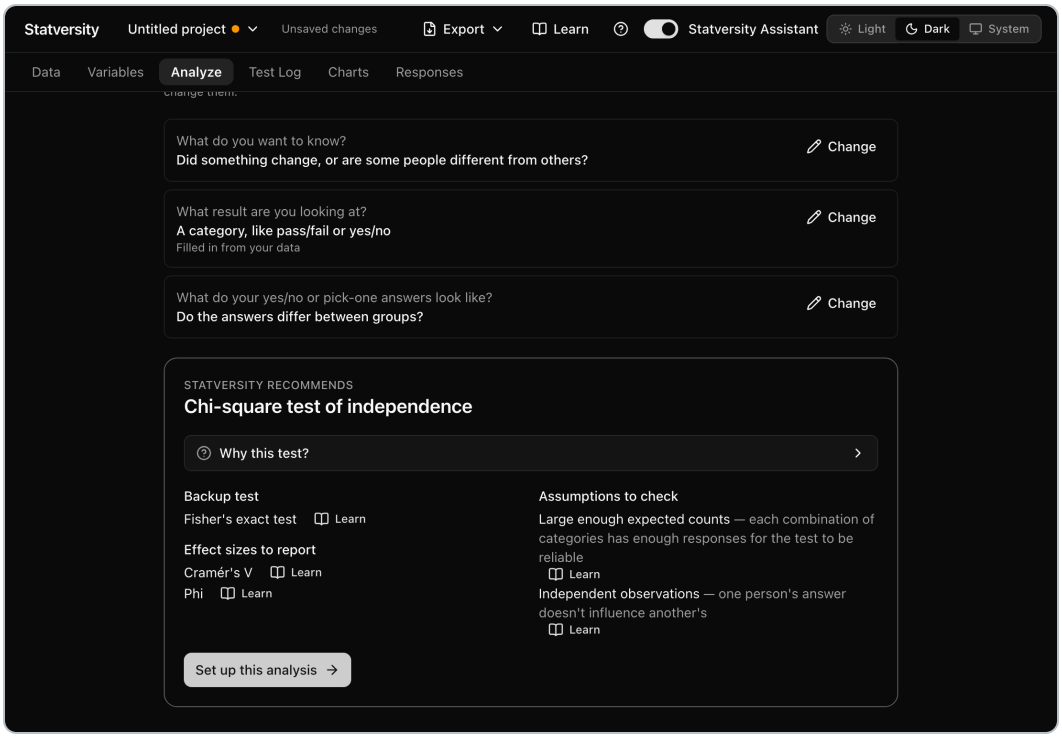


Figure 233. The Advisor again recommending the Chi-square test of independence, with Fisher's exact test as the backup test.

The Advisor again recommends **Chi-square test of independence**. Its recommendation is based on the design, which it can see before it looks at the counts. The counts come next.

5. Click **Set up this analysis**. The screen shows **Choose your variables**.
6. In **Rows**, choose **Did you use an extra-credit opportunity? (Q4)**.

7. In **Columns**, choose **Did you pass the course? (Q3)**.

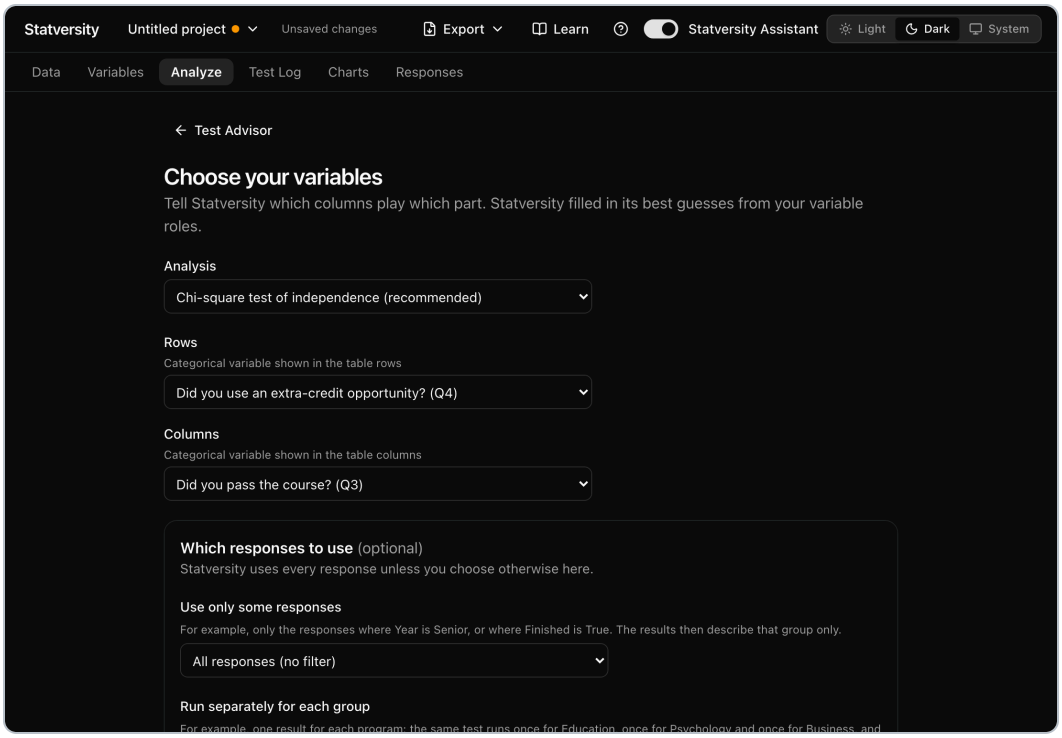


Figure 234. Choose your variables with Chi-square test of independence (recommended), Rows set to Did you use an extra-credit opportunity? (Q4) and Columns set to Did you pass the course? (Q3).

The same check, a different verdict

8. Click **Check the assumptions**. The check **Expected cell counts: Used an extra-credit opportunity × Passed the course** appears.

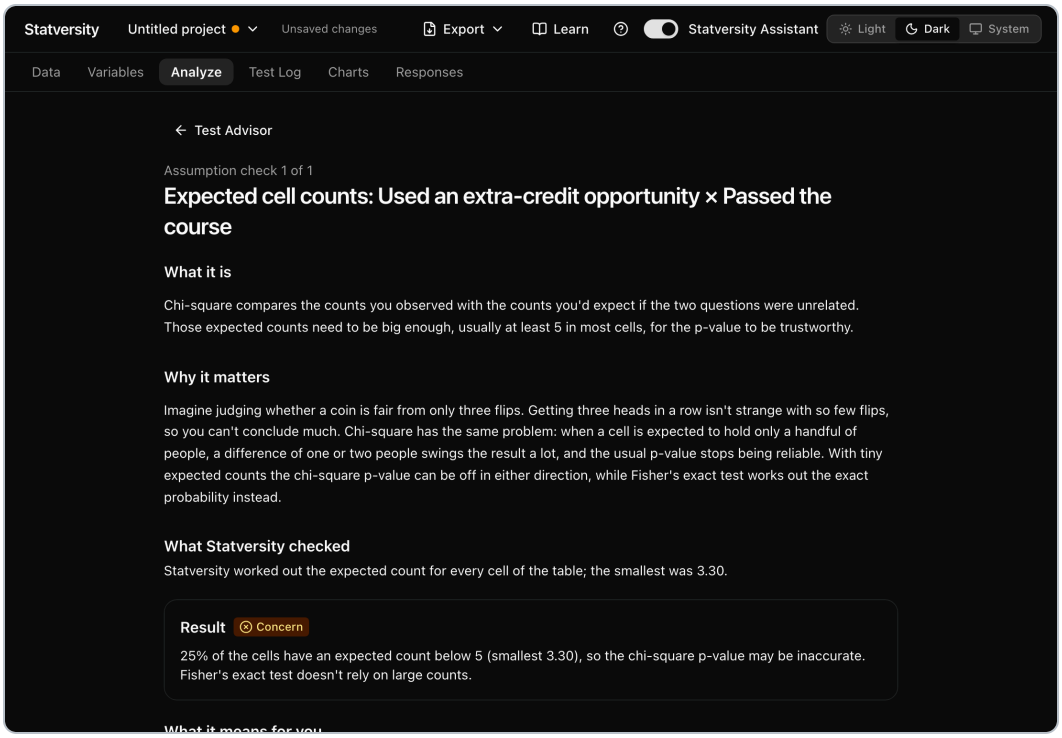


Figure 235. The expected cell counts check with a Concern verdict: 25% of the cells have an expected count below 5 (smallest 3.30), so the chi-square p-value may be inaccurate.

Now the verdict is **Concern**: “25% of the cells have an expected count below 5 (smallest 3.30), so the chi-square p-value may be inaccurate. Fisher’s exact test doesn’t rely on large counts.”

> **IN PLAIN LANGUAGE**
>
> Where does 3.30 come from? Almost nobody used extra credit: only 10 of the 300 students said Yes. Of the 300 students, 99 failed. If extra credit and passing were unrelated, you would expect the 10 extra-credit students to fail at the overall rate, 99 out of 300, which gives $10 \times 99 \div 300 =$ **3.3** people. A cell expected to hold only 3.3 people is a very small cell. The table has four cells (Yes or No for extra credit, by Fail or Pass), and one of them, “Yes and Fail”, is below 5, which is **25%** of the cells.
>
> So the question is not about your sample as a whole (300 is plenty) but about the smallest cell. When a group is rare, even a big sample leaves a tiny cell.

The Advisor recommended chi-square a moment ago because it could only see the design. Now Statversity has looked at the actual counts and found that the recommended test would be unreliable here. This is what the assumption check is for.

9. Click **Decide which test to use**. **Which test should you use?** appears.

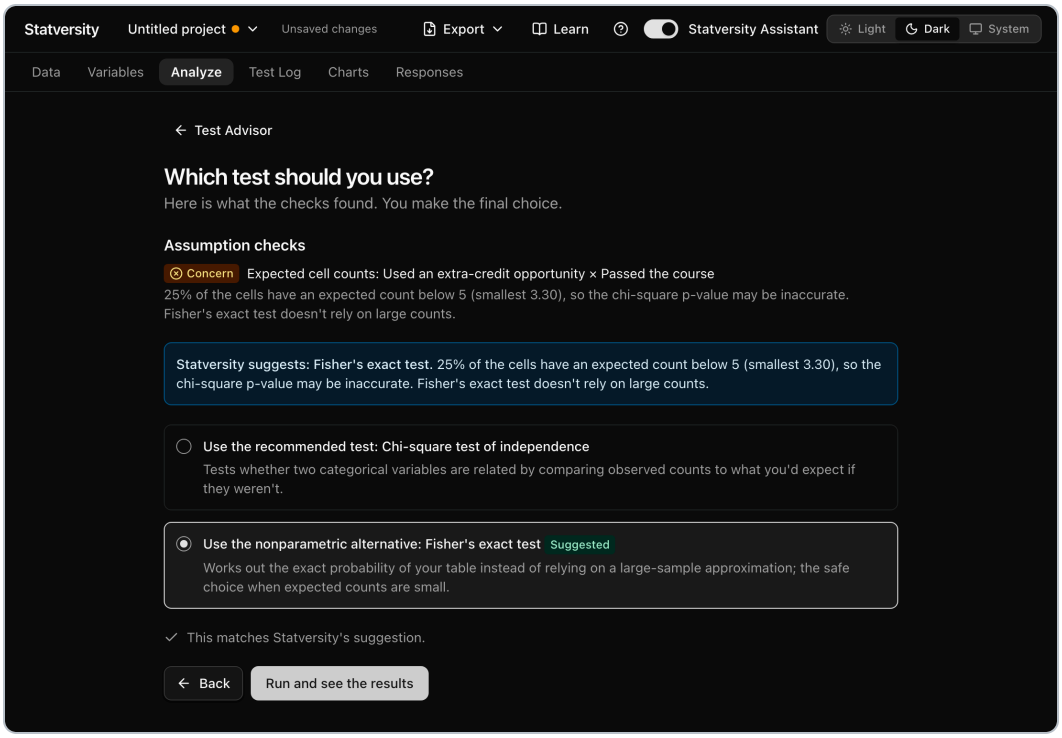


Figure 236. Which test should you use? with a Concern on expected cell counts, and Statversity suggesting Fisher's exact test. The Fisher's exact test option is selected and labelled Suggested.

Statversity now suggests **Fisher's exact test**: "25% of the cells have an expected count below 5 (smallest 3.30), so the chi-square p-value may be inaccurate. Fisher's exact test doesn't rely on large counts." On this screen the Fisher option is selected and marked **Suggested**, while the originally recommended chi-square is available but not selected.

> **IN PLAIN LANGUAGE**
>
> Chi-square uses a shortcut formula that works well when counts are large. **Fisher's exact test** does not use the shortcut. It works out the exact probability of seeing a table like yours (or more extreme) if the two variables were unrelated, so it stays accurate with small counts. It is the safe choice whenever a cell is small.

10. Leave Fisher's exact test selected and click **Run and see the results**.

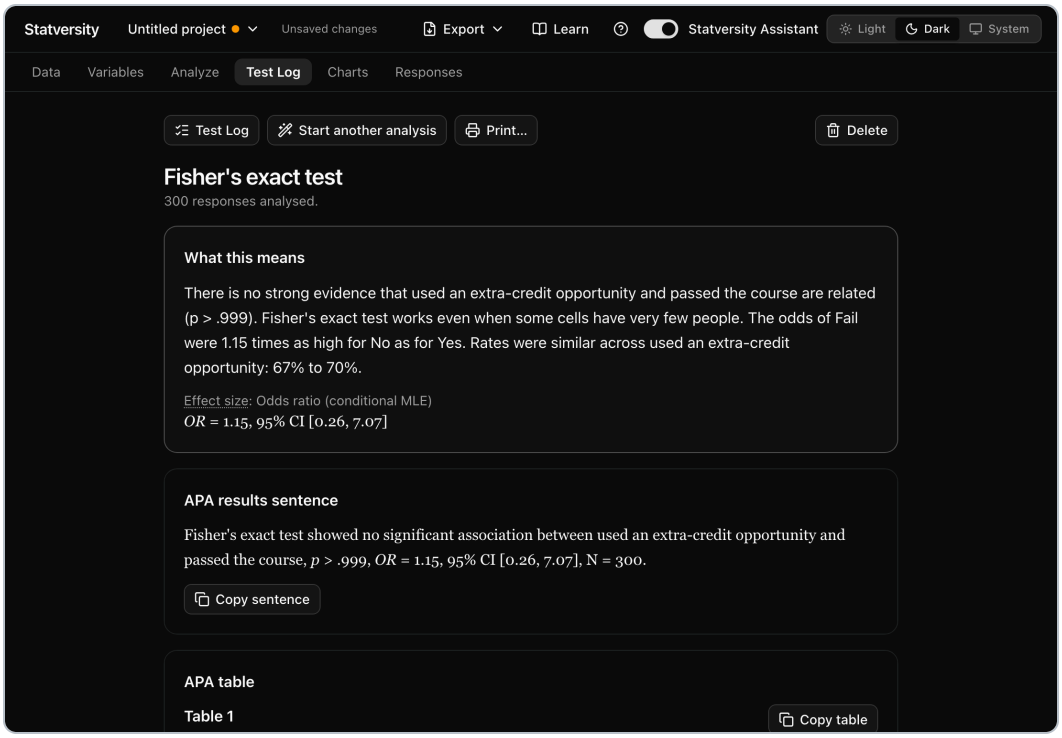


Figure 237. The Fisher's exact test result for 300 responses: There is no strong evidence that used an extra-credit opportunity and passed the course are related (p > .999), the odds ratio is 1.15, 95% CI [0.26, 7.07].

What Statversity found

Test	Fisher’s exact test
Responses analysed	300
p	> .999
Effect size	Odds ratio = 1.15, 95% CI [0.26, 7.07]
Also reported	Phi = .01, 95% CI [.00, .11], negligible
Rates	67% to 70% passed in both groups

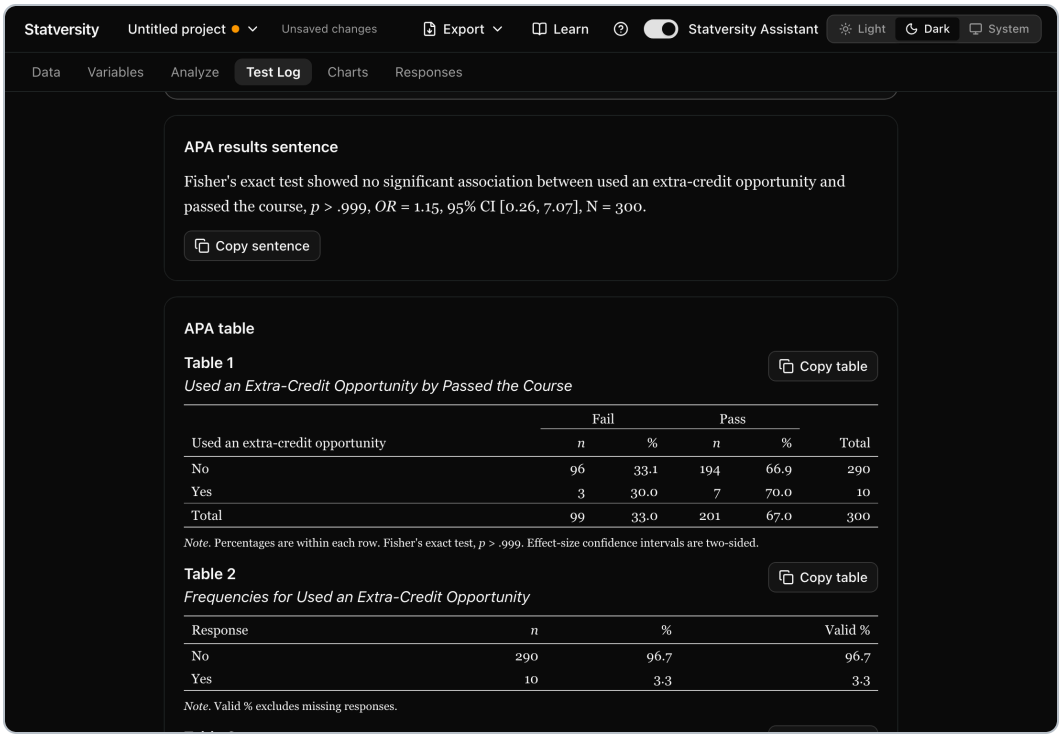


Figure 238. The APA sentence and Table 1, Used an Extra-Credit Opportunity by Passed the Course: No, 96 Fail and 194 Pass; Yes, 3 Fail and 7 Pass.

Used extra credit	Fail	Pass	Total
No	96 (33.1%)	194 (66.9%)	290
Yes	3 (30.0%)	7 (70.0%)	10

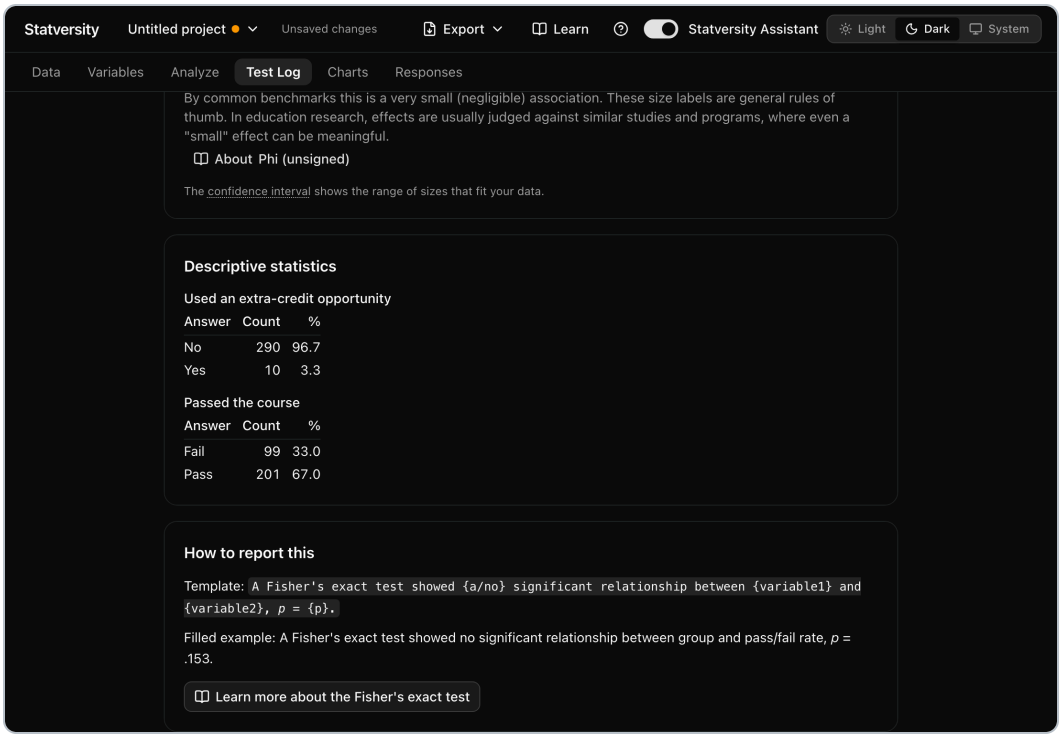


Figure 239. Effect sizes for the Fisher result, Odds ratio (conditional MLE) 1.15, Odds ratio (sample) 1.15 [0.29, 4.56] and Phi 0.01 labelled negligible, followed by Descriptive statistics.

What this means in plain English

There is no evidence that using extra credit is linked to passing. Among the students who used extra credit, 7 of 10 passed (70%); among those who did not, 194 of 290 passed (66.9%). Those rates are nearly the same, and the p-value, greater than .999, means the data are entirely consistent with there being no link.

Notice the **odds ratio**. It is 1.15 with a 95% confidence interval of **[0.26, 7.07]**, which is enormously wide. The odds ratio compares the odds of failing in one group with the odds in the other, with 1 meaning no difference. A range running from about a quarter to seven means the data cannot rule out a large effect in either direction, because only 10 people used extra credit. “No significant association” here means “no evidence either way”, not “proven to be unrelated”. The honest conclusion is that you would need far more extra-credit users to say anything firm.

What you would write in a paper

> Fisher’s exact test showed no significant association between used an extra-credit opportunity and passed the course, $p > .999$, $OR = 1.15$, 95% CI [0.26, 7.07], $N = 300$.

When you report this, say why you used Fisher’s exact test rather than chi-square: “because one expected cell count (3.3) was below 5”. And mention how few students used extra credit (10), because the wide confidence interval needs that context.

Analysis 3: Did office-hours use change? (McNemar)

Steps

1. Click **Start another analysis**.
2. In **Which score or answer do you want to analyse?**, choose **Did you use office hours after midterm? (Q5_post)**.
3. Under **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**.
4. Under **What do your yes/no or pick-one answers look like?**, choose **The same people's yes/no answer before and after.** and click **Continue**.

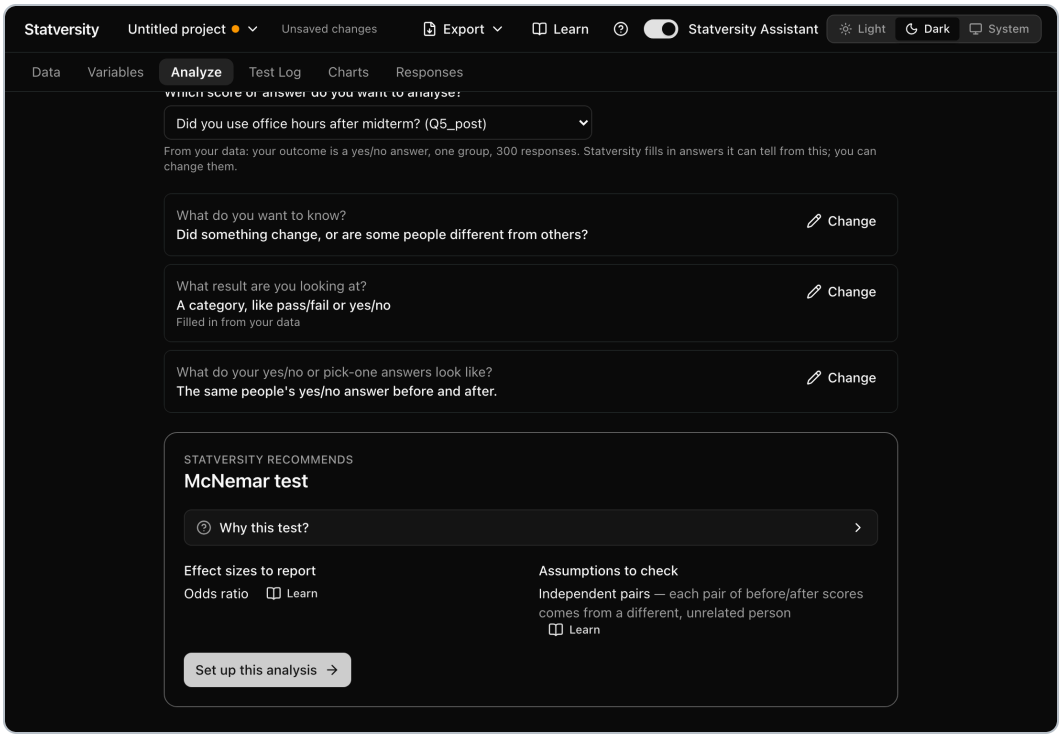


Figure 240. The Advisor recommending the McNemar test, with Odds ratio as an effect size and the independent pairs assumption.

> **IN PLAIN LANGUAGE**
>
> This time the same 300 students answered the question twice, once about the period before the midterm and once about after it. That makes the two answers **paired**: the "before" and "after" answers of one student belong together. A chi-square test treats every answer as coming from a different person, which would be wrong here, so Statversity switches to the **McNemar test**, which is designed for paired yes/no answers. It ignores the students whose answer did not change and asks only whether switches in one direction (No to Yes) are more common than switches in the other (Yes to No).

Statversity recommends the **McNemar test**. Its "Assumptions to check" box has a line about independent pairs, "each pair of before/after scores comes from a different, unrelated person". Read that as "each student's pair of answers is independent of the other students'", which holds here.

5. Click **Set up this analysis**. The screen shows **Choose your variables**.
6. Under **Measurements**, tick **Did you use office hours before midterm? (Q5_pre)** and **Did you use office hours after midterm? (Q5_post)**.

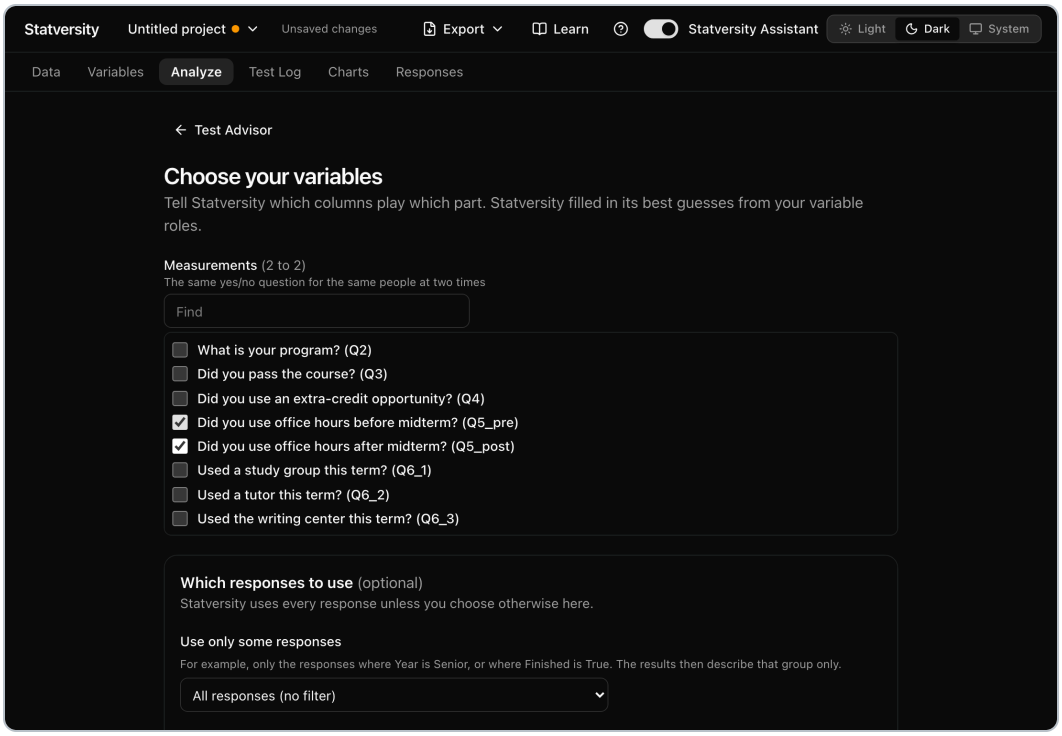


Figure 241. Choose your variables for McNemar with Measurements (2 to 2) and Q5_pre and Q5_post ticked.

The heading says **Measurements (2 to 2)**: "the same yes/no question for the same people at two times". Exactly two must be ticked.

7. Click **Check the assumptions**. The result for **McNemar test** appears directly, since there is no assumption check with a screen of its own for this test.

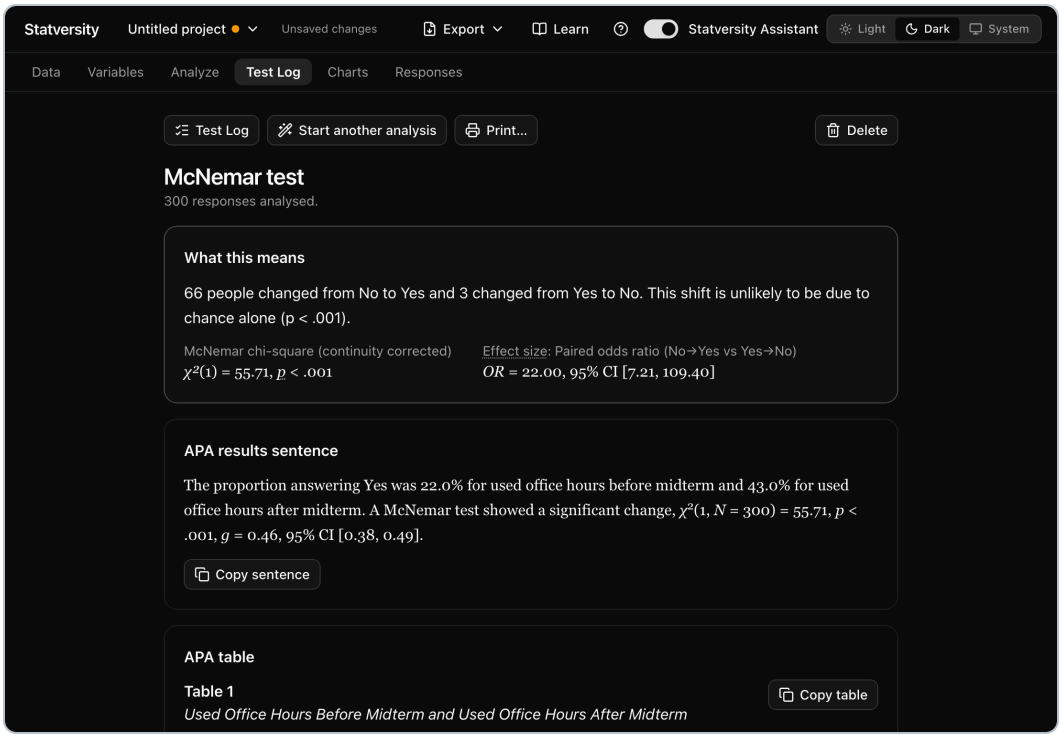


Figure 242. The McNemar test result for 300 responses: 66 people changed from No to Yes and 3 changed from Yes to No, chi-square(1) = 55.71, p < .001.

What Statversity found

Test	McNemar test (continuity corrected)
Responses analysed	300
Statistic	$\chi^2(1) = 55.71$
p	< .001
Effect size	Cohen's g = 0.46, 95% CI [0.38, 0.49]
Also reported	Paired odds ratio (No to Yes vs Yes to No) = 22.00, 95% CI [7.21, 109.40]
Share answering Yes	22.0% before, 43.0% after

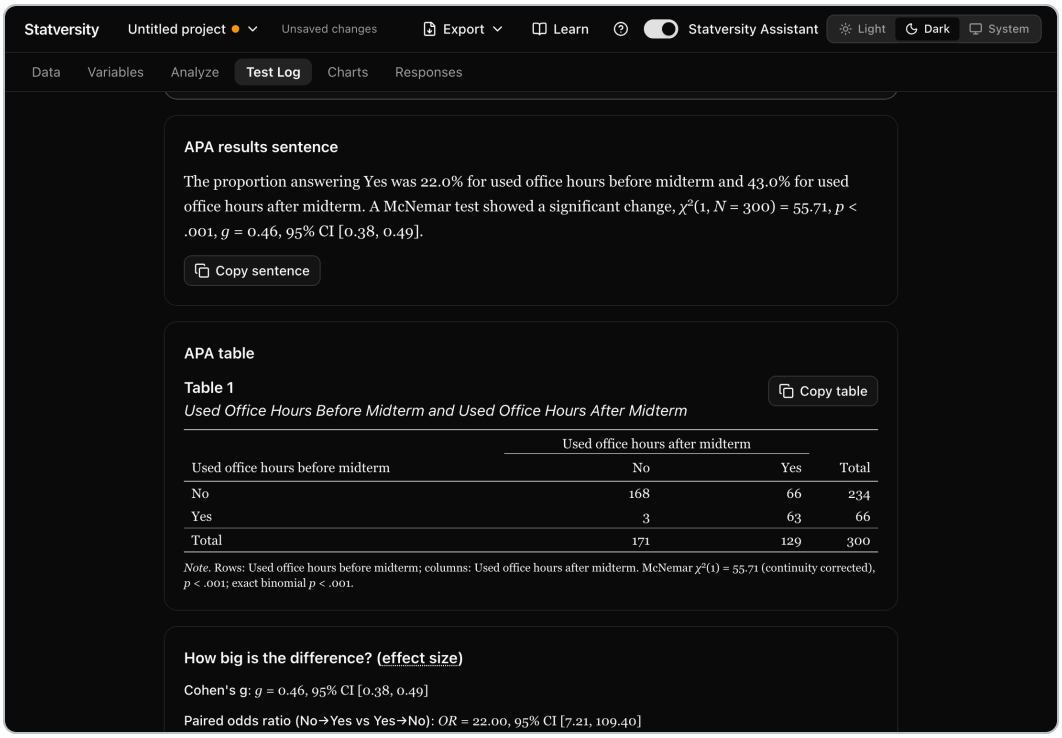


Figure 243. The APA sentence and Table 1: before No, after No 168, after Yes 66; before Yes, after No 3, after Yes 63; totals 234, 66, 171, 129, 300.

The key table cross-tabulates each student's "before" answer (rows) with their "after" answer (columns).

Before \ After	No	Yes	Total
No	168	66	234
Yes	3	63	66
Total	171	129	300

The two cells that matter are the ones where students **changed**: **66** went from No to Yes, and only **3** went from Yes to No. The 168 and 63 on the diagonal are students whose answer did not change, and McNemar ignores them.

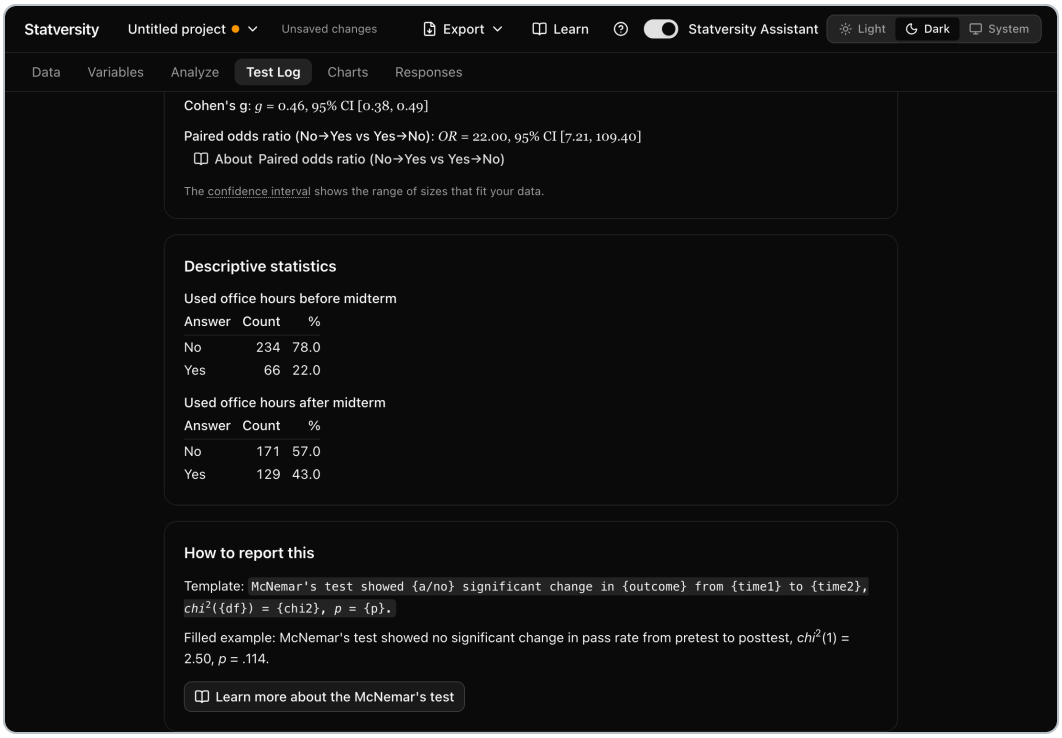


Figure 244. The effect sizes Cohen's g and the Paired odds ratio, and Descriptive statistics: before No 234 (78.0%) and Yes 66 (22.0%); after No 171 (57.0%) and Yes 129 (43.0%).

What this means in plain English

Many more students started using office hours after the midterm than stopped. The proportion saying Yes rose from 22% to 43%. Out of the 69 students who changed their answer, 66 changed toward using office hours, and only 3 changed away. If office-hours use were stable, you would expect switches in both directions to be about equally common, so a 66-to-3 split would be very unlikely by chance alone ($p < .001$). The paired odds ratio of 22.00 says the same thing in another form: switching to Yes was about 22 times as common as switching to No.

Notice that this analysis tells you that use changed, not why. The data cannot say whether the midterm, a reminder email, or something else prompted the change.

What you would write in a paper

> The proportion answering Yes was 22.0% for used office hours before midterm and 43.0% for used office hours after midterm. A McNemar test showed a significant change, $\chi^2(1, N = 300) = 55.71$, $p < .001$, $g = 0.46$, 95% CI [0.38, 0.49].

Adapt it by naming your two time points in your own words (“before the midterm” and “after the midterm”).

Analysis 4: Did use differ across three services? (Cochran's Q)

Steps

1. Click **Start another analysis**.
2. In **Which score or answer do you want to analyse?**, choose **Used a study group this term? (Q6_1)**.
3. Under **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**.
4. Under **What do your yes/no or pick-one answers look like?**, choose **The same people's yes/no answer at three or more times.** and click **Continue**.

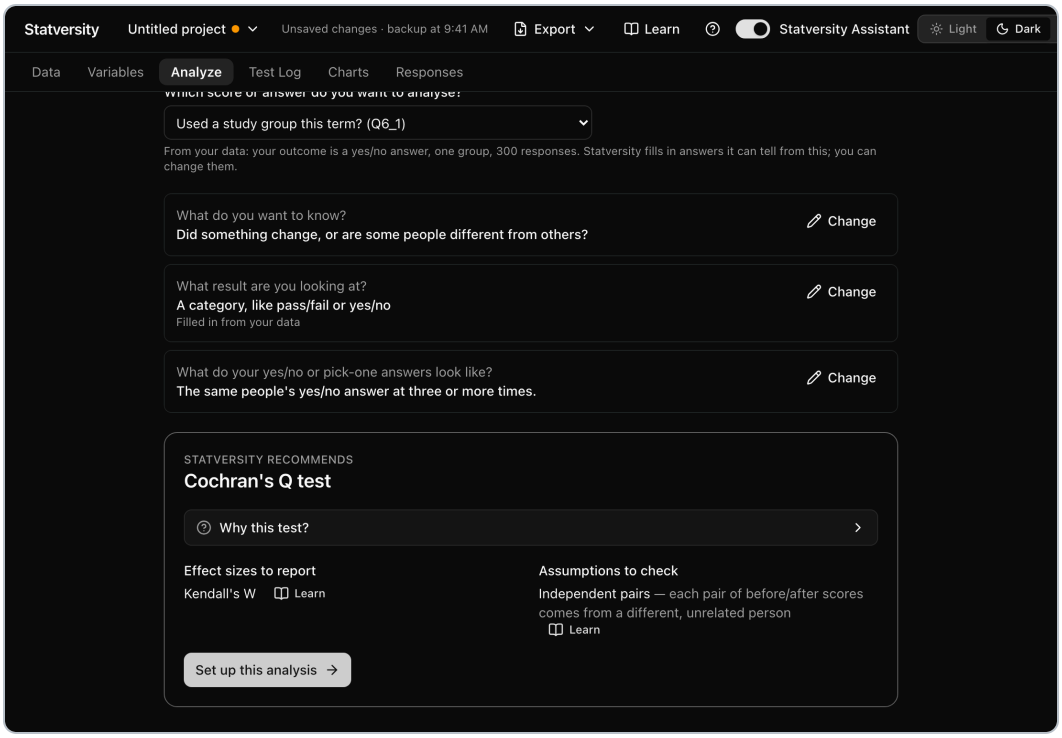


Figure 245. The Advisor recommending Cochran's Q test, with Kendall's W listed under effect sizes to report.

> **IN PLAIN LANGUAGE**
>
> Cochran's Q is McNemar's test extended to three or more yes/no measures on the same people. The wording of this option says "at three or more times", but the test does not care whether the three measures are three moments or three different questions. What matters is that the same 300 people answered each, so the answers are paired. Here they are three different services (study group, tutor, writing center), and the question is whether they are used equally often.

5. Click **Set up this analysis**. The screen shows **Choose your variables**.

6. Under **Measurements**, tick **Used a study group this term? (Q6_1)**, **Used a tutor this term? (Q6_2)** and **Used the writing center this term? (Q6_3)**.

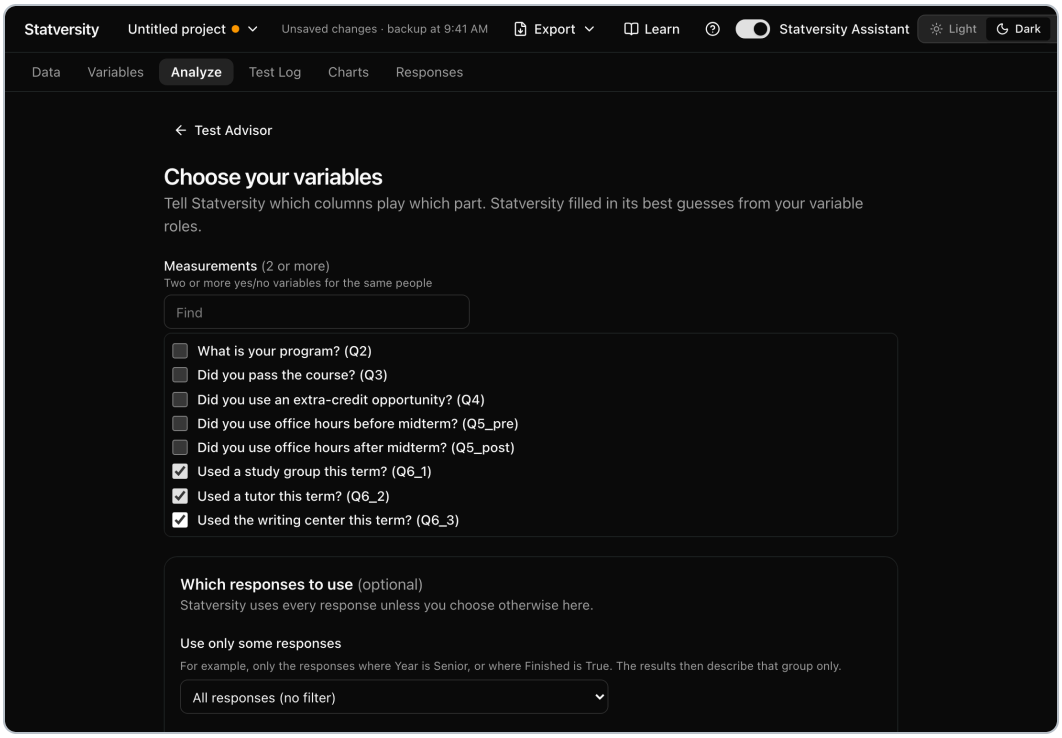


Figure 246. Choose your variables for Cochran's Q with Measurements (2 or more) and Q6_1, Q6_2 and Q6_3 ticked.

7. Click **Check the assumptions**. The result for **Cochran’s Q test** appears.

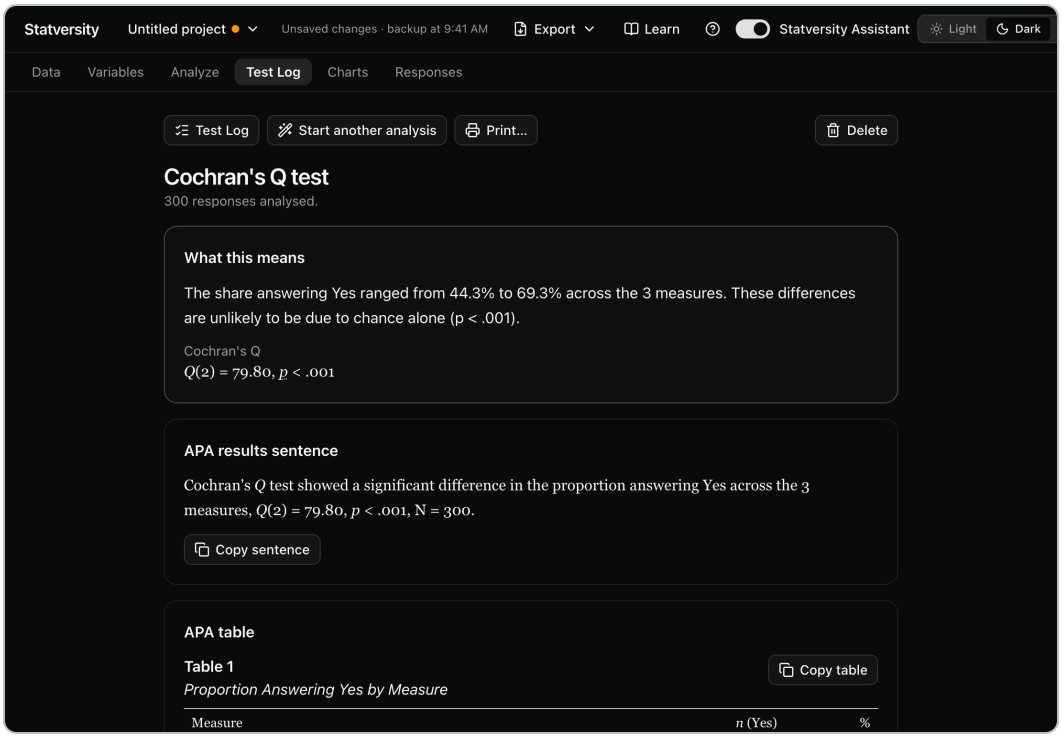


Figure 247. The Cochran's Q test result for 300 responses: the share answering Yes ranged from 44.3% to 69.3% across the 3 measures, Q(2) = 79.80, p < .001.

What Statversity found

Test	Cochran’s Q test
Responses analysed	300
Statistic	Q(2) = 79.80
p	< .001

Service	Said Yes (n)	%
Study group	133	44.3
Tutor	161	53.7
Writing center	208	69.3

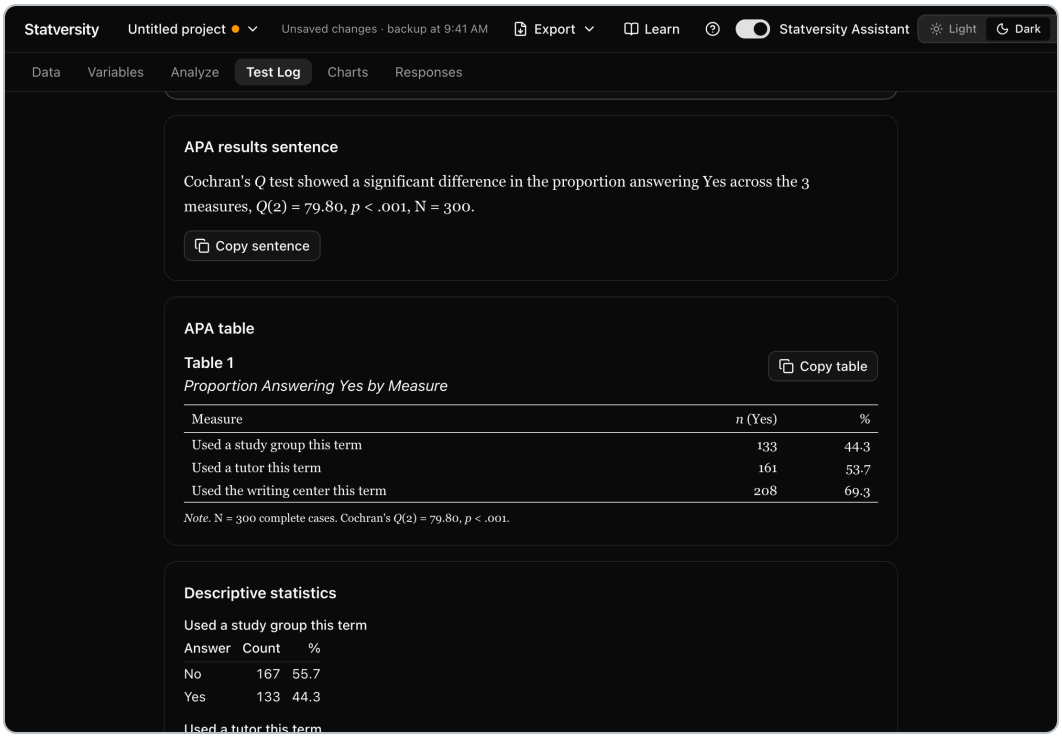


Figure 248. The APA sentence and Table 1, Proportion Answering Yes by Measure, with the counts for study group, tutor and writing center, followed by the start of Descriptive statistics.

What this means in plain English

The three services were not used equally. About 44% of students used a study group, 54% used a tutor, and 69% used the writing center, and gaps this large would be very unlikely by chance alone ($p < .001$). The test tells you that **at least one** service differs from the others; it does not say which pairs differ. Looking at the table, the writing center is clearly the most used, but a formal pairwise comparison would be needed to say, for example, whether the tutor and study group rates differ from each other.

What you would write in a paper

> Cochran's Q test showed a significant difference in the proportion answering Yes across the 3 measures, $Q(2) = 79.80$, $p < .001$, $N = 300$.

Add the three percentages in the sentence that follows so readers see the direction.

Your four results together

Click **Test Log** to see all four analyses listed together.

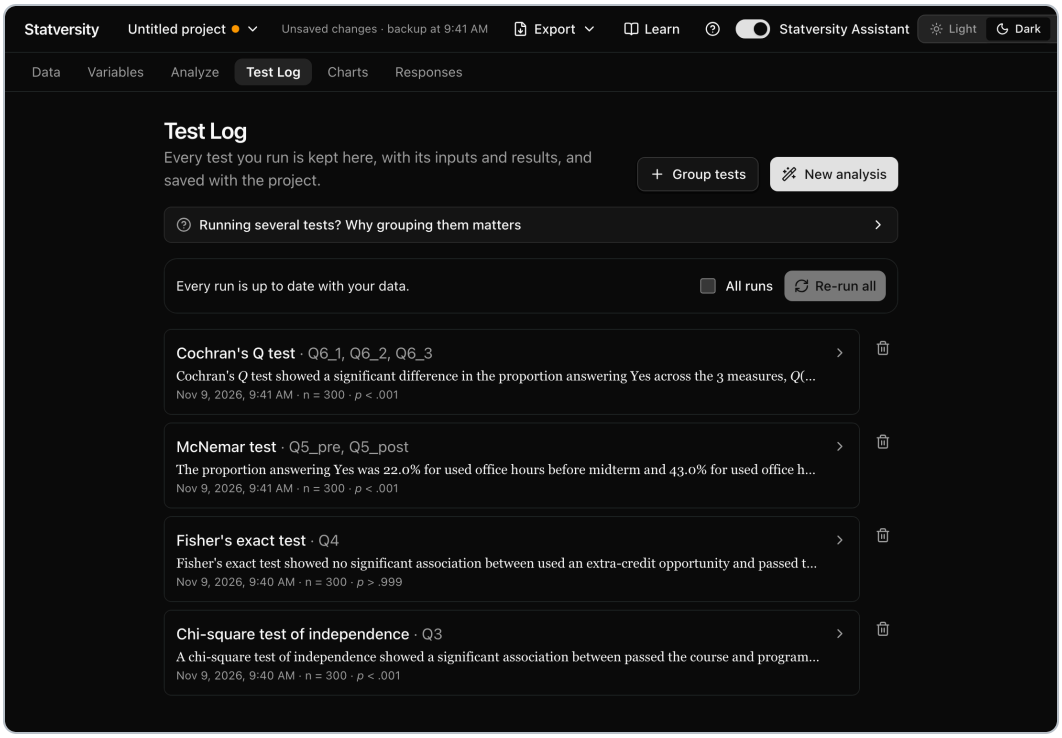


Figure 249. The Test Log listing Cochran's Q test, McNemar test, Fisher's exact test and Chi-square test of independence, each with its variables, its APA sentence, n = 300 and its p-value.

The Test Log keeps every test you have run, newest first, with the variables used, a shortened results sentence, the sample size and the p-value. Chapter 8 explains the Test Log in detail.

> **CAUTION**
> You have now run four tests on one dataset, and each reports its own p-value. When you run many tests, the chance that at least one is "significant" by luck grows. Here the four tests answer four separate, planned questions, so this is fine, but if you start running tests only to see what turns up, tell your readers how many you ran.

What if

- **The smallest expected count in Analysis 1 had been below 5.** Statversity would have shown a **Concern**, as in Analysis 2, and suggested Fisher's exact test (for a table larger than 2 by 2, one option is also to combine small categories before testing, as the check screen mentions).
- **The extra-credit group had been 100 students rather than 10.** The expected counts would have been comfortably above 5, the check would say **Looks fine**, and Statversity would have left chi-square as the suggested test.
- **The "before" and "after" answers had come from two different groups of students.** Then they would not be paired, and the right test would be chi-square, using **Do the answers differ**

between groups?. Choosing McNemar for unpaired data would give wrong answers, so check the design before picking.

- **You had only two yes/no questions from the same people, not three.** Choose **The same people's yes/no answer before and after.** and Statversity would use McNemar instead of Cochran's Q.
- **The effect size had been small.** Statversity would label Cramér's V as small instead of medium. With 300 students the test could still be significant, but you would describe the association as statistically significant but modest.

> **TRY IT YOURSELF**
>
> Run a fifth analysis on the same dataset. Ask whether **program (Q2)** is related to **using the extra-credit opportunity (Q4)**. Before you click **Decide which test to use**, predict whether the expected cell counts will look fine or raise a Concern. Then compare the smallest expected count with the "10 × 99 ÷ 300" arithmetic you saw in Analysis 2, and check what test Statversity suggests.

CHAPTER 17

Seeing how your survey branches

Many surveys do not show every question to everyone. A student who says they are enrolled in Course A sees questions about Course A; a student in Course B sees a different set. Survey tools call this **display logic**. It keeps a survey short for the person filling it in, but it scatters the answers across your data file, and that matters the moment you want to analyze them. This chapter shows how Statversity reads the branching from your Qualtrics survey file, draws it for you, and turns answers spread over several questions into one clean group you can compare.

The question

Which students have ever taken a course at another school, wherever in the survey they said so?

The survey asks it more than once. Each branch of the survey, one per course, has its own copy of "Which parts of the course have you used?", and each copy includes the answer **Course at another school**. So the information you want lives in six different columns, one per branch, and each student has an answer in at most one of them. The rest are blank, not because the student skipped them, but because the survey never showed them.

> IN PLAIN LANGUAGE
>
> **Why the blanks matter.** If you analyze one of those columns on its own, Statversity (or any program) sees mostly blank cells and treats them as missing answers. You would be comparing a handful of students against nobody. What you really want is one column that says, for every student, "yes, took a course elsewhere" or "no", built from all six places the question was asked. That column is a **group**, and it is what this chapter makes.

The data

The screenshots use a small made-up survey called **Synthetic course survey**, answered by 36 students. Nobody real is in it. It has:

- a first question, **Q1 Which course are you enrolled in?**, with the answers Course A, Course B, Course C, Course D, Course E and Not enrolled;
- six branch questions, **Q2** to **Q7**, all worded "Which parts of the course have you used?", one shown for each answer to Q1, each with the choices Lecture, Lab, Online module and Course at another school;

- a few follow-up questions, such as **Q8 Did that course include a lab component?**, which is shown to anyone who ticked Course at another school in any branch;
- some blocks of questions that everyone sees.

This survey is not one of the practice datasets, so you will not find it under **Open practice data**. Follow along with your own Qualtrics survey: import your data file together with the survey's .qsf file (chapter 5 shows where you add it on the **Choose files** screen). Without the survey file Statversity cannot see the branching, and none of the screens in this chapter appear.

A question that decides what comes next, like Q1, is called a **gate**. The questions it reveals are its **branches**.

Steps: seeing the branching in the Variable Interview

1. Import your data and your survey file. With the Assistant on, the Variable Interview opens after the import. Click **Continue** to reach **What is each question?**
2. Scroll to the gate question's row. Below its role choices, a box headed **Where this question leads** lists each answer to the gate, how many students gave it, and which later questions it shows.

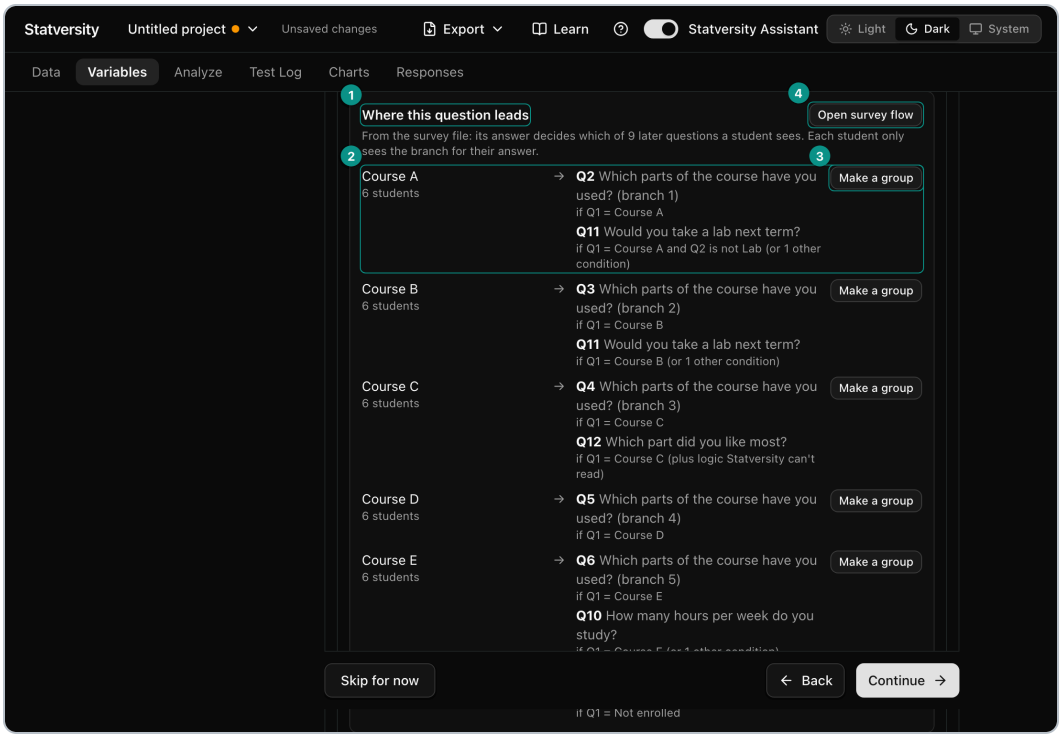


Figure 250. The Where this question leads panel on Q1. Each answer (Course A, Course B and so on) shows how many students chose it and the questions it reveals, with the condition under each one.

Number	Control	What it does
1	**Where this question leads**	Appears only on a question whose answers decide what is shown next. The line under it says how many later questions depend on the answer.
2	One answer's row	The answer, its count ("6 students"), an arrow, and each question it reveals, with the condition written out, for example "if Q1 = Course A". "(or 1 other condition)" means the question can also be reached another way.
3	**Make a group**	Starts a group from this answer (next section).
4	**Open survey flow**	Opens the whole survey as a diagram, with this question highlighted.

An answer that no student gave is greyed and reads **no one answered**. A question whose condition Statversity could only partly read says "(plus logic Statversity can't read)"; more on that below.

3. Scroll down to one of the branch questions. Under its wording, a line says when it is shown, for example **Shown only if Q1 = Course B**. The gate's name is a link: click it to jump back to the gate's row.

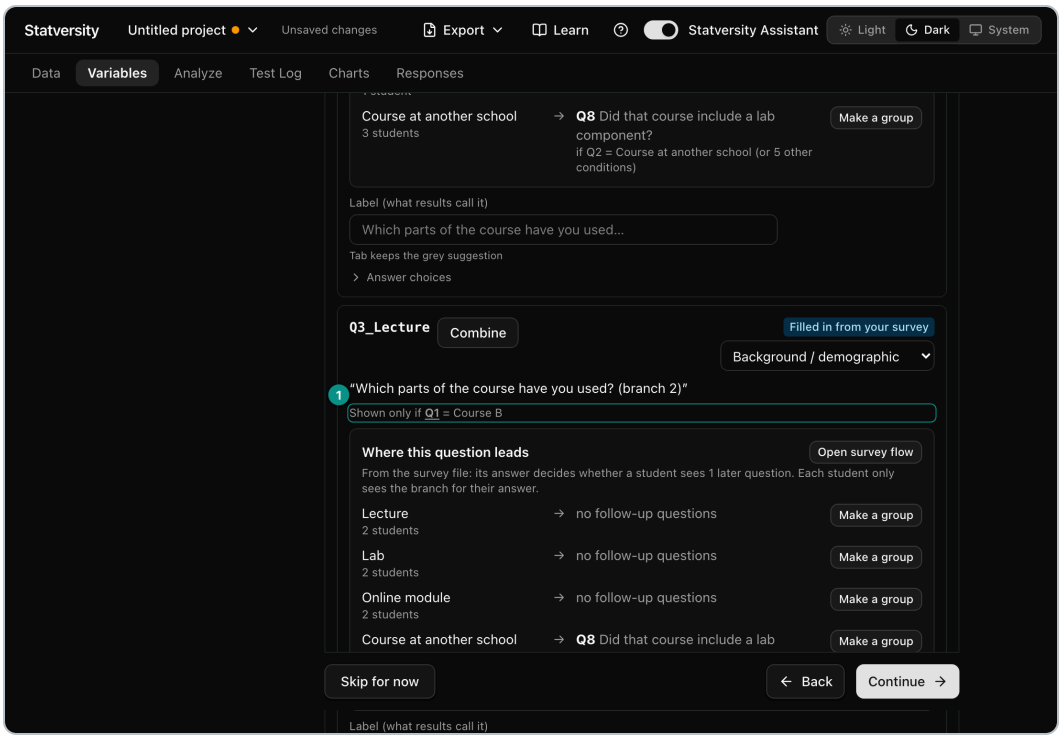


Figure 251. A branch question's row: "Shown only if Q1 = Course B", with Q1 as a link.

Notice that the branch question has its own **Where this question leads** box too. Ticking **Course at another school** in any branch is what shows Q8, so each branch question is a small gate of its own.

Steps: making a group from one answer

4. Go back to Q1's row and click **Make a group** beside **Course A**. A dialog opens: **New group: everyone who has taken Course A**.

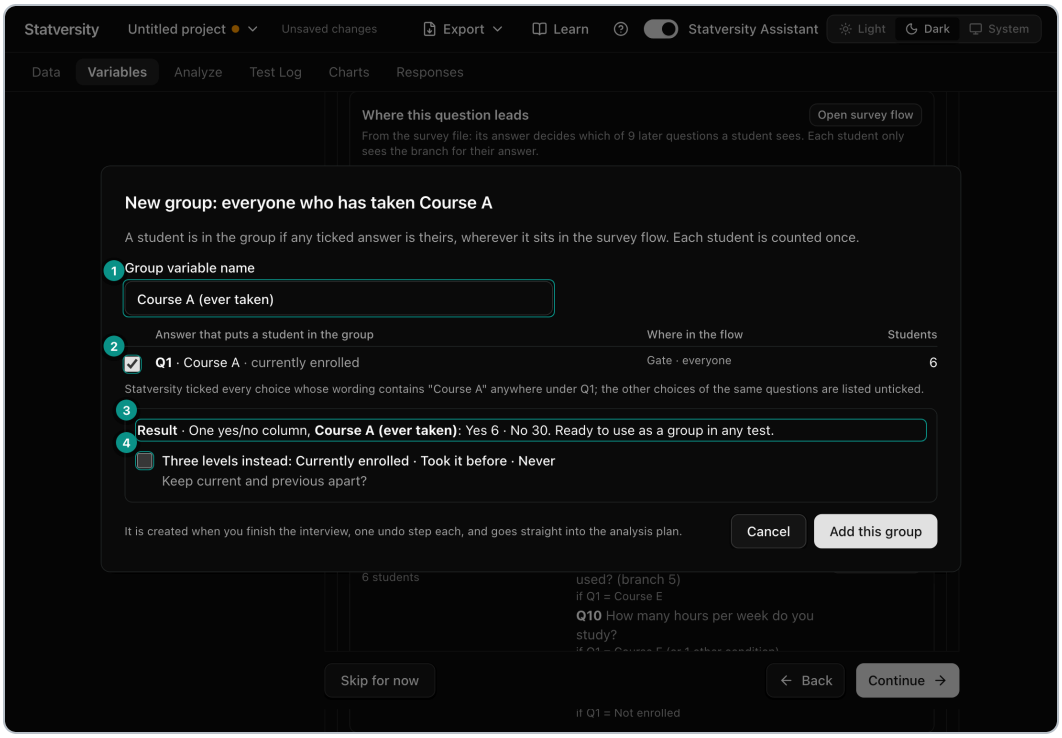


Figure 252. The New group dialog for Course A. Q1 Course A is ticked; the result line reads Yes 6, No 30.

Number	Control	What it does
1	**Group variable name**	The name of the new column. Statversity suggests "Course A (ever taken)"; type your own if you like.
2	The list of answers	Every answer that puts a student in the group. Ticked answers count; unticked ones do not. **Where in the flow** says where each answer sits, and the last column is how many students gave it.
3	**Result**	What you will get: one yes/no column and how many students land in each category, updated as you tick.
4	**Three levels instead**	Splits the group into Currently enrolled, Took it before and Never.

Statversity ticks every choice under the gate whose wording contains the answer you clicked, so in a survey that asks "Which of these courses have you taken before?" later on, an earlier **Course A** there is ticked for you as well. A choice that only looks similar is ticked but marked **wording differs, check it**.

Choices from the same questions that do not match are listed unticked, so you can tick any you want to add. In this small survey, Course A appears only in Q1, so only Q1 is ticked.

What “ever taken” means

The new column has three possible values for each student:

- **Yes**: at least one of the ticked answers is theirs, anywhere under the gate. A student counts once, however many ticked answers they gave.
- **No**: they answered the gate question, but none of the ticked answers.
- **Blank (missing)**: they never answered the gate question at all. Statversity does not count them as No, because it has no way to know. The result line says so, for example “blank 2 (did not answer Q1)”.

> **TIP**
> Tick **Three levels instead: Currently enrolled · Took it before · Never** when “enrolled now” and “took it in the past” mean different things for your question. **Currently enrolled** is the answer on the gate itself, **Took it before** is any other ticked answer, and **Never** is everyone else who answered the gate. A student who matches both is counted as currently enrolled.

When the answer you clicked also appears in other places under the gate, a box **Do the same for the other answers** lists the gate’s other answers. Tick the ones you want and the button becomes **Next**: Statversity makes one “(ever taken)” group per answer with the same rule, and shows you each list in turn so you can check the ticks (**Skip this one** leaves one out). In this small survey no course name repeats, so the box does not appear.

5. Click **Add this group**. The row now reads “Added Course A (ever taken). It is made when you finish the interview.” Nothing changes in your data yet. When you click **Finish**, Statversity creates the column, one undo step per group.

Steps: the Survey flow screen

6. Close or finish the interview and open the **Variables** tab. Click **Survey flow** in the toolbar. (In the interview, the **Open survey flow** buttons open the same screen.)

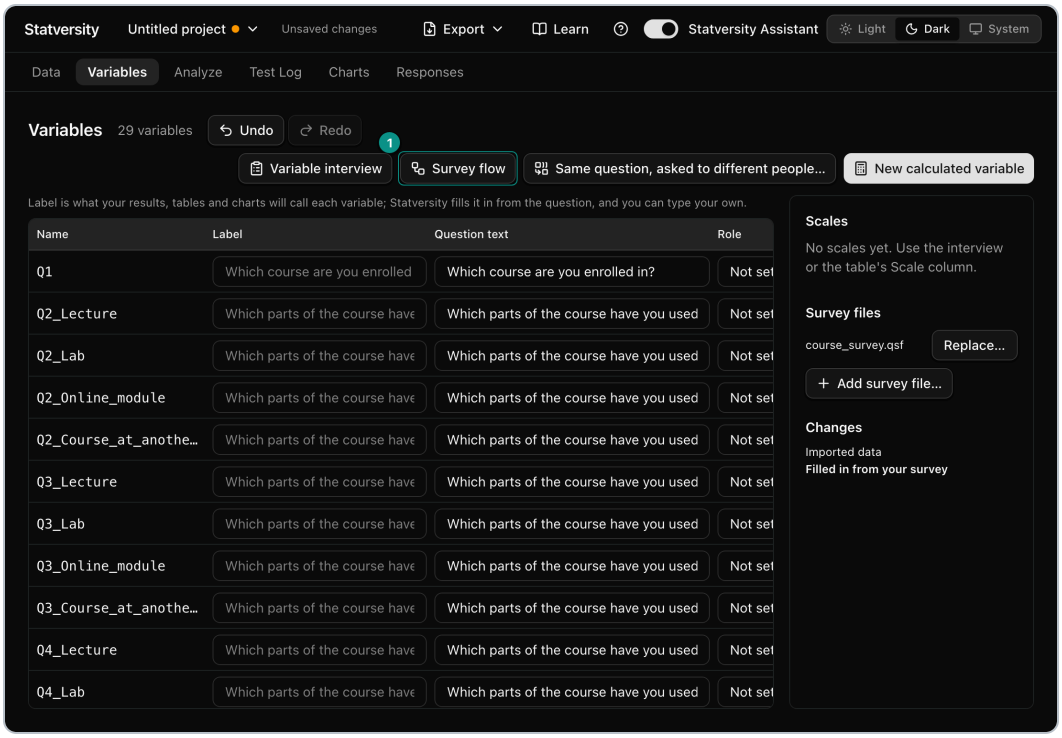


Figure 253. The Variables toolbar with the Survey flow button.

The screen fills the window. Its title names the survey, and the line under it counts the questions and how many are shown only to some students.

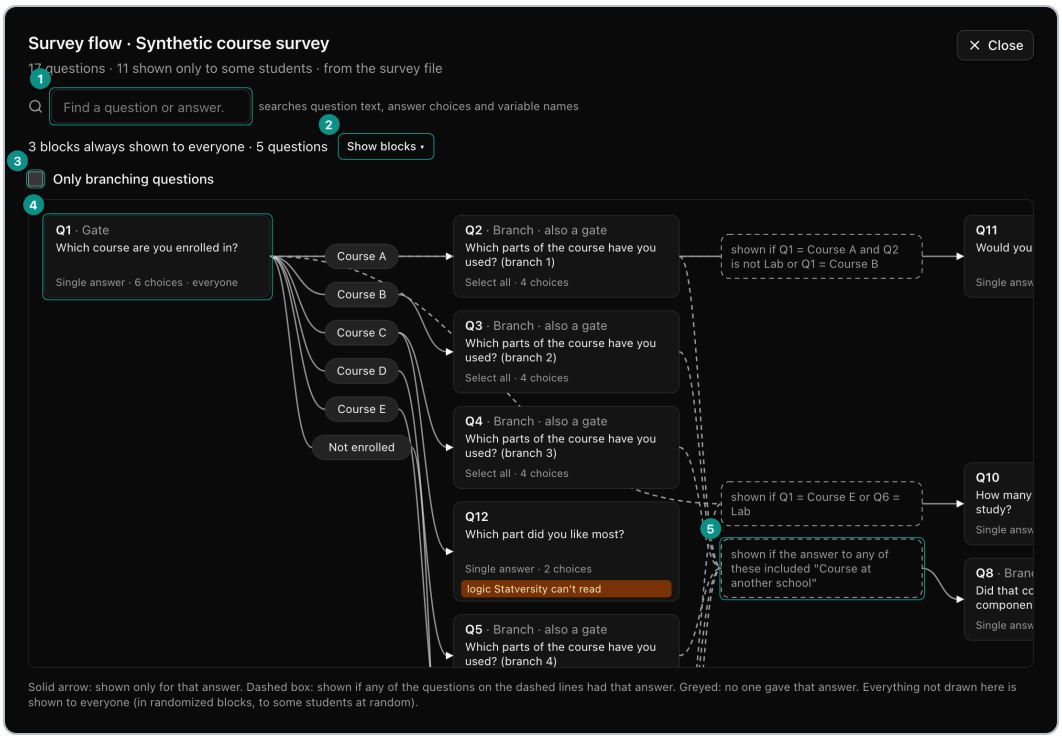


Figure 254. The Survey flow screen: Q1 on the left, a pill per answer, arrows to the branch questions, dashed boxes for questions reached from several places, and the badge on Q12.

Number	Control	What it does
1	**Find a question or answer…**	Searches question wording, answer choices and variable names (see the next section).
2	**Show blocks**	Lists the blocks of questions everyone sees. They are folded into one summary line, here “3 blocks always shown to everyone · 5 questions”, so the diagram stays about the branching.
3	**Only branching questions**	Hides everything except the questions with display logic.
4	A gate card	A question that decides what comes next, marked **Gate** (or **Branch · also a gate** when an earlier answer decides whether it is shown). Click any card to see its answers.
5	A dashed box	A question reached from several places (see below).

How to read the diagram

Read it from left to right, like the survey itself.

- **A gate card and its answer pills.** Q1 sits on the left. Each of its answers is a small rounded pill, and a **solid arrow** runs from the pill to the question that answer shows. Course A leads to Q2, Course B to Q3, and so on.
- **Dashed boxes.** Some questions can be reached from several places. Q8 is shown if the answer to any of Q2 to Q7 included Course at another school. Rather than draw six arrows, Statversity draws one dashed box that reads **shown if the answer to any of these included “Course at another school”**, with dashed lines from each question that feeds it, and one arrow from the box to Q8. A dashed box can also hold a mix, such as “shown if Q1 = Course E or Q6 = Lab”.
- **Greyed pills and boxes** read **no one answered**: the survey offers that path, but no student in your data took it.
- **The badge logic Statversity can’t read.** Statversity reads the common kinds of display logic: an answer was or was not chosen, combined with “and” and “or”. A condition built some other way, for example on answers carried forward from an earlier question, is badged rather than guessed at. Here Q12 has the badge. Check that question’s logic in Qualtrics before you build a group on it.
- **Chained gates** continue to the right: a branch question that is itself a gate has its own arrows.

The legend under the diagram repeats these rules. Everything not drawn is shown to everyone.

Click a card to select it. A panel below lists its answers with counts and a **Make a group** button for each; when you opened the screen from the interview, it also offers to take you to that question in the interview.

Steps: finding an answer across the survey

7. Type part of an answer into **Find a question or answer…**, here `another school`. The results appear above the diagram, grouped by question.

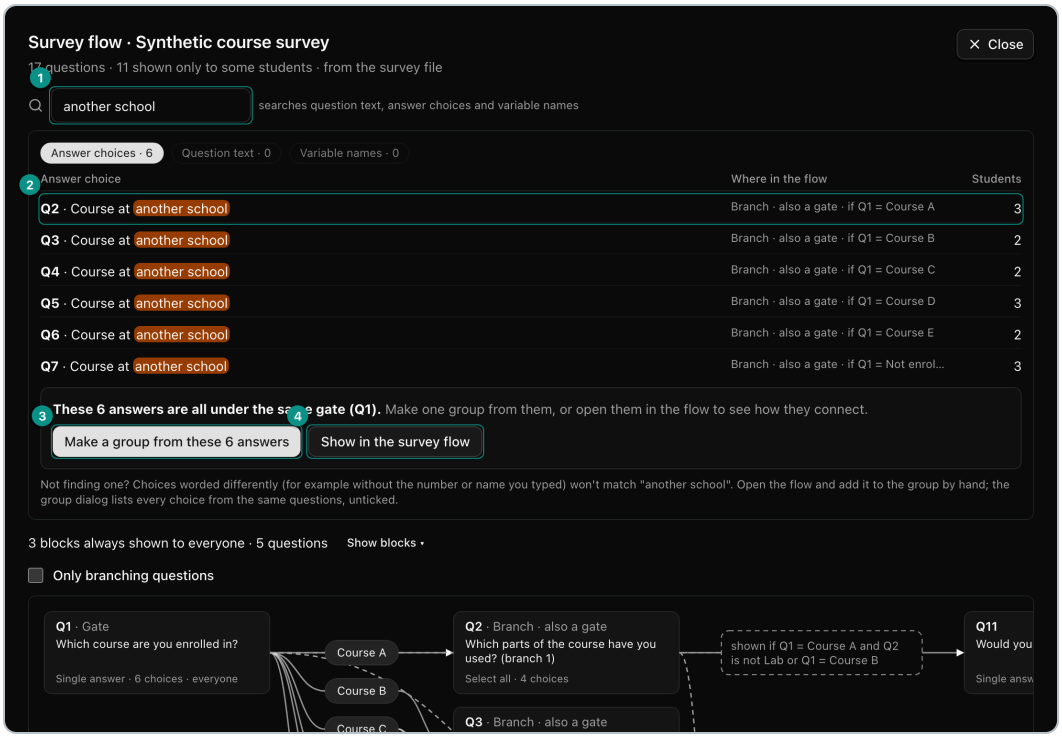


Figure 255. Search results for "another school": six matches, Q2 to Q7, each under a different branch of Q1, with Make a group from these 6 answers.

Number	Control	What it does
1	The search box	Matches anywhere in question wording, answer choices and variable names. The pills above the results (**Answer choices**, **Question text**, **Variable names**, each with a count) switch between kinds of match.
2	A result row	The question, the matching answer with your words highlighted, **Where in the flow** (“Branch · also a gate · if Q1 = Course A”, “always shown”), and how many students gave it.
3	**Make a group from these 6 answers**	Opens the group dialog with exactly these answers ticked.
4	**Show in the survey flow**	Highlights the questions in the diagram.

The line above the buttons says whether all the answers sit under one gate. Here it reads “These 6 answers are all under the same gate (Q1).” If they span more than one gate, Statversity asks **Which gate should the group be built on?**, because a group is built on one gate and only the answers under it are

ticked.

The same search box appears in the interview, on **What is each question?**, in a box called **Find an answer across the survey**.

8. Click **Make a group from these 6 answers**. The dialog title reads **New group: everyone who chose an answer with “another school”**. The six found answers are ticked, and the other choices of the same questions are listed unticked in case you want to add any.

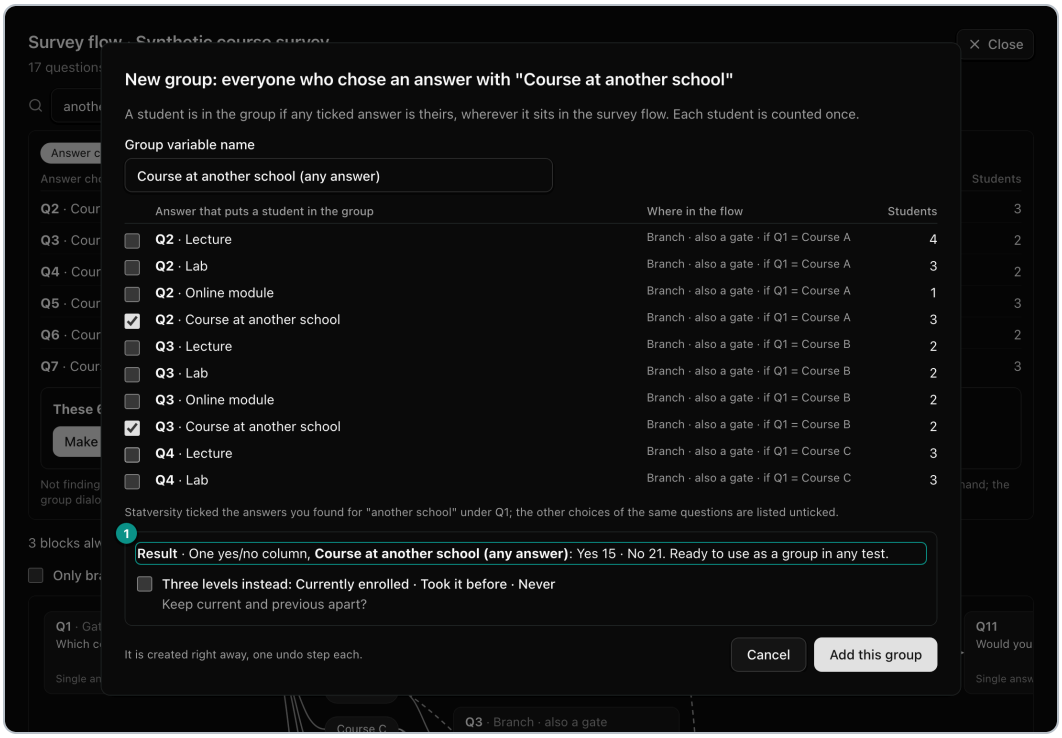


Figure 256. The group dialog from the search: the six Course at another school answers ticked, Lecture, Lab and Online module unticked. Result: Yes 15, No 21.

The result line reads **Yes 15 · No 21**: fifteen of the 36 students said, in whichever branch they saw, that they had taken a course at another school. That is the single column the question at the start of this chapter asked for. Rename it to something readers will understand, such as “Course elsewhere”, and click **Add this group**.

Where your new groups go

A group is an ordinary calculated column with the role **Group or cohort**:

- It appears on the **Variables** screen, marked **calculated**, like any other column you make (chapter 6).
- In **Plan my analyses** it is listed under **Groups to compare**, and in any test you can put it in the **Groups to compare** slot (chapter 7). Comparing students who took a course elsewhere with those

who did not is now one t test or chi-square away.

When the group is made depends on where you made it:

- **From the interview** (the **Where this question leads** box, or the search on **What is each question?**), groups are made when you click **Finish**, one undo step each. Until then you can still change your mind; **Cancel** in the dialog adds nothing.
- **From the Survey flow screen opened on Variables**, the group is made right away, again one undo step each. Click **Undo** on the Variables toolbar to remove it.

> **CAUTION**
> If a ticked question sits in a **randomized block**, the dialog warns that a blank there can mean the question was simply not shown to that student, not that the answer was "No". Think about whether No is fair for your question before you use the group.

> **NOTE**
> If your data stack several time points without linking students across them (chapter 5), the counts on these screens are responses, not students, and the screens say "responses". A student who answered twice is counted twice.

What if

What if Statversity shows no Where this question leads box? Either no survey file was added, or nothing in the survey uses display logic. On the Survey flow screen, a survey without branching says "No question in this survey is shown only to some students: everyone sees every question."

What if a course is worded differently in different places? Say the gate says "Course A" but a later question says "Course A (online)". Statversity ticks it and marks it **wording differs, check it**. Untick it if it is not the same course.

What if I want to group by something that is not under one gate? Use **Groups from answers** in **New calculated variable** (chapter 6). It lets you pick any answers from any questions, without the survey flow.

> **TRY IT YOURSELF**
> **Try it yourself.** On your own survey, search for a word that appears in several branches. Make a group from the results, then check the result line: do Yes, No and blank add up to the number of students you expect? Open **Plan my analyses** and find your new group under **Groups to compare**.

SECTION 4

Recipes

CHAPTER 18

What predicts the posttest score?

The question

“Which of the things I already know about my students (their program, how much they studied, how they did on the pretest) help explain how they did on the posttest?”

A novice asks this because a single comparison is rarely enough. Students who study more also tend to have scored higher on the pretest, and the two may be tangled together. Multiple regression untangles them by asking what each factor adds when the others are held steady.

> IN PLAIN LANGUAGE
>
> **Linear regression in one paragraph.** Think of drawing the best straight line through a cloud of dots, except that instead of one thing on the horizontal axis you have several. Statversity finds the combination of your predictors that comes closest to each student’s actual posttest score. Each predictor gets a number, called a coefficient (written b), that says how many points the posttest score changes when that predictor goes up by one unit and everything else stays the same. The model as a whole gets an R^2, the share of the differences in posttest scores that the predictors together account for.

The data

The practice set is 220 college students with a pretest, hours studied per week, GPA, a motivation survey and a posttest. It was built so that several things really do predict the outcome.

To open it, go to Home, click **Open practice data**, and find the card titled **What predicts the posttest score?**. The file is `survey.csv`.

Steps

1. On Home, click **Open practice data**, then click **Import this dataset** on the card **What predicts the posttest score?**. The Choose files screen opens with `survey.csv` already filled in.
2. Click **Read files**. The **About your study** screen appears.
3. Click **Skip these questions**. The **Check how we read them** screen appears.
4. Click **Continue**. The **Survey clean-up** screen appears.
5. Click **Continue**. The **Review and import** screen appears.

6. Click **Import**. The Variable Interview starts. Accept Statversity's defaults: click **Continue** on every interview screen, then **Finish**. You land on the Variables tab.
7. Click **Analyze**. The Test Advisor opens.
8. Under **Which score or answer do you want to analyse?**, choose **Posttest score (0-100) (Q7)**. The first question, **What do you want to know?**, appears.

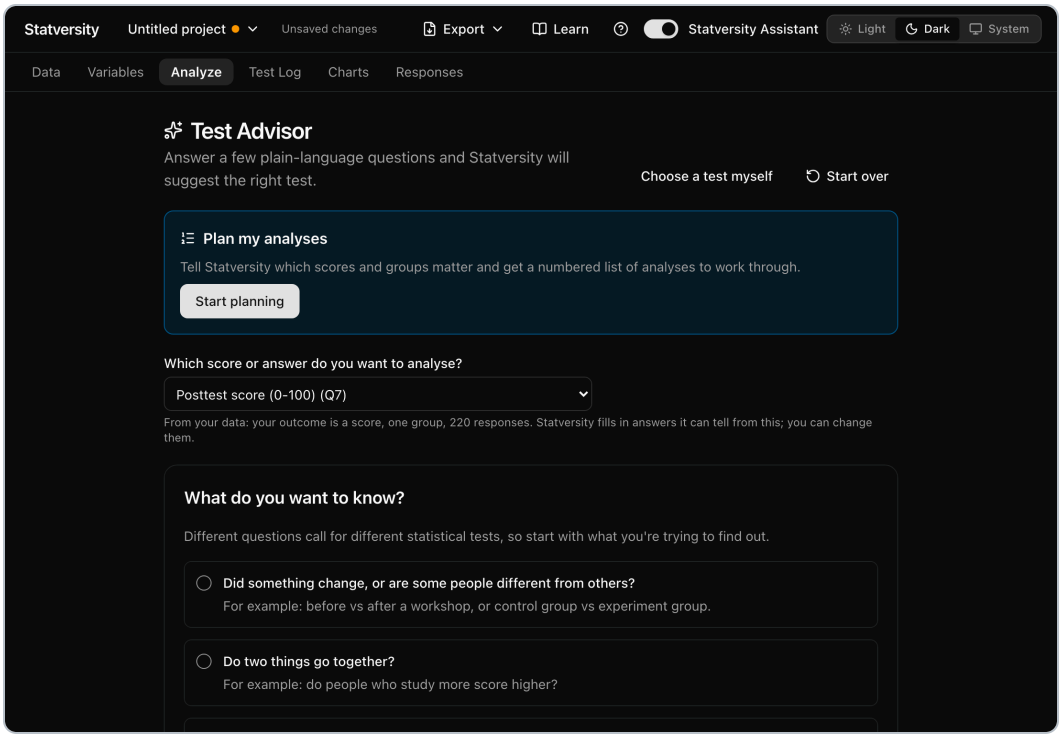


Figure 257. The Test Advisor with the posttest score chosen as the outcome. Statversity reports 220 responses and offers the first question.

9. Choose **Can some things I know help me estimate something else?** and click **Continue**.
10. For **How many things will you use to make the estimate?**, choose **Multiple (two or more)** and click **Continue**. Statversity now recommends **Linear regression**. (The screen also lists the effect sizes it will report, R^2 and Cohen's f^2, and the assumptions it will check.)

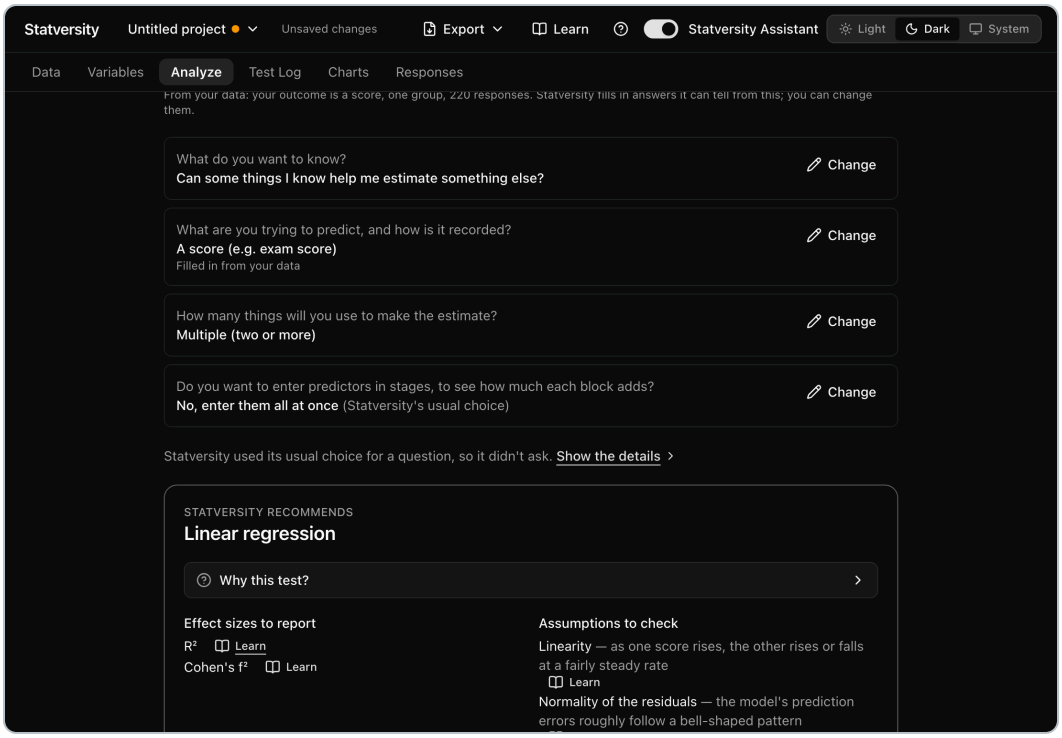


Figure 258. The recommendation card: Linear regression, with the question list above it.

11. Click **Set up this analysis**. **Choose your variables** opens. Under **Outcome (scores)**, Statversity has already filled in **Posttest score (0-100) (Q7)**.
12. Under the predictors list, tick **What is your program? (Q3)**, **About how many hours per week did you study? (Q4)** and **Pretest score (0-100) (Q6)**.

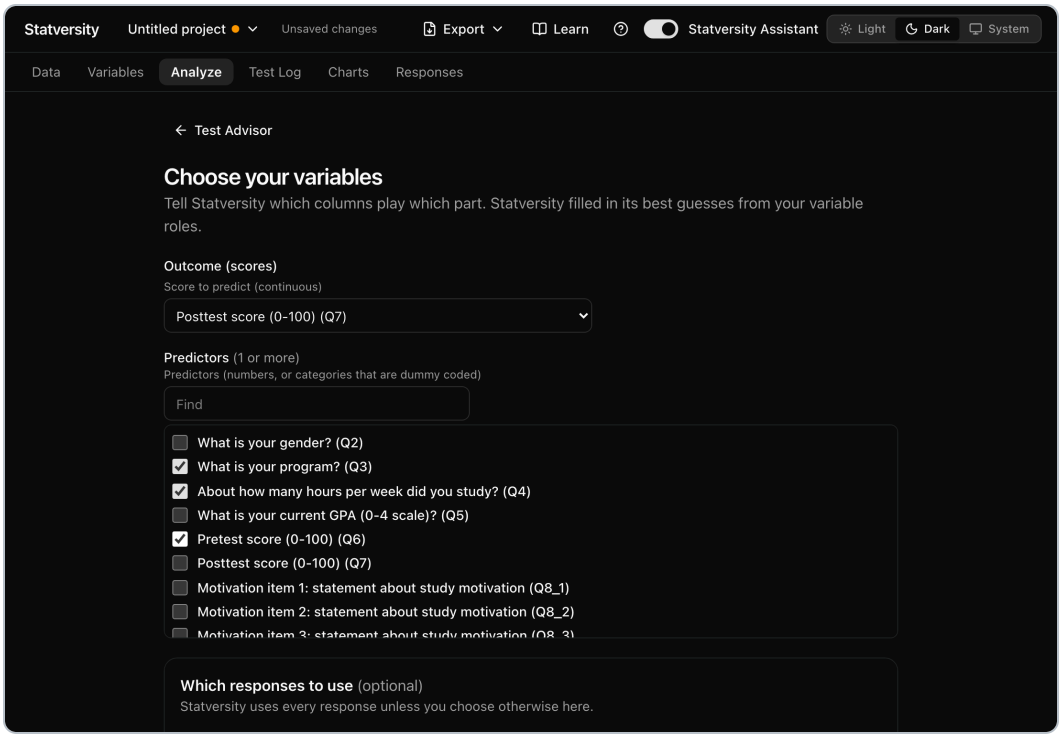


Figure 259. Choose your variables with the posttest as the outcome and three predictors ticked.

13. Click **Check the assumptions**. Statversity shows one screen per check. Click **Next check** to move through all five. The screens are described under “Watch out for” below.
14. On **Which test should you use?**, Statversity suggests **Linear regression** and pre-selects it. Click **Run and see the results**.

What Statversity found

The results page shows 220 responses analysed. The headline numbers are in the box at the top and in the table below it.

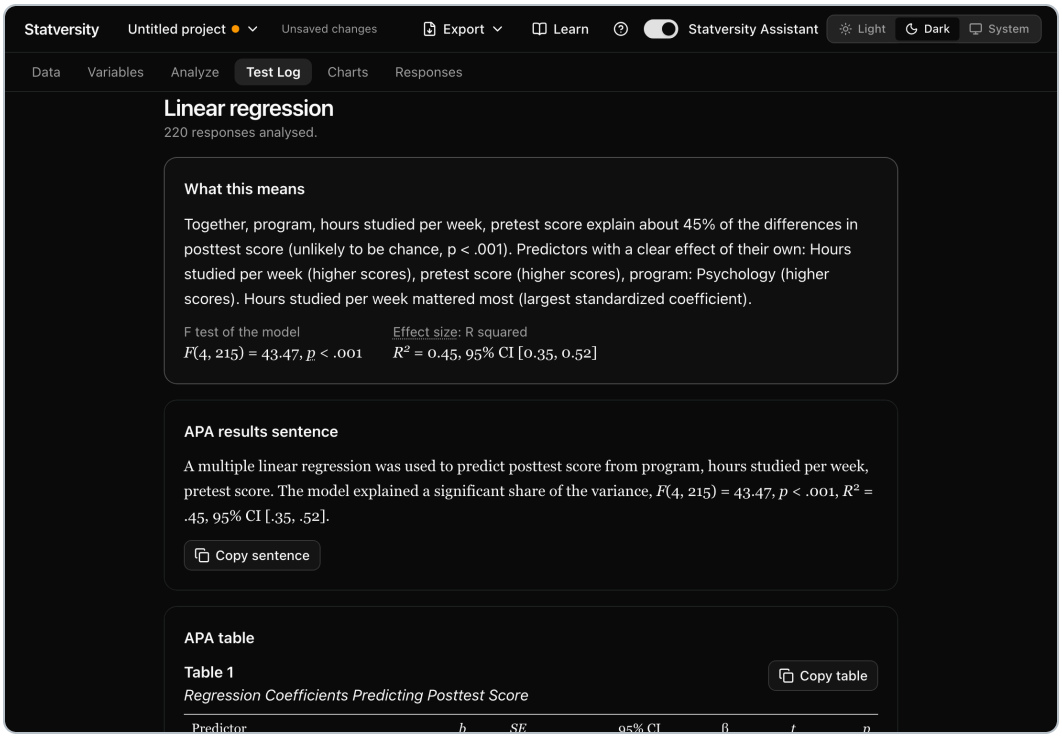
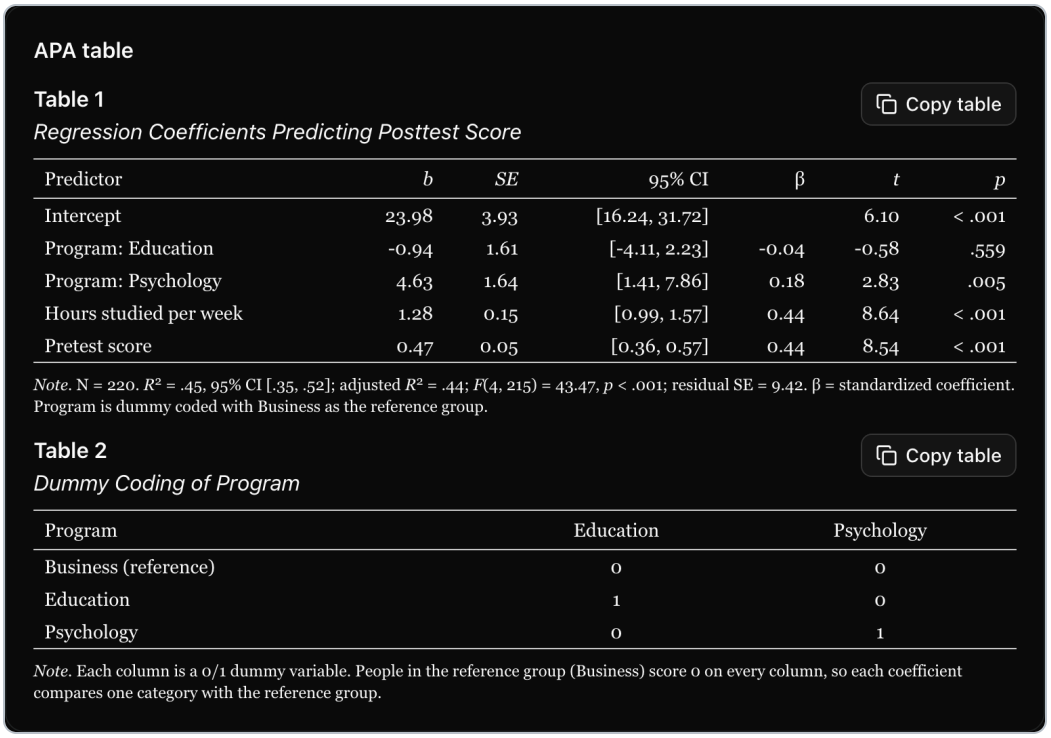


Figure 260. The top of the results: the What this means box, the APA sentence, and the start of the APA table.

Piece	b	95% CI	t	p
Program: Education	-0.94	[-4.11, 2.23]	-0.58	.559
Program: Psychology	4.63	[1.41, 7.86]	2.83	.005
Hours studied per week	1.28	[0.99, 1.57]	8.64	< .001
Pretest score	0.47	[0.36, 0.57]	8.54	< .001



APA table

Table 1

Regression Coefficients Predicting Posttest Score

Copy table

Predictor	*b*	*SE*	95% CI	β	*t*	*p*
Intercept	23.98	3.93	[16.24, 31.72]		6.10	< .001
Program: Education	-0.94	1.61	[-4.11, 2.23]	-0.04	-0.58	.559
Program: Psychology	4.63	1.64	[1.41, 7.86]	0.18	2.83	.005
Hours studied per week	1.28	0.15	[0.99, 1.57]	0.44	8.64	< .001
Pretest score	0.47	0.05	[0.36, 0.57]	0.44	8.54	< .001

Note. N = 220. R^2 = .45, 95% CI [.35, .52]; adjusted R^2 = .44; $F(4, 215)$ = 43.47, p < .001; residual SE = 9.42. β = standardized coefficient. Program is dummy coded with Business as the reference group.

Table 2

Dummy Coding of Program

Copy table

Program	Education	Psychology
Business (reference)	0	0
Education	1	0
Psychology	0	1

Note. Each column is a 0/1 dummy variable. People in the reference group (Business) score 0 on every column, so each coefficient compares one category with the reference group.

Figure 261. The full APA table: Table 1 with the coefficients and Table 2 showing how program was coded.

The whole model: $F(4, 215) = 43.47$, $p < .001$, $R^2 = .45$, 95% CI [.35, .52]. The table also shows an intercept of 23.98 (the predicted posttest score when every predictor is at zero, which is mostly a mathematical anchor rather than something to interpret). The standardized coefficients (β) for hours studied and pretest are both .44 on the screen. The table note adds adjusted $R^2 = .44$ and a residual SE of 9.42.

The **How big is the effect?** section labels $R^2 = .45$ (95% CI [0.35, 0.52]) as large, and Cohen's $f^2 = 0.81$ (95% CI [0.53, 1.10]) as large. The standardized coefficients are labeled medium for hours (β = .44, 95% CI [.34, .54]) and pretest (β = .44, 95% CI [.33, .54]), small for Psychology (β = .18, 95% CI [.05, .30]) and negligible for Education (β = -.04, 95% CI [-.16, .09]).



Figure 262. The effect size section: R squared, adjusted R squared, Cohen's f squared and each standardized coefficient with its label.

> **NOTE**
>
> **Why two program rows, not three?** A program has three levels, but a regression needs one level as the reference. Table 2 on the screen shows that Business is the reference group (it scores 0 on both columns), and each listed program is compared with Business.

What this means in plain English

Together, program, hours studied and pretest score explain about 45% of the differences in posttest scores, and that is very unlikely to be chance ($p < .001$). Following the wording in the reading-results chapter, the p-value tells you a pattern this strong would be very unlikely if none of the predictors were related to the posttest, and the R^2 tells you how big it is: a bit under half of the variation, which leaves a lot unexplained.

Looking at the individual pieces, holding the others steady:

- Each extra hour studied per week went with about 1.28 more points on the posttest.
- Each extra point on the pretest went with about 0.47 more points on the posttest.
- Psychology students scored about 4.63 points higher than Business students.
- The Education row (also compared with Business) has a p of .559 and a confidence interval that runs from below zero to above zero, so the data do not show a reliable difference there. That is

> not the same as proving there is none.

Statversity's summary box says hours studied "mattered most" because it has the largest standardized coefficient. On screen hours and pretest both round to .44, so treat them as about equally important.

What you would write in a paper

Statversity's APA sentence is ready to copy:

> A multiple linear regression was used to predict posttest score from program, hours studied per week, pretest score. The model explained a significant share of the variance, $F(4, 215) = 43.47$, $p < .001$, $R^2 = .45$, 95% CI [.35, .52].

Add a sentence or two about the individual predictors from the APA table (use **Copy table** to paste the whole table into your document).

Watch out for

Statversity ran five checks on the residuals (the leftover errors of the model). Two raised concerns.

> **CAUTION**
>
> **Normality of residuals: caution.** The Shapiro-Wilk test gave $W = 0.91$, $p < .001$. With 220 people this test flags tiny differences, and regression holds up well, but Statversity asks you to look at the Q-Q plot and histogram. On the results page, under Assumption checks, Statversity repeats the check with its plots. Linear regression has no results chart of its own in Statversity, so the assumption plots are the only pictures.

> **CAUTION**
>
> **Linearity: concern.** The Ramsey RESET test gave $F(2, 213) = 6.54$, $p = .002$, which means the relationship looks slightly curved rather than straight. Statversity suggests a transformed or squared predictor might fit better.

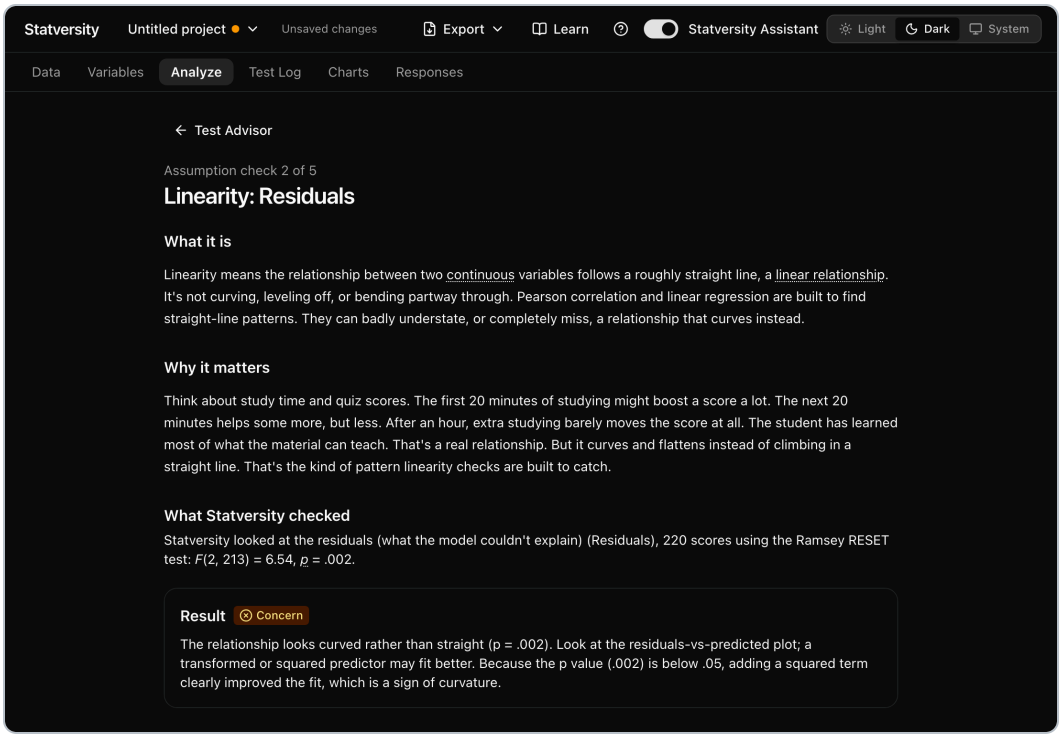


Figure 263. The Linearity check flagged as a concern on its own screen.

Equal spread of residuals passed ($p = .287$) and so did multicollinearity (predictors that are too closely related to each other). The influential cases check (a few unusual people pulling the line) was also flagged. The practice data intentionally contains two outliers, one very low posttest and one very high hours-studied value.

Because there is no rank-based alternative to regression, Statversity's advice was: "run it and mention the concern when you report it." A sentence in your methods such as "residuals departed slightly from normality and linearity" is honest and enough for a first look.

> **TIP**
> The screen labels an assumption problem "Concern" or "Caution". A "Concern" does not stop the test from running. It tells you what to mention.

CHAPTER 19

Can we trust the graders?

The question

“If several people grade the same essays, do they agree with each other closely enough that I can trust the grades?”

A novice asks this the moment a second grader appears. If two raters give the same essay a 2 and a 5, the grade says more about the rater than the essay. The statistic for this question is the **intraclass correlation**, or **ICC**.

> IN PLAIN LANGUAGE
>
> **What ICC measures.** Imagine every essay has a true quality, and each rater’s grade is that quality plus some personal noise (this rater is stricter, that one is sloppy). ICC is the share of all the variation in grades that comes from real differences between essays rather than from rater noise. It runs from 0 (grades are all noise) up to 1 (raters agree perfectly). Higher is better.

The data

The practice set has 60 student essays, each graded 1 to 5 by three raters. A few cells are blank on purpose: a rater skipped an essay. The card also offers a second file for a different exercise (two raters sorting essays into four categories, for Cohen’s kappa). This recipe uses only the essay grades.

To open it, go to Home, click **Open practice data**, and find the card titled **Can we trust the graders?**. It asks **Which file do you want to import?** and lists two choices. Pick **Essay grades from 3 raters (for ICC)**, which is `essay_ratings.csv`.

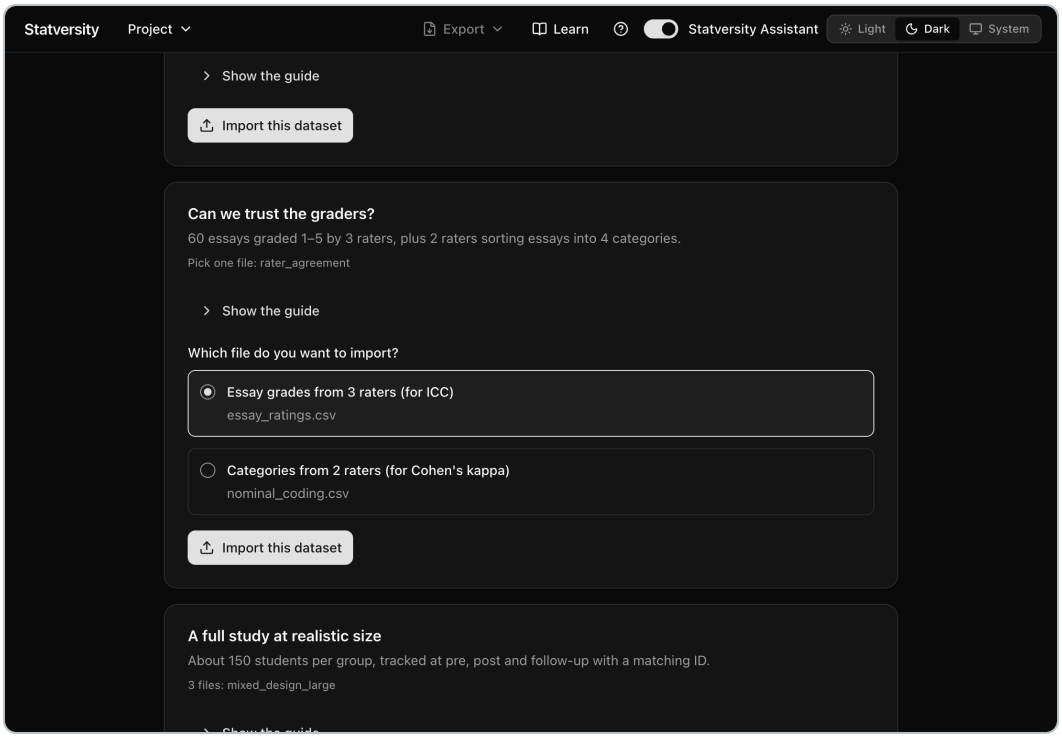


Figure 264. The practice card for Can we trust the graders?, with the essay grades file selected.

Steps

1. On Home, click **Open practice data**. On the card **Can we trust the graders?**, choose **Essay grades from 3 raters (for ICC)**.
2. Click **Import this dataset**. The Choose files screen opens.
3. Click **Read files**, then **Skip these questions** on **About your study**.
4. On **Check how we read them**, click **Continue**. On **Survey clean-up**, click **Continue**.
5. On **Review and import**, click **Import**.
6. Accept the defaults in the Variable Interview: **Continue** on every screen, then **Finish**. You land on Variables.
7. Click **Analyze**. The Test Advisor opens.
8. Under **Which score or answer do you want to analyse?**, choose **rater1**. The first question appears.
9. For **What do you want to know?**, choose **Do my rating questions consistently measure one idea?** and click **Continue**.
10. For **What kind of questions are they?**, choose **Do different raters or coders agree?** and click **Continue**.
11. For **What do the ratings look like?**, choose **Scores that can be averaged, like two or more raters scoring the same items on a 1-10 scale** and click **Continue**.

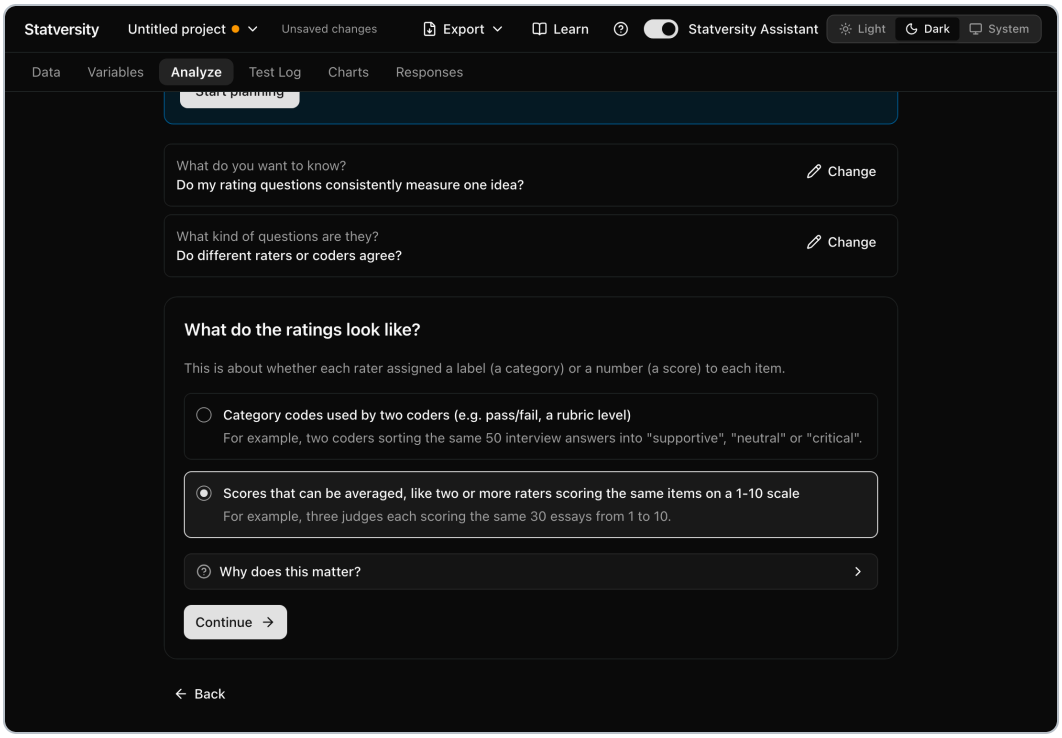


Figure 265. The third Advisor question, with the averaged-scores answer selected.

12. Statversity recommends **Intraclass correlation (ICC)**. Click **Set up this analysis**.

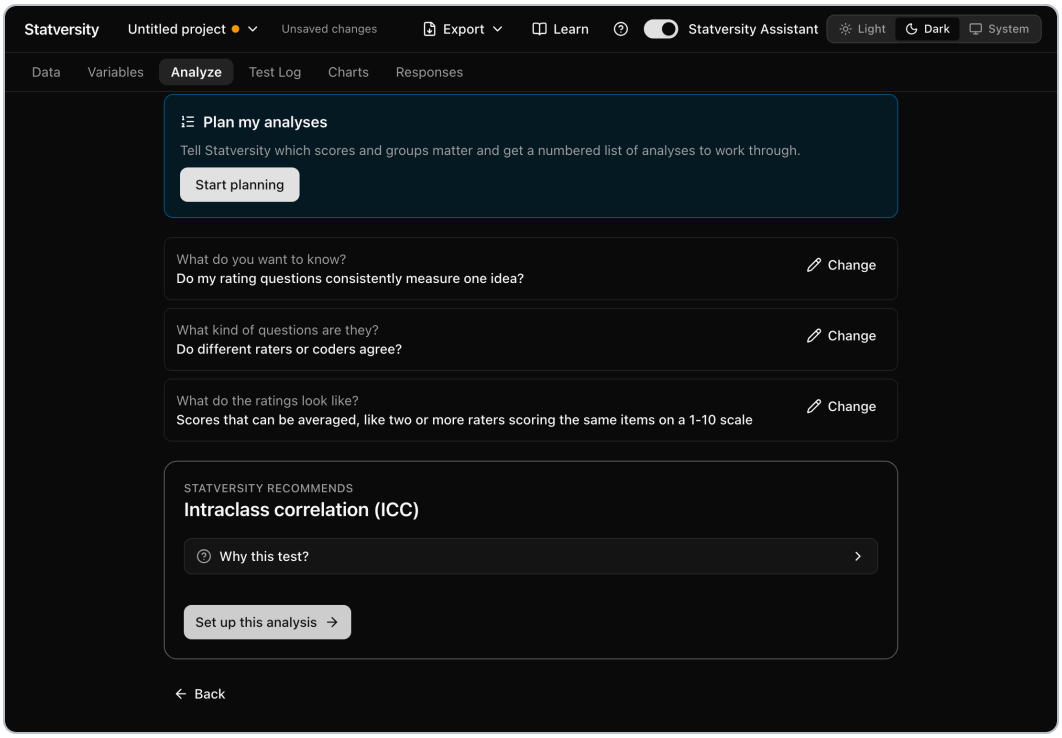


Figure 266. The recommendation card for Intraclass correlation (ICC) after the three answers.

13. On **Choose your variables**, under **Raters**, tick **rater1**, **rater2** and **rater3**. Leave **essay_id** unticked; it only names the essay.

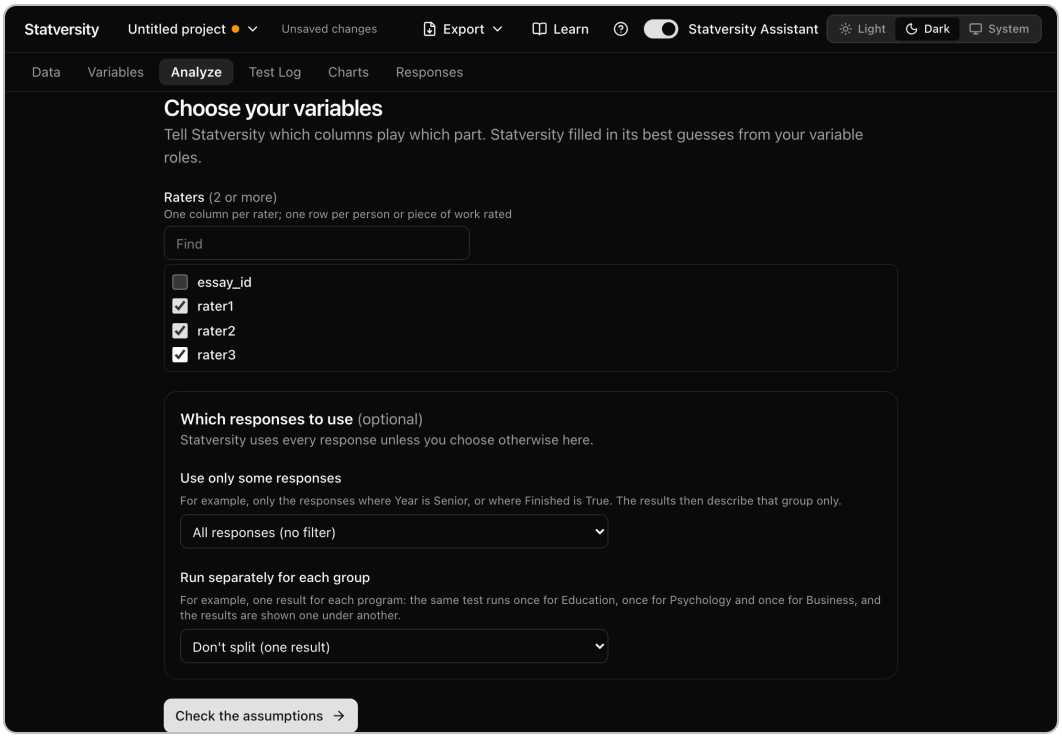


Figure 267. Choose your variables with the three rater columns ticked.

14. Click **Check the assumptions**. The results page opens; the capture shows no separate assumption screens for this test.

> **NOTE**
>
> **The wording does not match perfectly.** The first Advisor answer says “rating questions consistently measure one idea”, which sounds like a survey scale. Do not let it put you off: the next question, “Do different raters or coders agree?”, is where you tell Statversity this is about raters. The last answer also mentions a 1-10 scale as its example, while this dataset’s grades run from 1 to 5. Both are fine for ICC.

What Statversity found

The results are headed **Intraclass correlation (ICC)** with the line “55 responses analysed, 5 left out for missing answers.”

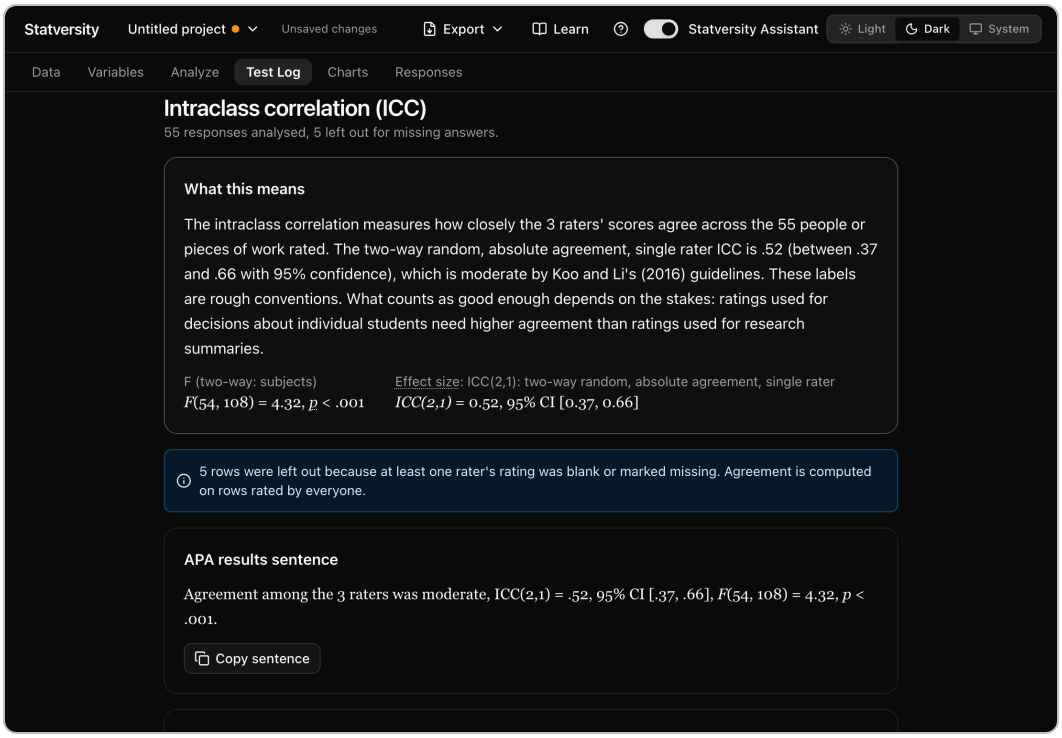


Figure 268. The top of the results: the What this means box, the notice about rows left out, and the start of the APA sentence and table.

Piece	Value
ICC(2,1): two-way random, absolute agreement, single rater	.52
95% confidence interval	[.37, .66]
F(54, 108)	4.32
p	< .001
Label Statversity gives it	moderate

What this means in plain English

The three raters' scores agree moderately well across the 55 essays that everyone graded. The ICC of .52 means that roughly half of the differences in the grades reflect real differences between essays, and the rest is rater disagreement or noise. The confidence interval says the true value is plausibly anywhere from .37 to .66, which is wide. That is the honest range for 55 essays.

The tiny *p*-value (< .001) only tells you the agreement is better than none at all. It does not say the agreement is good enough. For that, use the label. Statversity calls .52 moderate by Koo and Li's (2016) guidelines, and adds that these labels are rough conventions. How much agreement you need depends

on the stakes: grades that decide something about an individual student need higher agreement than grades used for a research summary.

What you would write in a paper

> Agreement among the 3 raters was moderate, ICC(2,1) = .52, 95% CI [.37, .66], $F(54, 108) = 4.32$, $p < .001$.

You should also say how many essays were graded by all three raters (55 here) and why the others were dropped; the next section covers that.

Watch out for

> **CAUTION**
>
> **Five rows were left out.** Above the APA sentence Statversity shows a notice: “5 rows were left out because at least one rater’s rating was blank or marked missing. Agreement is computed on rows rated by everyone.” The practice file has 60 essays, so 55 were used. Statversity never guesses a missing grade. If many essays are missing a rating, your result could differ from one computed on all of them, so report the number left out.

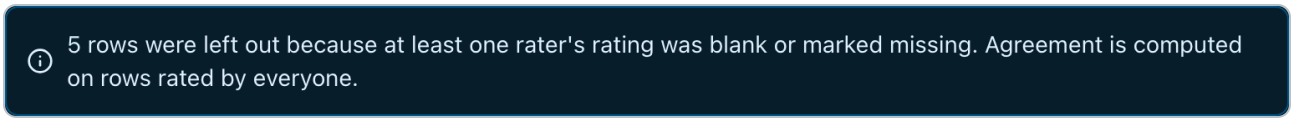


Figure 269. The results again, showing the notice about the five rows that were left out.

Two other things to know:

- The APA table lists six forms of ICC (see below). The sentence uses ICC(2,1), described on screen as two-way random, absolute agreement, single rater. Quote the form the sentence names.
- The ICC results have no chart, so the tables and numbers are the whole story.

APA table

Table 1

Intraclass Correlation Coefficients

Copy table

Form	ICC	95% CI	F	df_1	df_2	p
ICC(1,1): one-way random, single rater	.52	[.37, .66]	4.26	54	110	< .001
ICC(2,1): two-way random, absolute agreement, single rater	.52	[.37, .66]	4.32	54	108	< .001
ICC(3,1): two-way mixed, consistency, single rater	.52	[.37, .67]	4.32	54	108	< .001
ICC(1,k): one-way random, average of raters	.77	[.63, .86]	4.26	54	110	< .001
ICC(2,k): two-way random, absolute agreement, average of raters	.77	[.63, .86]	4.32	54	108	< .001
ICC(3,k): two-way mixed, consistency, average of raters	.77	[.64, .86]	4.32	54	108	< .001

Note. n = 55 rated with all 3 raters. Shrout and Fleiss (1979) forms; headline ICC(2,1). Rater effect: $F(2, 108) = 1.75$, $p = .179$.

Figure 270. The APA table listing all six ICC forms, with the note that 55 essays were rated by all three raters.

The three single-rater forms all give .52 ([.37, .66], or [.37, .67] for the two-way mixed form). The three average-of-raters forms give .77 ([.63, .86], or [.64, .86] for two-way mixed). The “single rater” rows tell you how reliable one grader is; the “average of raters” rows tell you how reliable the mean of all three graders is. Use the single-rater value if future essays will be graded by one person. The table note also reports the rater effect, $F(2, 108) = 1.75$, $p = .179$, which means the raters’ average strictness did not differ reliably.

CHAPTER 20

A full study at realistic size

The question

“We measured three groups of students (a control group and two interventions) before, right after, and some weeks later. Did the groups change differently over time?”

A novice asks this because it is the shape of most real studies: more than one group, more than one time point, and a few people who dropped out along the way. The test that fits is a **mixed ANOVA**.

> IN PLAIN LANGUAGE
>
> **Mixed ANOVA in plain words.** “ANOVA” compares averages. “Mixed” means the design mixes two kinds of comparison. *Between* people: Control versus Intervention A versus Intervention B (different people in each group). *Within* people: Pre versus Post versus Follow-up (the same people measured three times). The test gives you three answers: do the groups differ overall, does the score change over time overall, and, the interesting one, does the change over time depend on the group? That last one is called the **interaction**.

The data

The practice set is a large study: about 150 students per group, in three files (`pre.csv`, `post.csv` and `followup.csv`), each carrying a matching ID so one person can be followed across the three times. It has the groups (Control, Intervention A, Intervention B), a score column, and some students who drop out between time points, as in real life.

To open it, go to Home, click **Open practice data**, and find the card **A full study at realistic size**. Its file list shows pre.csv, post.csv and followup.csv.

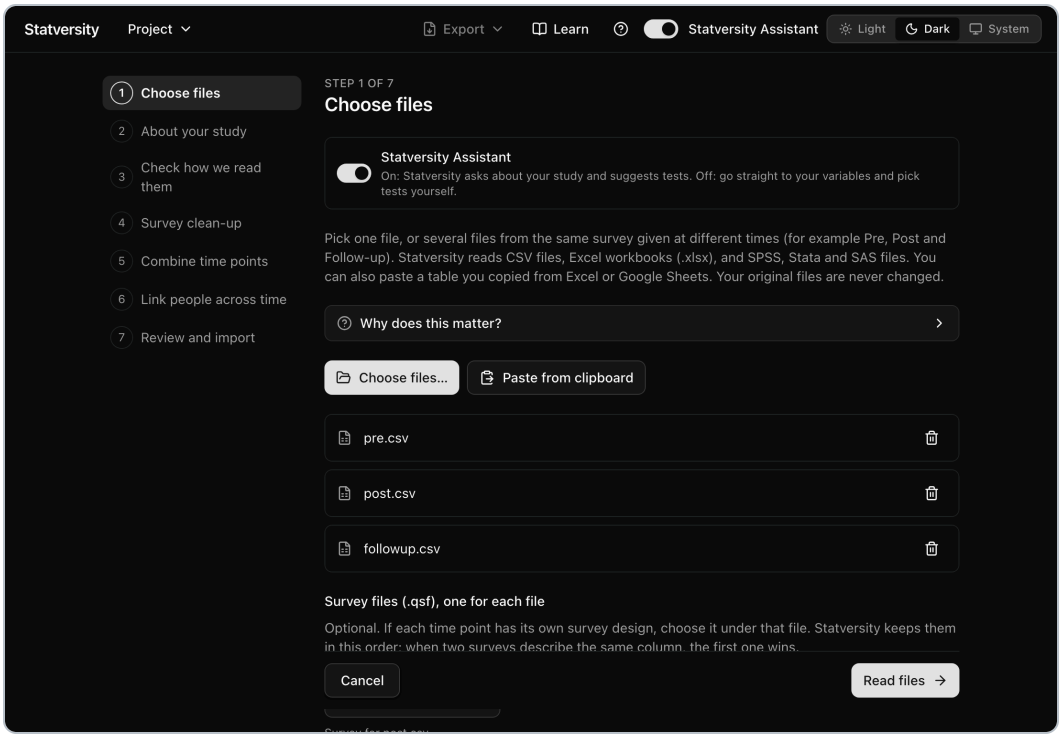


Figure 271. Choose files with the three files filled in from the practice card.

Steps

1. On Home, click **Open practice data**, then **Import this dataset** on the card **A full study at realistic size**.
2. On Choose files, click **Read files**. On **About your study**, click **Skip these questions**.
3. On **Check how we read them**, click **Continue**. On **Survey clean-up**, click **Continue**.
4. On **Combine time points**, click **Continue**.
5. On **Link people across time**, choose **Link people by an ID**. A box asks **Which column identifies each person?**
6. In **ID column**, choose **Q1 - Please enter your unique ID (first two initials + birth day,**. (The label is cut off on screen because the question text is long.) Leave the two ticked boxes, **Ignore extra spaces** and **Ignore capital letters**, as they are.

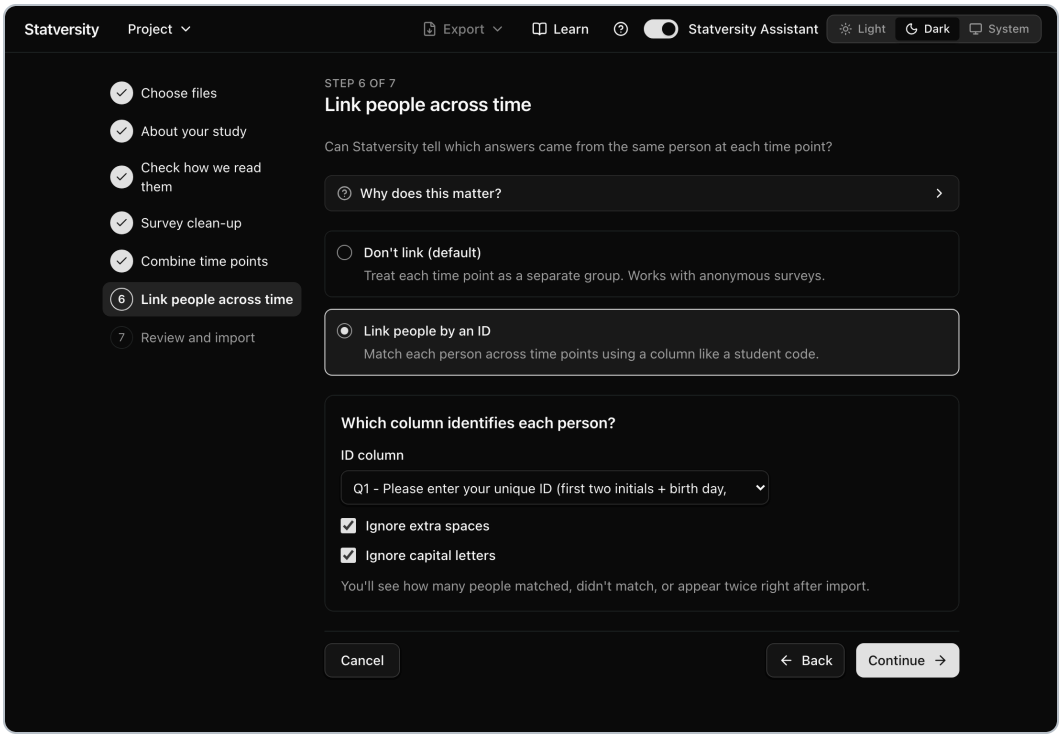


Figure 272. Link people across time with Link people by an ID chosen and the ID column picked.

7. Click **Continue**, then on **Review and import**, click **Import**.
8. **Import complete** appears with a box called **How the IDs matched**. Read it before you move on (see below). Click **Set up variables**.

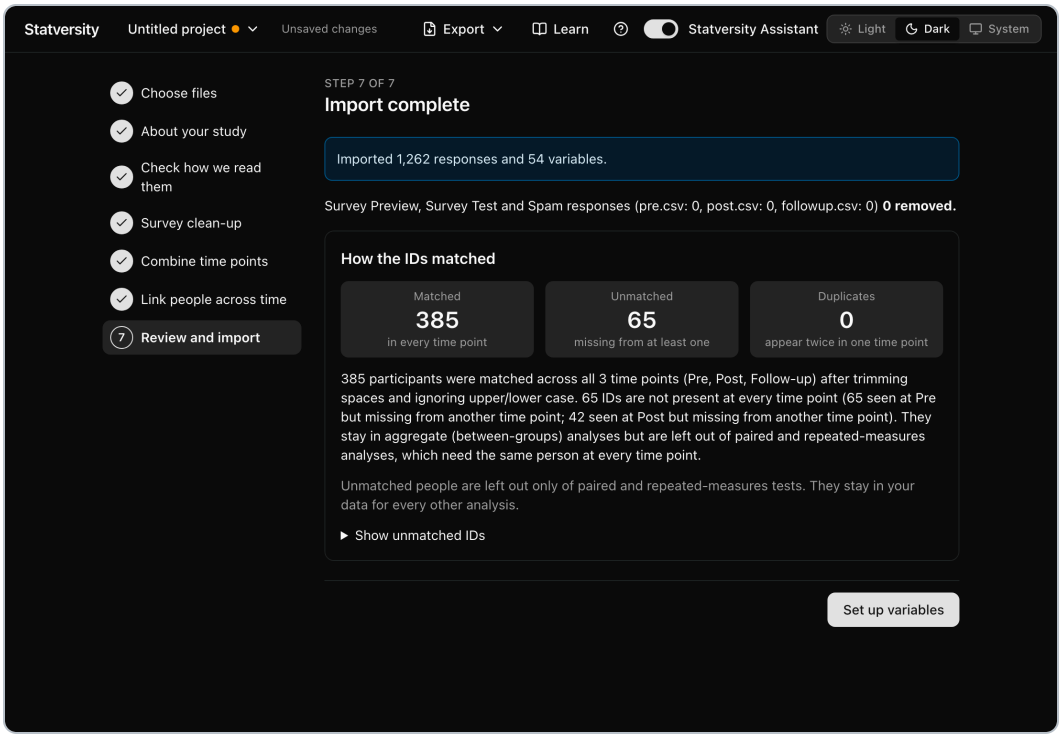


Figure 273. Import complete, with the ID matching summary: 385 matched, 65 unmatched, 0 duplicates.

9. Accept the defaults in the Variable Interview: **Continue** on every screen, then **Finish**. You land on Variables.
10. Click **Analyze**. Under **Which score or answer do you want to analyse?**, choose **Qualtrics score (SC0)**.
11. For **What do you want to know?**, choose **Did something change, or are some people different from others?** and click **Continue**.
12. For **What are you comparing?**, choose **The same people at different times, and I can match their answers.** and click **Continue**.
13. For **Were there also different groups, each measured at those same times?**, choose **Yes, groups measured across the same time points** and click **Continue**. Statversity recommends **Mixed ANOVA (between x within)**.
14. Click **Set up this analysis**. **Choose your variables** opens already filled in.

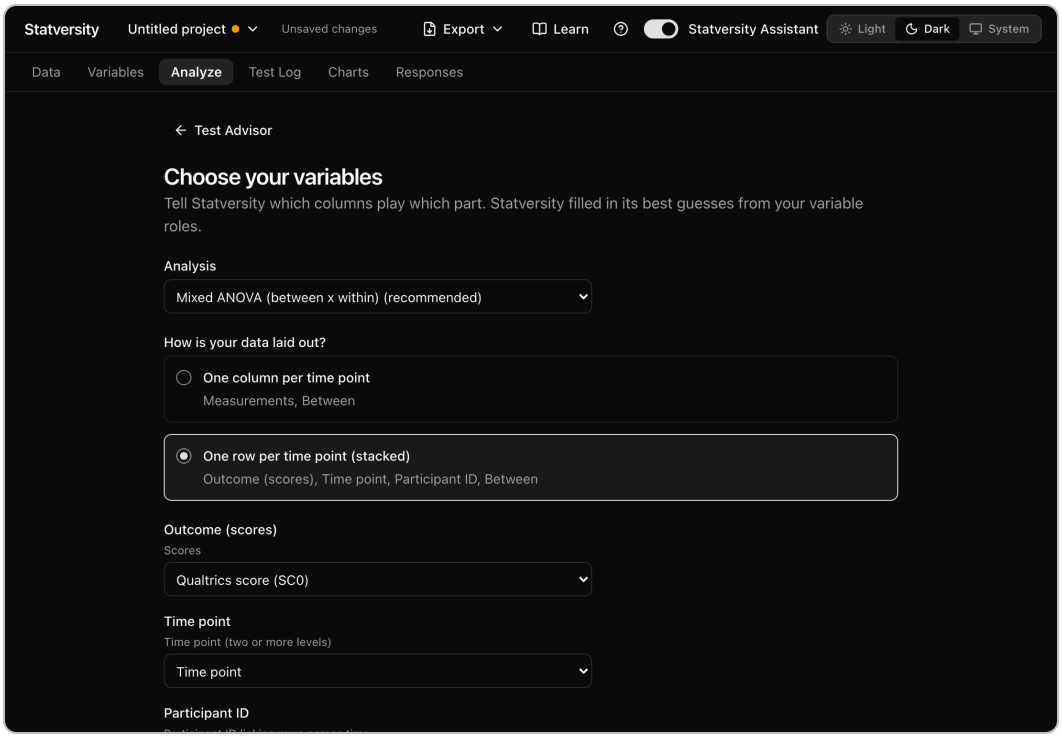


Figure 274. Choose your variables for the mixed ANOVA: one row per time point, Qualtrics score as the outcome, Group (Q2) as the between factor.

On the screen, the data layout is **One row per time point (stacked)**, the outcome is **Qualtrics score (SC0)**, the time variable is **Time point**, the **Participant ID** is the Q1 column, and **Between** is **Group (Q2)**.

15. Click **Check the assumptions**. There are eight screens. Click **Next check** through all of them. (Details below.)
16. On **Which test should you use?**, Statversity suggests **Mixed ANOVA (between x within)** and says the cautions are mild and common with larger samples. Keep it and click **Run and see the results**.

> **NOTE**
>
> **Do not be thrown by the name "Qualtrics score (SC0)".** It is simply the label this practice file's score column carries in Statversity. It is the outcome you are analysing.

What Statversity found

The results are headed **Mixed ANOVA (between x within)**. The line under the heading reads "385 responses analysed, 107 left out for missing answers (Control: 125; Intervention A: 128; Intervention B: 132)."

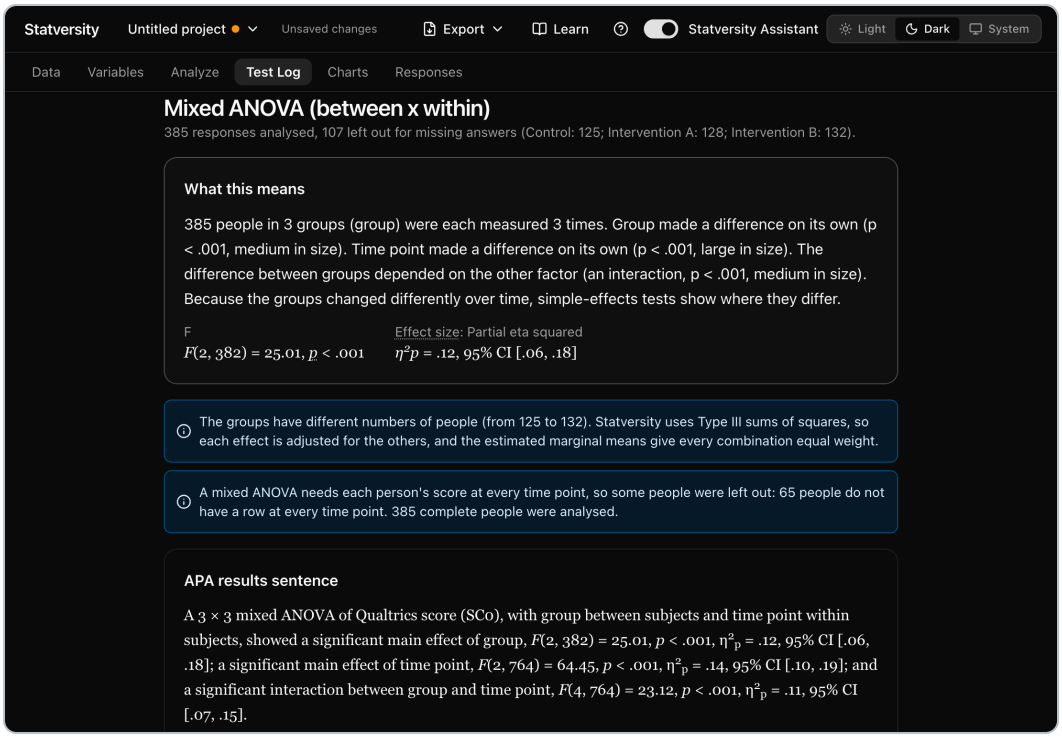


Figure 275. The top of the results: What this means, two notices, and the APA sentence.

Effect	*F*	*p*	Partial η^2	95% CI	Size label
Group (between people)	*F*(2, 382) = 25.01	< .001	.12	[.06, .18]	medium
Time point (within people)	*F*(2, 764) = 64.45	< .001	.14	[.10, .19]	large
Group × time point (interaction)	*F*(4, 764) = 23.12	< .001	.11	[.07, .15]	medium

The size labels follow the benchmarks in the reading-results chapter: .06 to under .14 is medium and .14 or more is large. The table note adds generalized η^2_G values (group .06, time point .07, interaction .05), Mauchly's $W = 1.00$ ($p = .919$) for sphericity, and Greenhouse-Geisser and Huynh-Feldt corrections that equal the uncorrected test here ($\varepsilon = 1.00$), so you can quote the uncorrected F values.

What this means in plain English

All three effects are statistically significant: differences this large would be very unlikely if they were due to chance alone. The group mattered on its own, time mattered on its own, and, most importantly, the groups changed differently over time. That is what the interaction says. Statversity's own words: "Because the groups changed differently over time, simple-effects tests show where they differ."

An interaction is the answer to "did the intervention work?" because a successful intervention means that one group's scores moved differently from the others'. A main effect of time alone would only say that everyone changed, for example because students learn anyway.

The descriptive table (Table 2 in the APA table, repeated under Descriptive statistics) shows the averages for each group and time point:

Group	Pre	Post	Follow-up
Control	12.43	12.69	12.36
Intervention A	11.88	13.84	13.81
Intervention B	11.73	16.95	15.84

Control stayed flat near 12.4. Intervention A rose by about two points and held. Intervention B rose the most (from 11.73 to 16.95 at Post) and slipped a little by Follow-up (15.84). That is the interaction in numbers.

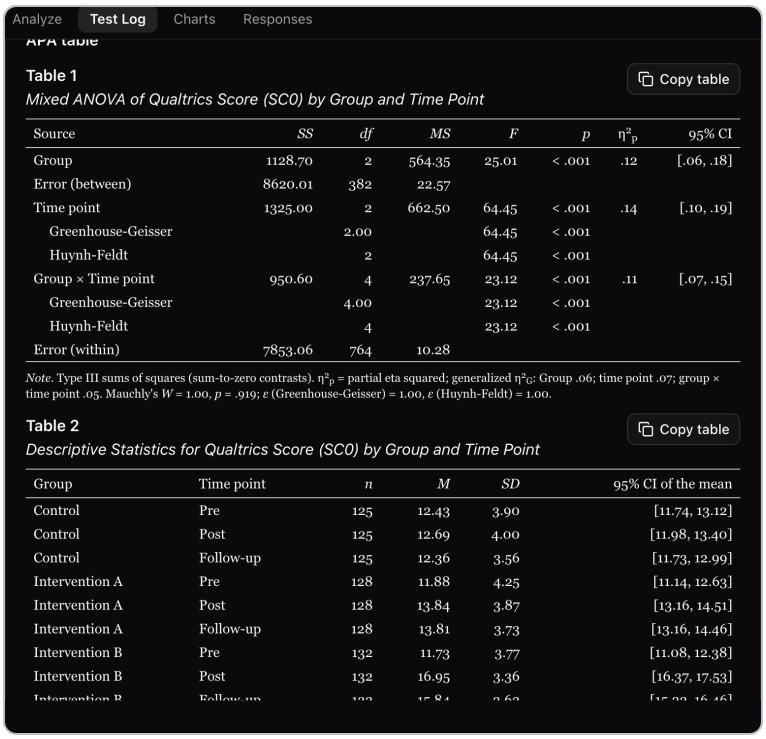


Figure 276. The full APA table: the mixed ANOVA table and the means and standard deviations for all nine group-by-time combinations.

Where the difference comes from

A significant interaction says the groups changed differently, but not which groups, or when. Because the interaction is significant, Statversity adds two follow-up sections under the ANOVA table. Each looks at the same interaction from one side.

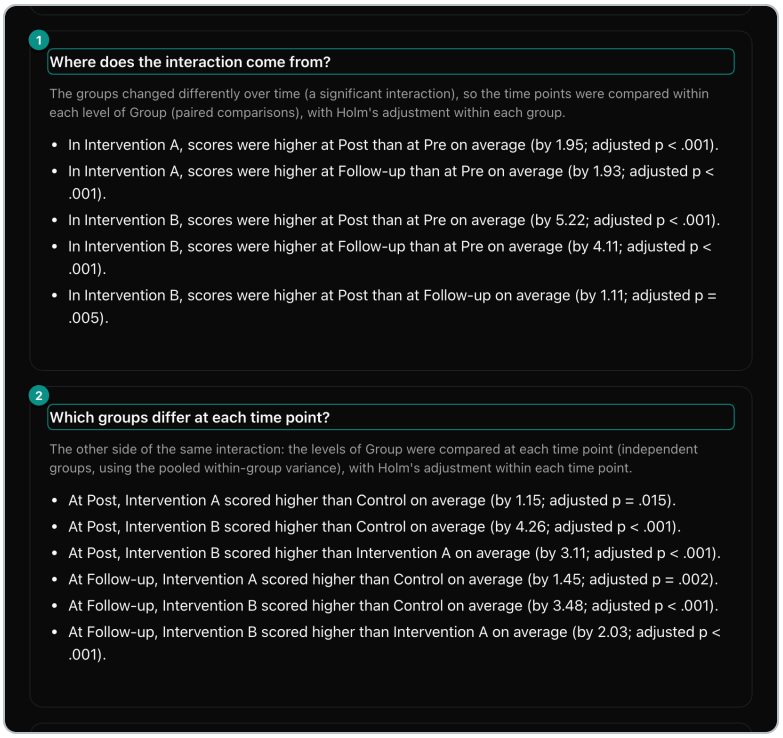


Figure 277. The two follow-up sections: 1, Where does the interaction come from? 2, Which groups differ at each time point?

1. **Where does the interaction come from?** compares the time points *within each group*: did this group's scores really change between Pre, Post and Follow-up? Here Control has no lines at all, because its scores did not change reliably. Intervention A rose from Pre to Post (by 1.95) and from Pre to Follow-up (by 1.93), and did not move between Post and Follow-up. Intervention B rose by 5.22 from Pre to Post, stayed 4.11 above Pre at Follow-up, and slipped by 1.11 from Post to Follow-up (adjusted $p = .005$).
2. **Which groups differ at each time point?** compares the groups *against each other* at one time point at a time. At Pre no group differed, which is what you want at the start of a study. At Post and at Follow-up both interventions scored higher than Control, and Intervention B scored higher than Intervention A.

Every p value is adjusted (Holm's method) for the number of comparisons made together, so the more comparisons there are, the harder each one has to work to count. Quote the adjusted values.

> **IN PLAIN LANGUAGE**
>
> **Why two sections?** A table of nine averages can be read in two directions. Reading across a row asks "did this group change?" Reading down a column asks "did the groups differ at that moment?" Both are fair questions, and neither answer can be worked out from the other, so Statversity answers both.

The two comparison figures

Each follow-up section has its own **Comparison figure**, a line for each group with brackets over the pairs that differ. Both show the same lines. They differ in which pairs get a bracket.

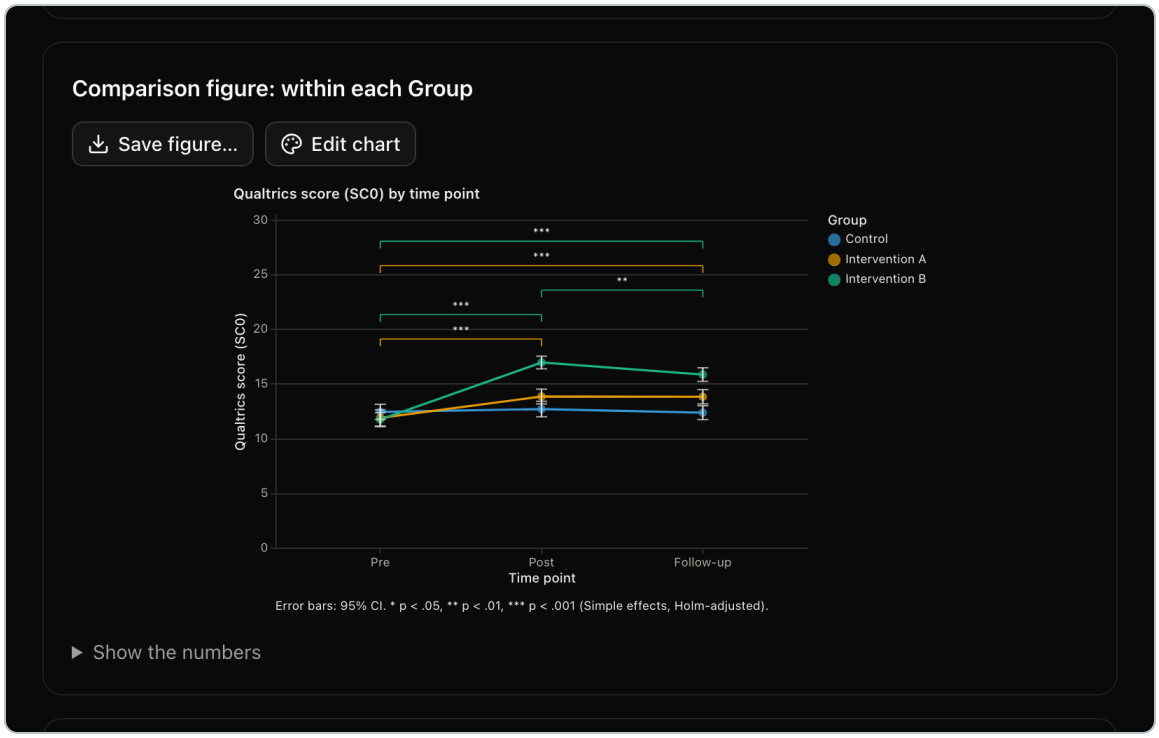


Figure 278. The comparison figure within each Group: horizontal brackets across time points, drawn in each group's colour.

Comparison figure: within each Group draws a horizontal bracket, in the group's own colour, between two time points of that group. Intervention A has two brackets (Pre to Post, Pre to Follow-up), Intervention B has three, and Control has none. Read it as "this group changed". Use it when your claim is about change: "Intervention B improved after the programme and held most of the gain."

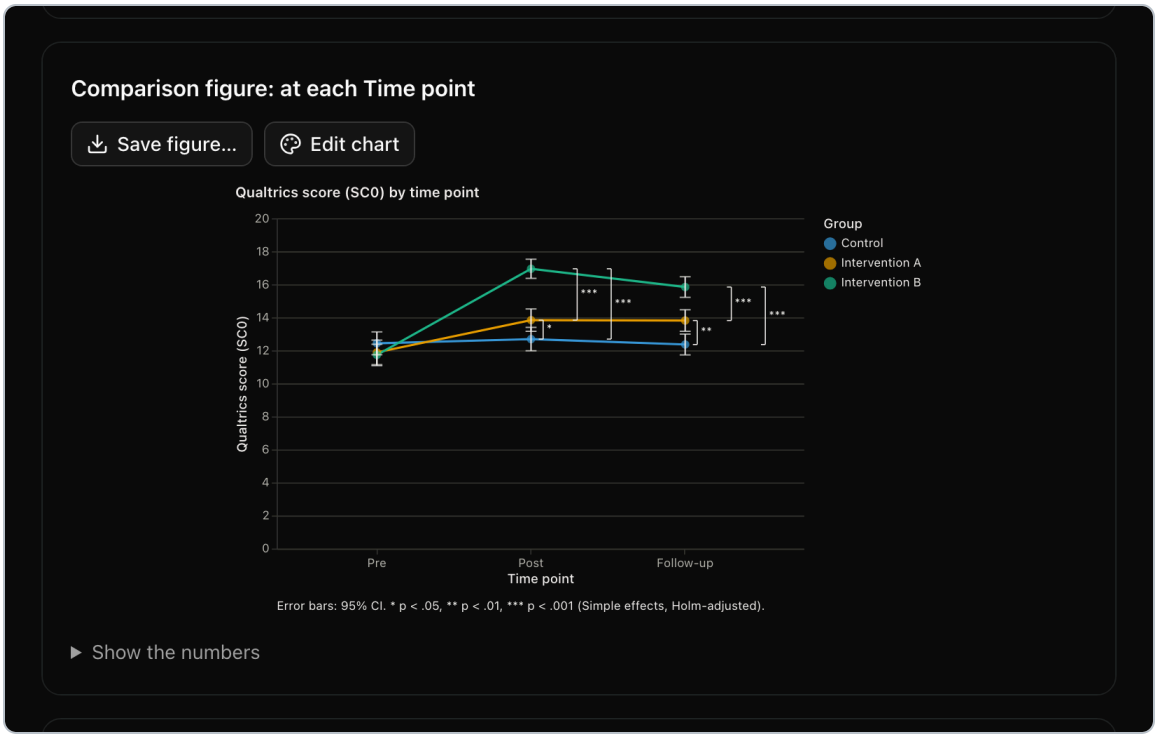


Figure 279. The comparison figure at each Time point: short vertical brackets between groups at Post and Follow-up.

Comparison figure: at each Time point draws a short vertical bracket between two groups at one time point. At Post and at Follow-up it shows the three pairs of groups; the stars say how strong the evidence is. Read it as "these groups differ at this moment". Use it when your claim is about groups: "after the programme, Intervention B students scored higher than Control students and higher than Intervention A students."

Which one to report depends on the question you asked. "Did the intervention work?" is mostly a question about change, so start with the first figure, and add the second if you want to show that the groups ended up different. Many papers report both. The stars mean the same in both: one star is $p < .05$, two are $p < .01$ and three are $p < .001$, from the adjusted values. If one figure is too crowded for a paper, you can show both directions on one figure with the **Brackets** option in **Edit chart** (see the Charts chapter).

What you would write in a paper

> A 3 × 3 mixed ANOVA of Qualtrics score (SC0), with group between subjects and time point within subjects, showed a significant main effect of group, $F(2, 382) = 25.01$, $p < .001$, $\eta^2_p = .12$, 95% CI [.06, .18]; a significant main effect of time point, $F(2, 764) = 64.45$, $p < .001$, $\eta^2_p = .14$, 95% CI [.10, .19]; and a significant interaction between group and time point, $F(4, 764) = 23.12$, $p < .001$, $\eta^2_p = .11$, 95% CI [.07, .15].

Replace “Qualtrics score (SC0)” with the name of your measure.

Watch out for

People who were not measured at every time point. After import, Statversity shows that 385 participants were matched across all three time points, 65 were not present at every time point, and none appeared twice. A mixed ANOVA needs each person’s score at every time, so Statversity says on the results page: “65 people do not have a row at every time point. 385 complete people were analysed.” Those 65 stay in your data for other analyses.

> CAUTION
>
> **Two numbers for “left out”.** The line under the heading says 107 rows were left out, while the notice says 65 people. They count different things: one counts people (65 who do not have a row at every time point) and the other counts rows that were set aside. Report the number of people (65) and the number analysed (385).

Groups of unequal size. A notice says the groups have different numbers of people (from 125 to 132), so Statversity uses Type III sums of squares: each effect is adjusted for the others, and every group-time combination gets equal weight. You do not need to do anything.

Normality cautions. The eight assumption screens were: sphericity (passed), equal covariance matrices (passed), equal spread at Pre, Post and Follow-up (all passed), and normality at Pre (passed), Post (caution) and Follow-up (caution). For Post the Shapiro-Wilk test gave $W = 0.99$, $p = .010$. With 385 scores these tests flag tiny harmless bumps, so Statversity says to trust the plots as much as the test.

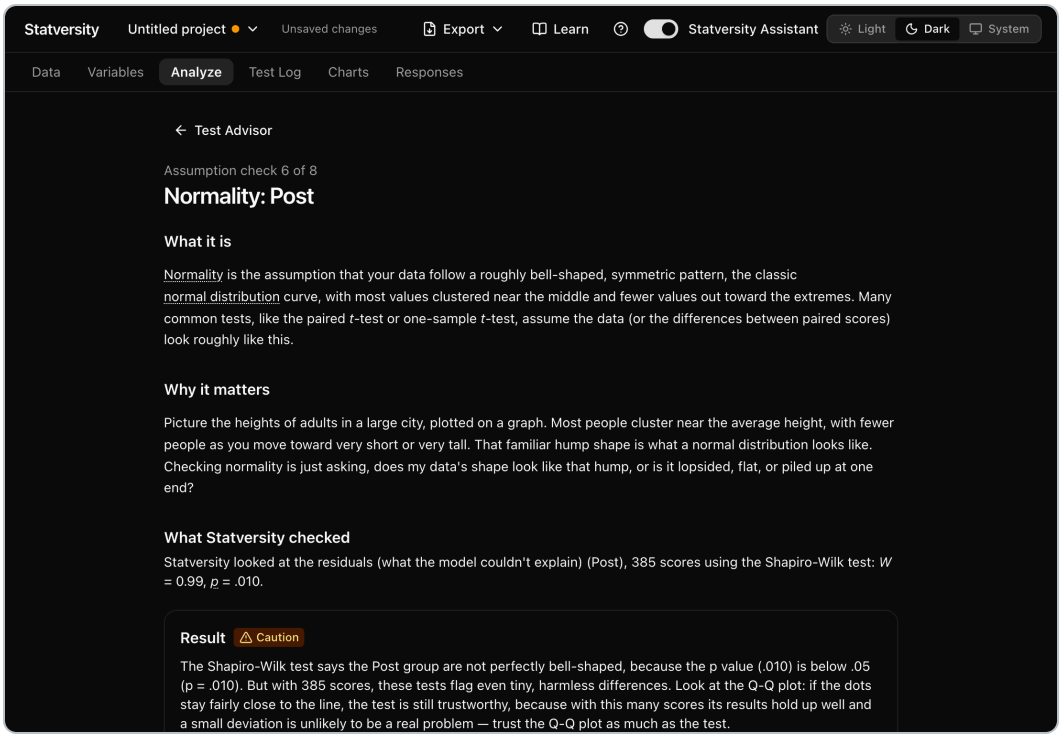


Figure 280. The Normality: Post check flagged as a caution.

On the decision screen Statversity says the checks “raised only mild cautions, which are common with larger samples” and that the recommended test is usually fine.

CHAPTER 21

Does tutoring work through confidence?

The question

Two related questions about the same students: “Does the payoff of studying depend on how confident you feel?” and “Does tutoring raise exam scores *through* making students more confident?”

A novice asks these because the first answer to “does it work?” is usually followed by “how?” and “for whom?”. The first question is **moderation** (“it depends on a third thing”). The second is **mediation** (“it works through a third thing”).

> IN PLAIN LANGUAGE
>
> **Moderation versus mediation.** Moderation: the link between two things is stronger for some people than for others. Example: studying helps everyone, but it may help the nervous students more than the confident ones. Mediation: one thing leads to a second thing, which leads to a third. Example: tutoring makes students more confident, and confident students score higher. In mediation the middle step is the “how”. Neither one, on its own, proves that anything causes anything else.

The data

240 students answered four questions: whether they attended tutoring sessions this term (yes or no), how confident they feel about the final exam (a 1 to 5 rating), about how many hours per week they studied, and their final exam score (0 to 100). The same four columns support both questions.

To open it, go to Home, click **Open practice data**, and find the card **Does tutoring work through confidence?**. It has one file.

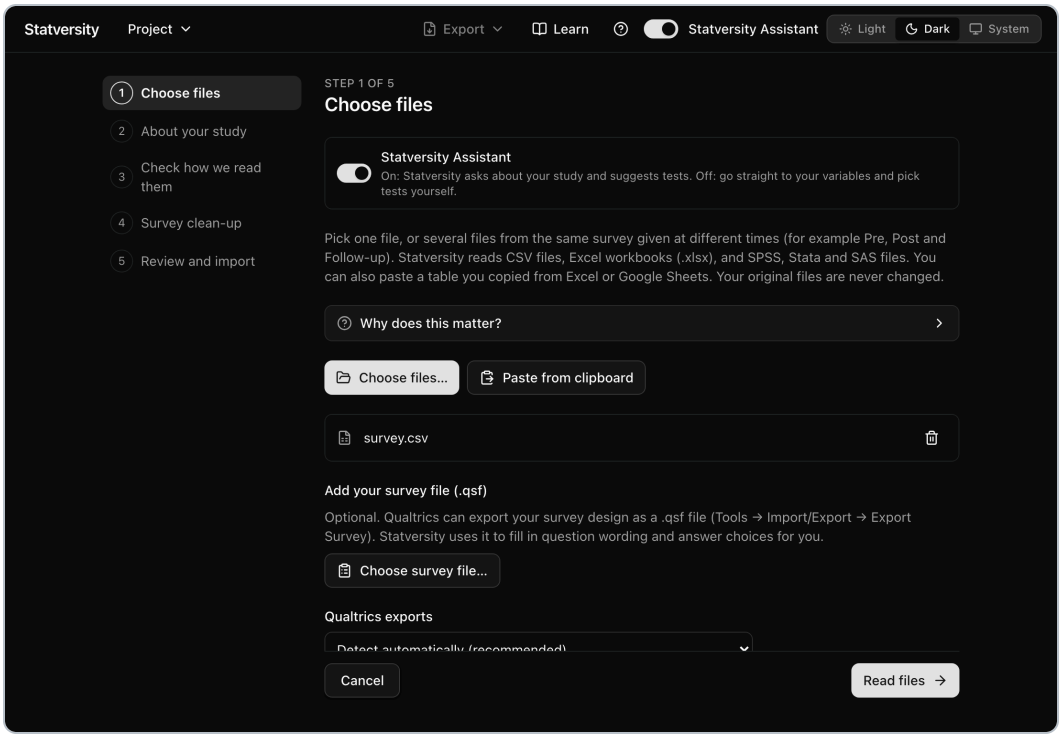


Figure 281. Choose files with the survey file filled in from the practice card.

Steps: import

1. On Home, click **Open practice data**, then **Import this dataset** on the card **Does tutoring work through confidence?**.
2. Click **Read files**, then **Skip these questions**, then **Continue** on **Check how we read them**, **Continue** on **Survey clean-up**, and **Import** on **Review and import**.
3. Accept the defaults in the Variable Interview: **Continue** on every screen, then **Finish**. You land on Variables.

Part A: moderation

Does the payoff of study time depend on confidence?

4. Click **Analyze**. Under **Which score or answer do you want to analyse?**, choose **Final exam score (0-100) (Q5)**.
5. For **What do you want to know?**, choose **Do two things go together?** and click **Continue**.
6. For **Which other variable?**, choose **About how many hours per week did you study? (Q4)**.
7. For **Does a third thing come into it?**, choose **Does the link depend on a third thing?** and click **Continue**. Statversity recommends **Moderation analysis**.

8. Click **Set up this analysis**. Statversity has already filled in the **Outcome (scores)** (the final exam score) and the **Predictor** (hours studied per week).
9. In **Moderator**, choose **How confident do you feel about the final exam? (1 = Not at all confident, 5 =… (Q3)**.

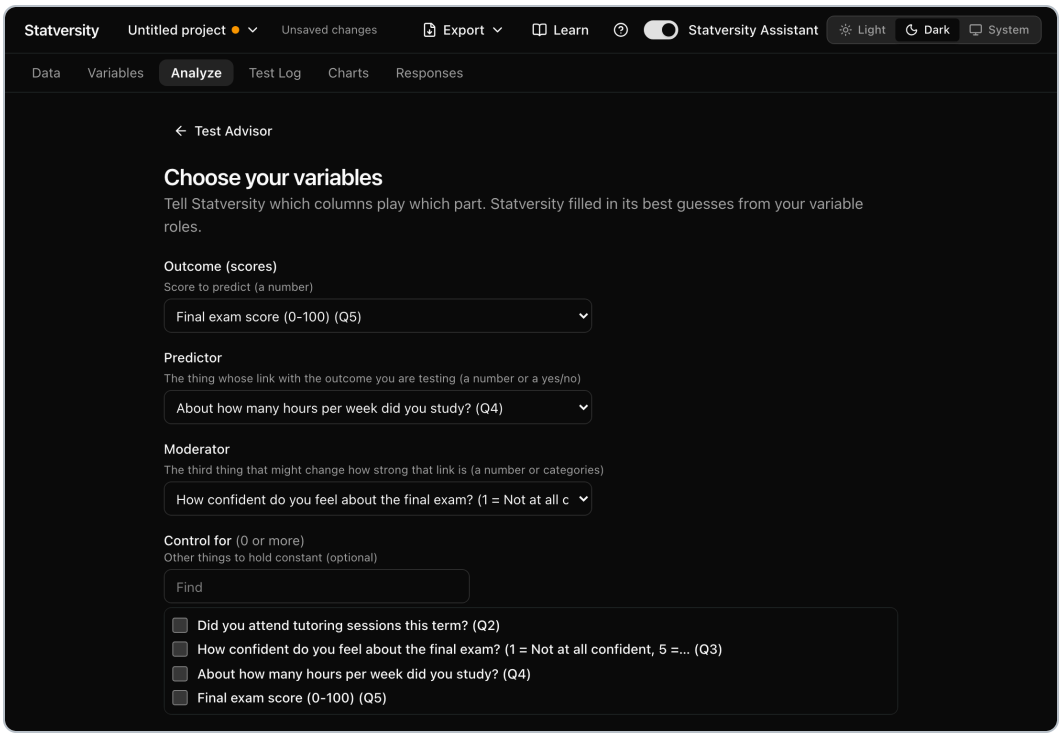


Figure 282. Choose your variables for moderation: outcome final exam score, predictor hours per week, moderator confidence.

10. Click **Check the assumptions** and then **Next check** through the five screens. On **Which test should you use?**, keep **Moderation analysis** and click **Run and see the results**.

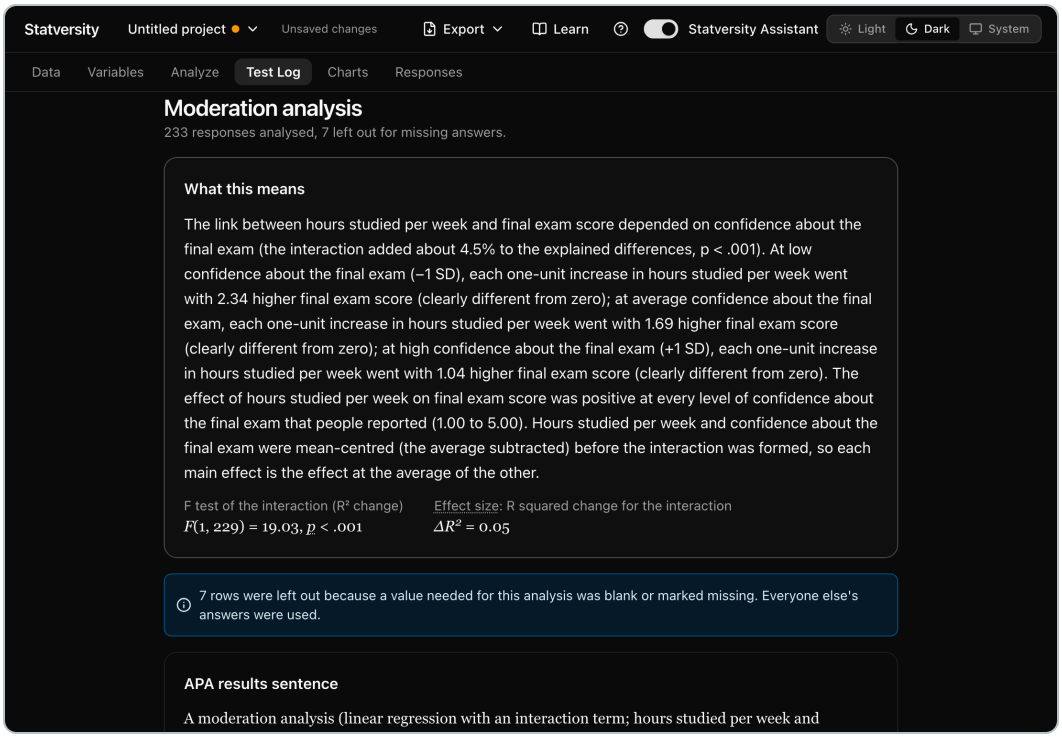


Figure 283. The top of the moderation results: What this means, the notice about seven rows, and the start of the APA sentence.

What Statversity found (moderation)

The page says 233 responses analysed, 7 left out for missing answers.

Piece	Value
Interaction (hours × confidence)	b = -0.63, $t(229)$ = -4.36, p < .001
Extra variance explained by the interaction	ΔR^2 = .05 (Statversity's label: small), $F(1, 229)$ = 19.03
Slope of hours at low confidence (-1 SD, 2.30)	2.34, p < .001
Slope of hours at average confidence (3.33)	1.69, p < .001
Slope of hours at high confidence (+1 SD, 4.35)	1.04, p < .001

The APA table also reports the whole model: R^2 = .46, with an intercept of 88.63, hours at 1.69 (95% CI [1.39, 1.99]) and confidence at 3.43 (95% CI [2.58, 4.29]).

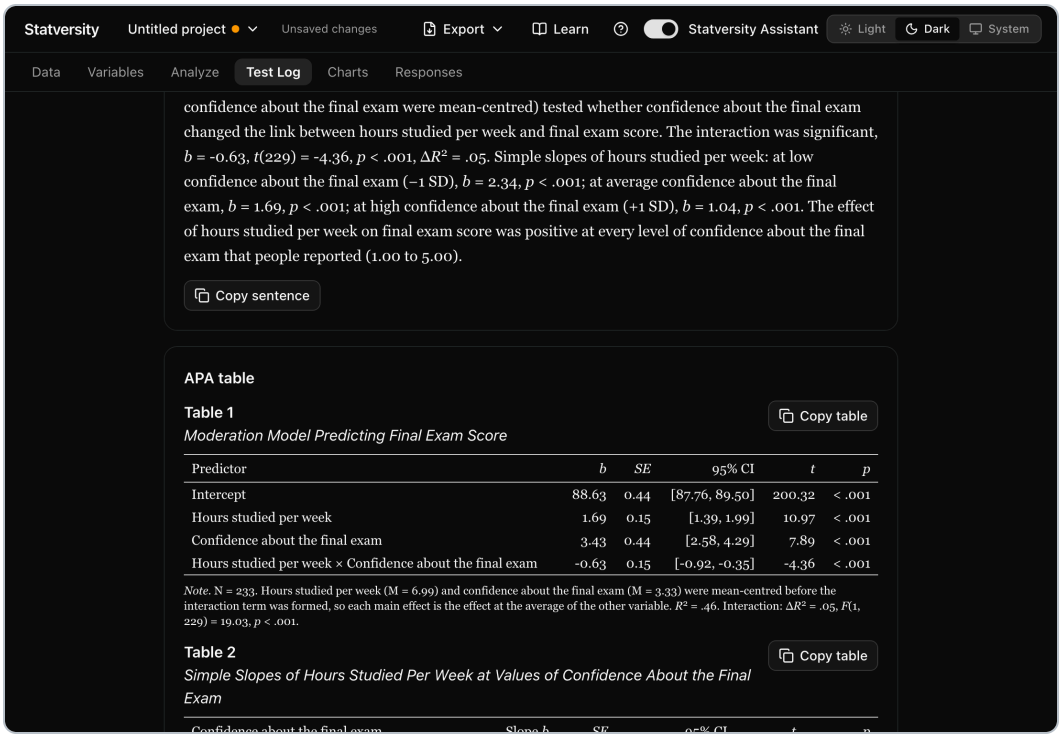


Figure 284. The APA sentence and table for the moderation, including the simple slopes.

In plain English. More study time went with higher exam scores at every level of confidence, but the payoff was bigger for the less confident. At low confidence each extra hour went with about 2.34 more points; at high confidence, about 1.04. The negative interaction (-0.63) measures that difference: for each one-point rise in confidence, the payoff of an extra hour shrinks by about 0.63 points. The interaction adds about 4.5% to the variance explained, which Statversity calls small, yet $p < .001$ says a change in slope this large would be very unlikely by chance alone. Statversity's note adds that the effect of hours was positive at every confidence level people actually reported (1.00 to 5.00).

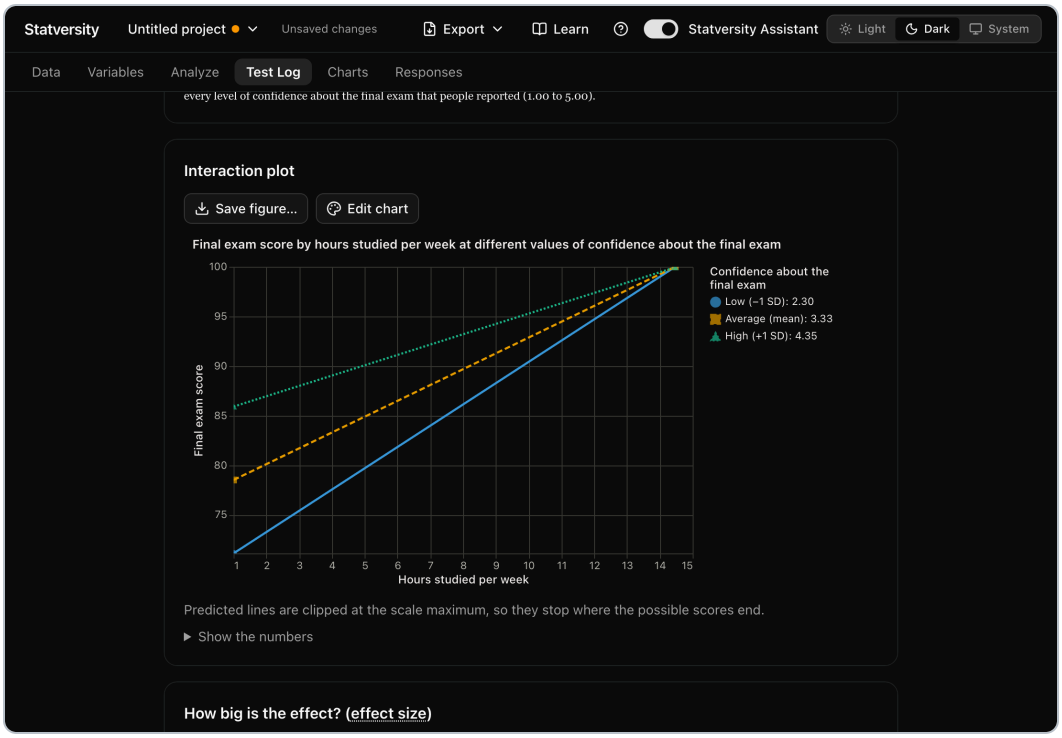


Figure 285. The interaction plot: three lines for low, average and high confidence, with the lines clipped at the top of the scale.

In the plot, the steepest line is low confidence (the blue line) and the flattest is high confidence (the green dotted line). The note under the plot says “Predicted lines are clipped at the scale maximum, so they stop where the possible scores end.”

Part B: mediation

Does tutoring work through confidence?

11. Click **Analyze**, then **Start over** to clear the Advisor.
12. Choose **Final exam score (0-100) (Q5)** again. For **What do you want to know?**, choose **Do two things go together?** and click **Continue**.
13. For **Which other variable?**, choose **Did you attend tutoring sessions this term? (Q2)**.
14. For **What kind of things are they?**, choose **One is a score and the other is a yes/no.** and click **Continue**.
15. For **Does a third thing come into it?**, choose **Does one thing work through another?** and click **Continue**. Statversity recommends **Mediation analysis**.
16. Click **Set up this analysis**. The **Outcome (scores)** and **Predictor** (tutoring) are already filled in. In **Mediator**, choose **How confident do you feel about the final exam? (1 = Not at all confident, 5 =… (Q3)**.

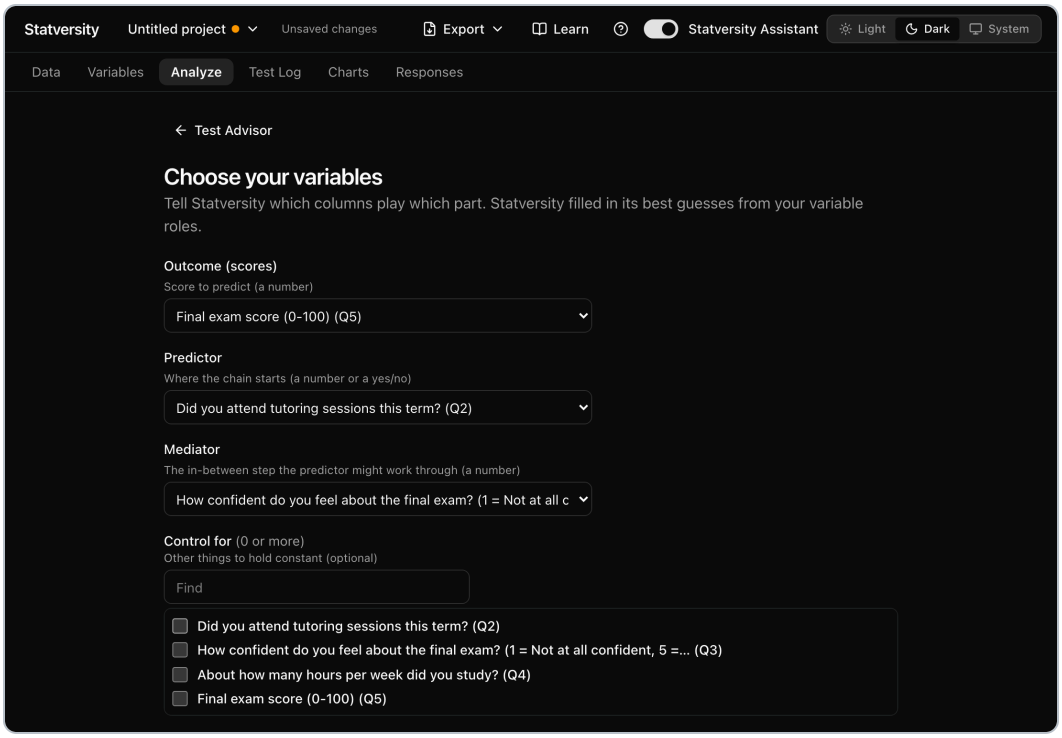


Figure 286. Choose your variables for mediation: outcome final exam score, predictor tutoring, mediator confidence.

17. Click **Check the assumptions**, then **Next check** through the four screens. Keep **Mediation analysis** and click **Run and see the results**.

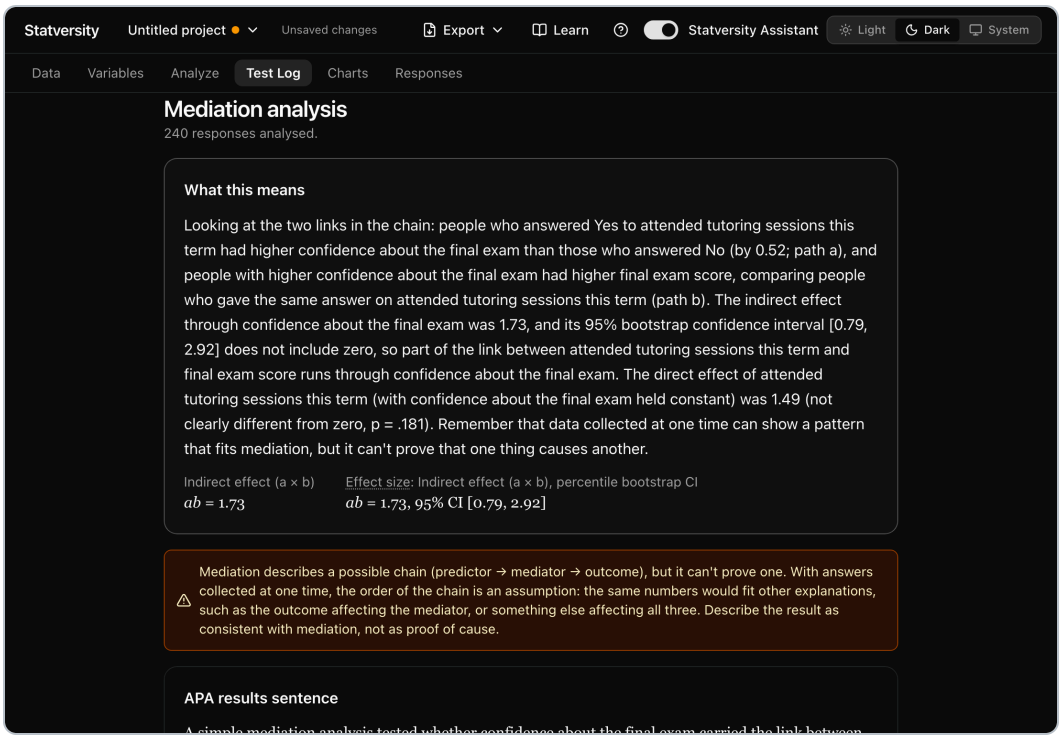


Figure 287. The top of the mediation results, with Statversity's warning that mediation cannot prove cause.

What Statversity found (mediation)

240 responses were analysed. Statversity used a percentile bootstrap with 5,000 resamples (it re-draws the sample many times to build the confidence interval).

Piece	Value
Path a: tutoring to confidence	b = 0.52, $t(238)$ = 4.04, p < .001
Path b: confidence to exam score (tutoring held constant)	b = 3.33, $t(237)$ = 6.16, p < .001
Indirect effect (a × b)	1.73, 95% CI [0.79, 2.87]
Direct effect (c′)	1.49, p = .181
Total effect ©	3.22, p = .006

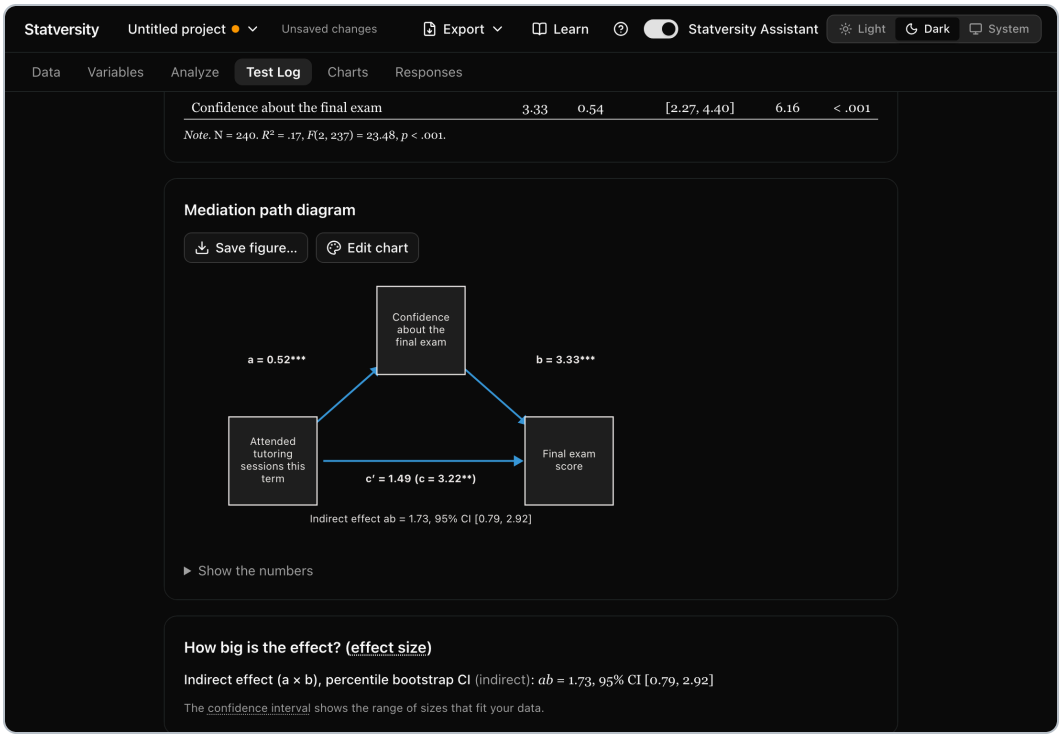


Figure 288. The mediation path diagram: tutoring to confidence (a = 0.52), confidence to score (b = 3.33), and the direct path (c′ = 1.49).

In plain English. Students who attended tutoring were more confident (by 0.52 on the 1 to 5 scale), and more confident students scored higher. The indirect route through confidence is 1.73 points, and its interval [0.79, 2.87] does not include zero, so part of the tutoring link runs through confidence. Once confidence is accounted for, the leftover direct effect of tutoring is 1.49 and not clearly different from zero (p = .181).

What you would write in a paper

For moderation:

> A moderation analysis (linear regression with an interaction term; hours studied per week and confidence about the final exam were mean-centred) tested whether confidence about the final exam changed the link between hours studied per week and final exam score. The interaction was significant, b = -0.63, $t(229)$ = -4.36, $p < .001$, $\Delta R^2 = .05$.

Follow it with the three simple slopes from the table above. For mediation, copy Statversity's sentence, which ends with: "The indirect effect was significant, ab = 1.73, 95% CI [0.79, 2.87]. The direct effect was c' = 1.49, p = .181, and the total effect was c = 3.22, p = .006."

Watch out for

> **CAUTION**
>
> **Mediation cannot prove cause.** Statversity shows this warning on the results: "Mediation describes a possible chain (predictor → mediator → outcome), but it can't prove one. With answers collected at one time, the order of the chain is an assumption: the same numbers would fit other explanations, such as the outcome affecting the mediator, or something else affecting all three. Describe the result as consistent with mediation, not as proof of cause." In a paper, write "consistent with mediation", not "tutoring raised scores by making students confident".

Seven rows left out in the moderation. The notice says "7 rows were left out because a value needed for this analysis was blank or marked missing. Everyone else's answers were used." That is why the moderation has 233 people and the mediation has 240.

Assumption cautions in both parts. Normality of residuals was a caution each time, and equal spread of residuals (homoscedasticity) failed each time. Linearity passed; multicollinearity passed in the moderation; influential cases was a caution. For both, Statversity said there is no rank-based alternative: "run it and mention the concern when you report it." Unequal spread means the scatter of errors is not the same everywhere, so standard errors and p-values may be off a little. Mention it as a limitation.

> **TIP**
>
> In the **Control for** list on Choose your variables, Statversity also lists the outcome, predictor and other columns. You do not need to tick anything there for this recipe.

CHAPTER 22

Durations and dates

The question

“How long did people take, and does age relate to the score?”

A novice asks this because survey exports are full of dates: when someone started, when they finished, when they enrolled, their birthday. Dates cannot be averaged or correlated directly. You first turn them into numbers (minutes taken, age in years), and then you can analyse them like any other score.

The data

Eighty made-up responses to an anatomy workshop survey. Each person has a StartDate and EndDate (with times), the date they Enrolled, a BirthDate, a Cohort (Lecture, Movement or Motion capture) and a Score from 0 to 100. Two BirthDate cells say “unknown”, and two people have an EndDate before their StartDate, on purpose.

To open it, go to Home, click **Open practice data**, and find the card **How long did people take, and does age relate to the score?**. It has one file, `survey.csv`.

Steps

1. On Home, click **Open practice data**, then **Import this dataset** on the card **How long did people take, and does age relate to the score?**.
2. Click **Read files**. On **About your study**, click **Skip these questions**.
3. On **Check how we read them**, find the **Dates** section. It lists how Statversity read each date column: StartDate and EndDate as year-month-day with the time of day, Enrolled as month/day/year, BirthDate with the month written out. A warning above says “2 values in BirthDate could not be read as dates and were left blank.”

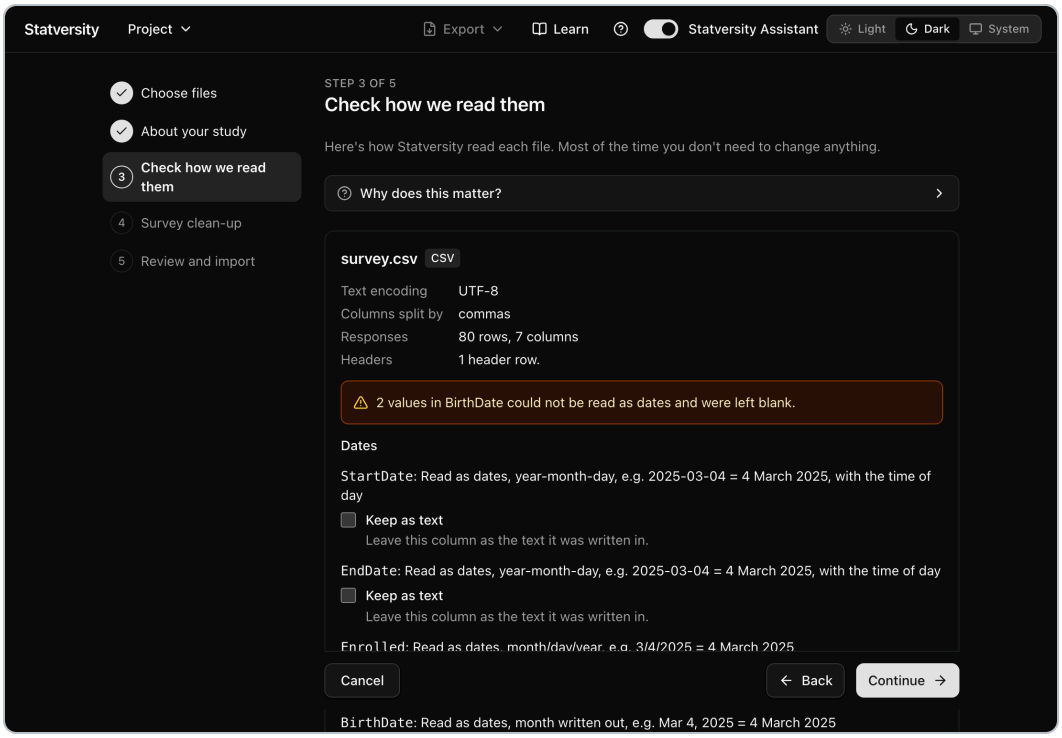


Figure 289. Check how we read them, with the Dates section and the BirthDate warning.

4. Click **Continue**. On **Survey clean-up**, click **Continue**. On **Review and import**, click **Import**.
5. Accept the defaults in the Variable Interview: **Continue** on every screen, then **Finish**. You land on Variables.
6. Click the **Variables** tab. Under each date column's name Statversity shows its type: **Date and time** for StartDate and EndDate, **Date** for Enrolled and BirthDate, with the reminder "Dates can't be averaged directly. Use New calculated variable → Time between two dates to turn them into days, weeks, months or years."

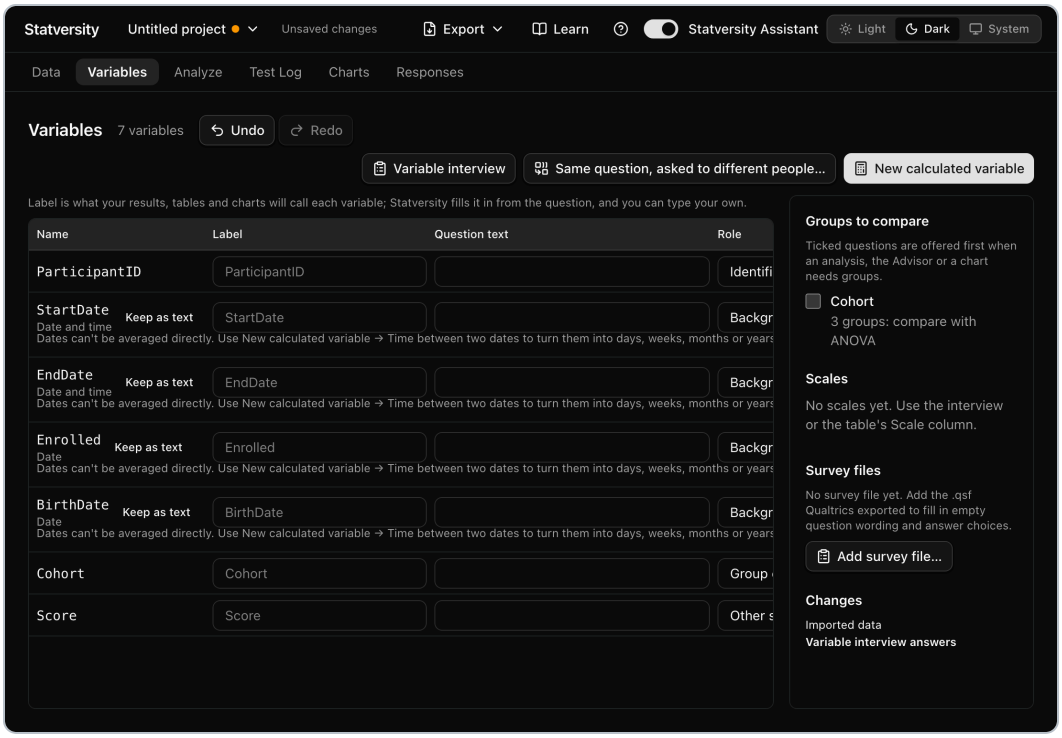


Figure 290. The Variables list with the date types and the reminder under each date column.

7. Click **New calculated variable**, and choose **Time between two dates**. Set **Later date** to **EndDate**, **Earlier date** to **StartDate**, **Unit** to **Minutes**, tick **Whole numbers only**, and type **Minutes** in **Name**. The preview shows the first 10 rows (11, 24, 23, 38, 28, 11, 23, 29, 34, 19) and says “80 rows get a value, 0 are missing.”

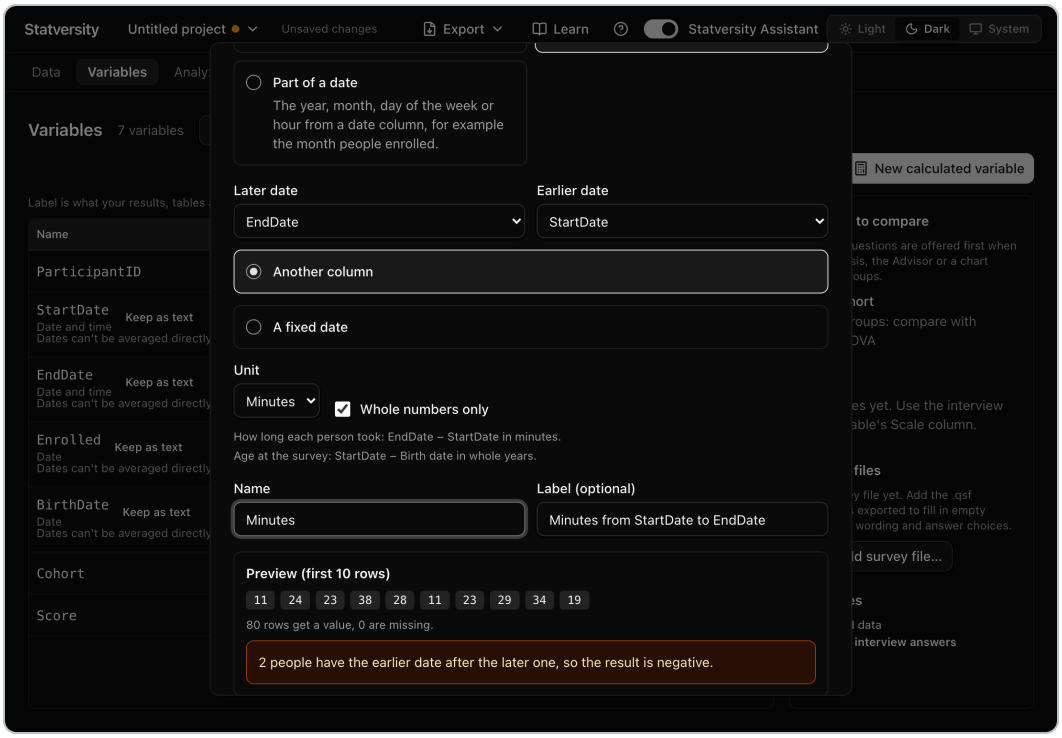


Figure 291. The New calculated variable box for Minutes, with its preview and the negative-duration warning.

8. Read the warning in the preview: “2 people have the earlier date after the later one, so the result is negative.” That is the two planted people whose EndDate is before their StartDate. Click **Add variable**. **Minutes** appears on Variables.

9. Click **New calculated variable** again, choose **Time between two dates**, set **Later date** to **StartDate**, **Earlier date** to **BirthDate**, **Unit** to **Years**, tick **Whole numbers only**, and name it **Age**.

10. Click **Add variable**. **Age** appears on Variables.

11. Go to the Data tab, click **Filter rows**, then **New filter**. For **Column to filter by** choose **Enrolled**, for **Enrolled is** choose **on or after**, and type **2026-02-01**.

12. Click **Apply**. The heading now reads “41 of 80 rows · Filter: Enrolled on or after 1 Feb 2026.” Rows outside the filter stay in the table but appear dimmed. Nothing is deleted; **Clear** brings every row back.

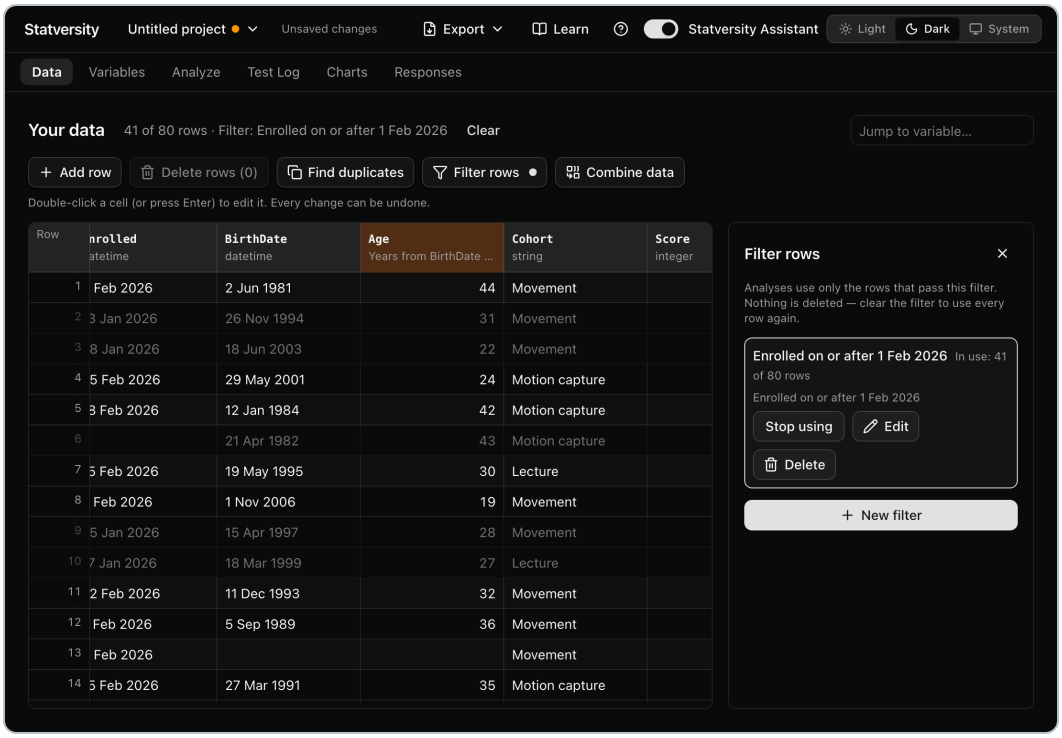


Figure 292. The Data tab with the filter applied: 41 of 80 rows, new Minutes and Age columns visible.

13. Click **Analyze**. Under **Which score or answer do you want to analyse?**, choose **Score**. For **What do you want to know?**, choose **Do two things go together?** and click **Continue**.

14. For **Which other variable?**, choose **Years from BirthDate to StartDate**. Statversity has filled in "Both are scores or totals of some kind" itself.

15. For **Does a third thing come into it?**, leave **No, just these two** and click **Continue**. Statversity recommends **Pearson correlation**.

16. Click **Set up this analysis**. **Choose your variables** shows Score and Years from BirthDate to StartDate, and "Also using the saved filter Enrolled on or after 1 Feb 2026."

17. Click **Check the assumptions**, then **Next check** through the two screens.

18. On **Which test should you use?**, Statversity suggests **Spearman's rank correlation**. Keep it and click **Run and see the results**.

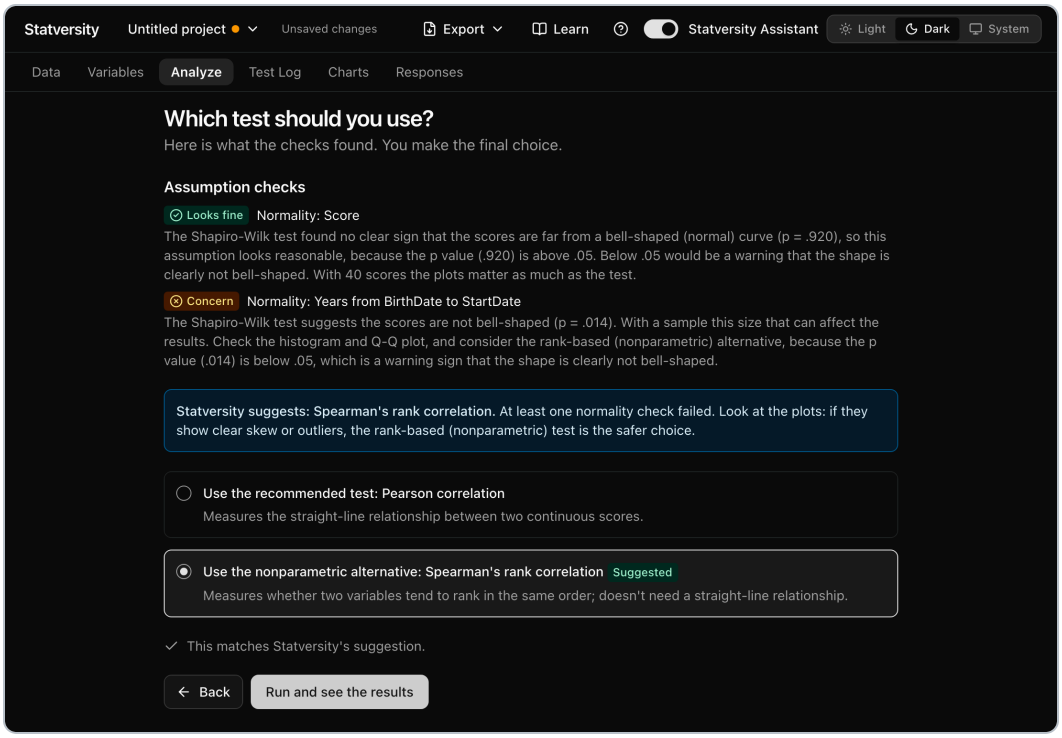


Figure 293. Which test should you use? Statversity suggests Spearman's rank correlation instead of Pearson.

> **NOTE**
>
> **A label that may surprise you.** The variable you named **Age** shows up in the Advisor, the assumption checks and the results as **Years from BirthDate to StartDate**. That is its label, filled in by Statversity when you made the calculation. It is the same column. You can see the name Age and the label side by side in the Data tab.

What Statversity found

The results page says “Enrolled on or after 1 Feb 2026 · n = 40”, and under it “40 responses analysed, 1 left out for missing answers.”

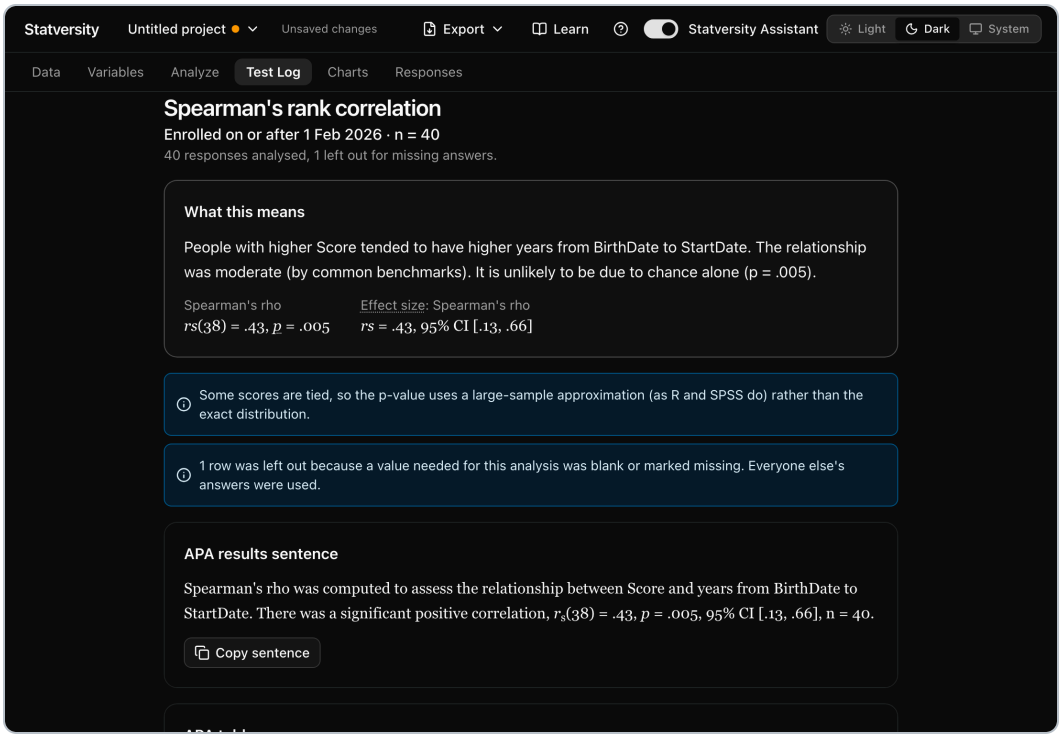


Figure 294. The top of the results: What this means, two notices, the APA sentence and the start of the table.

Piece	Value
Spearman's rho, $r_s(38)$	.43
95% confidence interval	[.13, .66]
p	.005
n	40
Label Statversity gives it	moderate

APA table

Table 1

Spearman's Rho Between Score and Years From BirthDate to StartDate

Copy table

Variables	*n*	r_s	95% CI	*p*
Score and Years from BirthDate to StartDate	40	.43	[.13, .66]	.005

Note. CI = confidence interval (Fisher z with the Fieller et al., 1957, standard error). p-value from the t approximation (tied ranks).

Figure 295. The full APA table for the Spearman correlation, with its note about the confidence interval and tied ranks.

The Descriptive statistics section gives the context for these 40 people: Score has a mean of 56.20 (SD = 10.49, range 35 to 83), and years from BirthDate to StartDate has a mean of 32.20 (SD = 8.36, range 19 to 45). Spearman's correlation has no chart in Statversity.

What this means in plain English

Among people who enrolled on or after 1 February 2026, those who were older tended to score higher. The relationship is moderate by common benchmarks (a rank correlation between .30 and under .50 counts as medium in the reading-results chapter), and it is unlikely to be chance alone (p = .005). The interval [.13, .66] is wide, which is what you expect from 40 people: the true relationship could be modest or fairly strong.

> IN PLAIN LANGUAGE
>
> **Spearman versus Pearson.** Pearson's correlation asks "do the dots follow a straight line?" and works best when both variables are roughly bell-shaped. Spearman's replaces each value by its rank (1st, 2nd, 3rd, and so on) and asks "when one goes up, does the other tend to go up too?" That makes it less thrown off by skewed data or a few extreme values.

What you would write in a paper

> Spearman's rho was computed to assess the relationship between Score and years from BirthDate to StartDate. There was a significant positive correlation, $r_s(38)$ = .43, p = .005, 95% CI [.13, .66], n = 40.

Say in your methods that the analysis was limited to people who enrolled on or after 1 February 2026, and rename "years from BirthDate to StartDate" to "age" for readers.

Watch out for

> CAUTION
>
> **The normality check failed for age, and Statversity switched tests.** Normality of Score looked fine (p = .920), but for Years from BirthDate to StartDate the Shapiro-Wilk test gave W = 0.93, p = .014, flagged as a "Concern". Statversity's advice: "At least one normality check failed. Look at the plots: if they show clear skew or outliers, the rank-based (nonparametric) test is the safer choice." That is why the suggestion changed from Pearson correlation (still shown as the recommended test on the setup screen) to Spearman's rank correlation, and why the "Which test should you use?" screen has Spearman pre-selected, marked **Suggested**.

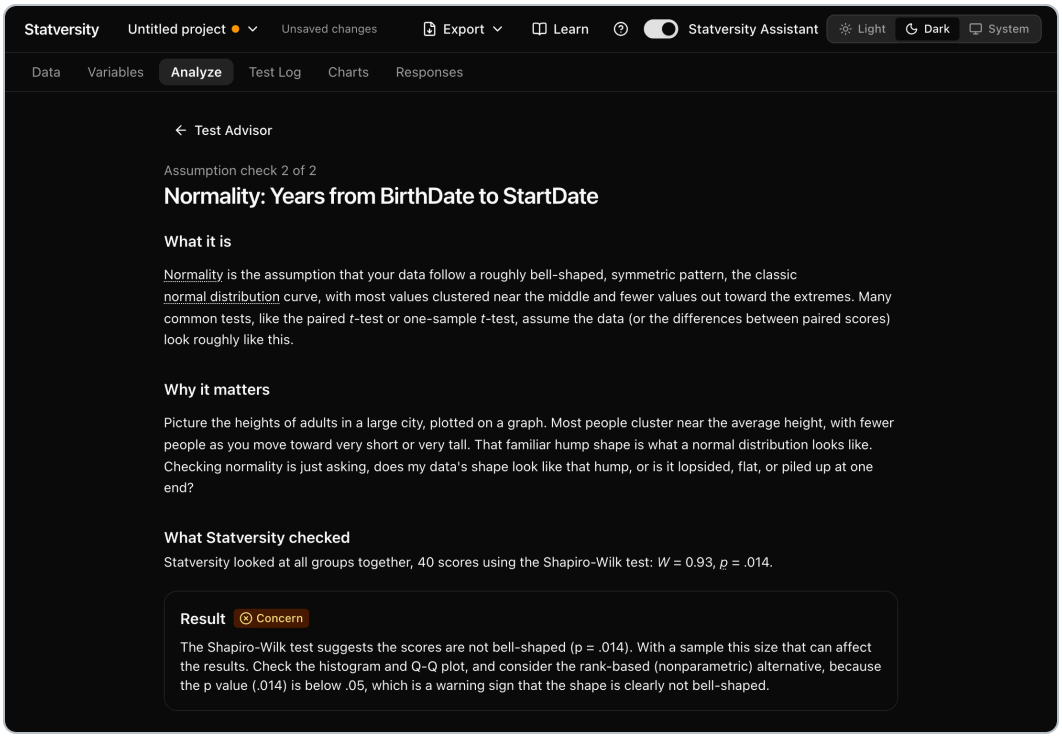


Figure 296. The normality check for age, flagged as a concern.

Two notices on the results. The first says “Some scores are tied, so the p-value uses a large-sample approximation (as R and SPSS do) rather than the exact distribution.” Ties are common with whole-number data and the approximation is standard. The second says “1 row was left out because a value needed for this analysis was blank or marked missing.” That is a person whose BirthDate was one of the two “unknown” values, so their Age was blank.

Negative durations. The two people with an EndDate before their StartDate have a negative number of minutes (for example -28 shows in one row of the filtered data). The Minutes column is not used in the correlation here, but fix or exclude those people before you average the Minutes column.

Filters apply to analyses. The results only describe the filtered group (n = 40 here), not all 80 people. The caption above the results reminds you of this.

SECTION 5

Understanding what you see

CHAPTER 23

Reading a result when you are new to statistics

Statversity does the arithmetic for you. What it cannot do is decide what the numbers mean for your study. This chapter explains the handful of ideas you need in order to read any Statversity result with confidence: p-values, effect sizes, confidence intervals, assumption checks, and sample size. It also shows you how to read the **What this means** box, the APA sentence, and the APA table.

The examples here are small made-up classroom stories, chosen so the arithmetic is easy to follow. They are not from the practice datasets, so do not expect to see these exact numbers on screen.

> **NOTE**
> You do not need to memorize any of this. Skim the chapter once, then come back to the section you need. Every technical term is also defined in the Glossary chapter.

The big picture in one paragraph

Every test answers two different questions. The first is "Could this result be a fluke?" The **p-value** speaks to that. The second is "How big is the result?" The **effect size** speaks to that. A good reading of a result always looks at both, and then asks a third question that no number can answer for you: "Does a difference this size matter in my classroom, course, or program?"

What a p-value is

Suppose you teach two sections of the same course. Section A used a new review game and Section B did not. On the final quiz, Section A averaged 78 and Section B averaged 72. Six points is a gap, but gaps appear by luck all the time. Students differ, quiz days differ, and a few strong students may have landed in one section by chance.

A statistical test starts from a deliberately boring assumption called the **null hypothesis**: the review game made no difference at all, and the six-point gap is just luck. The test then asks how often luck alone would produce a gap at least this big. The answer is the **p-value**.

If Statversity reports $p = .03$, that means: "If the review game truly did nothing, you would see a gap this large (or larger) only about 3 times in 100 by chance alone." That is fairly surprising, so you lean toward the idea that something real is going on.

IN PLAIN LANGUAGE

Think of a p-value as a surprise meter. A big p-value (say .40) means "this result would not surprise me if nothing were happening." A small p-value (say .003) means "this would be very surprising if nothing were happening." Small p-values are evidence against the boring explanation. They are not proof of anything.

The cutoff: alpha

Before you run a test you pick a line in the sand called **alpha**, almost always .05. If the p-value falls below alpha, the result is called **statistically significant**. If it does not, the result is "not statistically significant." Statversity uses .05 unless you change it.

Alpha is also your accepted false-alarm rate. At .05, you are agreeing to be fooled by luck about 1 time in 20 when nothing is really going on. A false alarm of this kind is called a **Type I error**.

What a p-value is not

CAUTION

Common misreadings of a p-value. Each of these is wrong, and each is very common.

- *"$p = .03$ means there is a 3% chance the result is a fluke."* No. The p-value assumes the null hypothesis is true and asks how surprising your data are. It is not the probability that the null hypothesis is true.
- *"$p = .03$ means there is a 97% chance the review game works."* No, for the same reason.
- *"$p = .20$ means there is no effect."* No. It means your data were not surprising enough to rule out luck. The effect might be real but too small for your sample to detect.
- *"A smaller p-value means a bigger effect."* No. A tiny difference in a huge sample can have a tiny p-value. Size is the job of the effect size.
- *"$p = .049$ is a win and $p = .051$ is a loss."* The line at .05 is a convention, not a law of nature. Those two results are nearly identical in what they tell you.

How Statversity shows very small p-values

Statversity follows APA style. A p-value that rounds to zero is written as $p < .001$ rather than $p = .000$, because a probability is never exactly zero. Other p-values are written to three decimal places with no leading zero, for example $p = .034$.

Two-tailed tests

When you compare two groups you can ask two kinds of question. A **one-tailed** question is “Is Section A higher than Section B?” It only looks for a difference in one direction you chose in advance. A **two-tailed** question is “Is Section A different from Section B, higher or lower?” It watches both directions.

Statversity runs two-tailed tests by default. That is the standard choice in most journals and theses, and it protects you: if the review game accidentally made things worse, a two-tailed test can still notice. A one-tailed test is only defensible when you decided, before seeing the data, that a difference in the other direction would be meaningless, and you can say so in your methods section.

> **IN PLAIN LANGUAGE**
> Picture a bell-shaped curve with a tail on each side. A two-tailed test saves a little of its “surprise budget” for each tail, so it asks, “Is my result in either extreme tail?” A one-tailed test spends the whole budget on one tail. That makes it easier to reach significance in that direction and impossible in the other, which is why reviewers are skeptical of it.

Effect sizes

A p-value says whether a result could be luck. An **effect size** says how big the result is, on a standard scale you can compare across studies. Statversity reports one for nearly every test, picked to match the test you ran.

Why not just use the raw difference, like “6 points”? Because 6 points means something very different on a 10-point quiz than on a 200-point exam. Effect sizes remove the units. Most of them divide the difference by how spread out the scores are, so they answer: “How big is the gap compared with ordinary student-to-student variation?”

Cohen’s d and Hedges’ g (differences between means)

Cohen’s *d* is the difference between two group means divided by a typical standard deviation. Back to the review-game example: the gap is 6 points, and scores in both sections spread out by about 12 points (the standard deviation). Then $d = 6 / 12 = 0.50$. The groups are half a standard deviation apart.

Hedges’ *g* is d with a small correction that matters when samples are small (under about 20 per group). With large samples g and d are nearly identical. For two independent groups, Statversity leads with Hedges’ g and also shows Cohen’s d. For one group compared with a fixed value, it leads with Cohen’s d and also shows g. For paired data (the same people measured twice), it leads with d_{av}, which divides by

the typical spread of the scores, and also shows d_z, which divides by the spread of the people's change scores and is often larger (chapter 14 shows both). You read all of them the same way, but say which one you report.

The benchmark labels Statversity uses, from Cohen (1988), look at the size ignoring the sign:

Size of *d* or *g* (sign ignored)	Label
under 0.20	negligible (very small)
0.20 to under 0.50	small
0.50 to under 0.80	medium
0.80 and above	large

So the review game's $d = 0.50$ is a medium effect. The sign only tells you direction, which group was higher.

Eta squared and partial eta squared (ANOVA)

When you compare three or more groups with an ANOVA, the effect size is usually **eta squared** (η^2). It is the share of all the variation in scores that is linked to group membership. If $\eta^2 = .09$, then 9% of the differences between students' scores lines up with which group they were in, and 91% is everything else: prior knowledge, sleep, luck.

Partial eta squared (partial η^2) is used when the design has more than one factor, for example teaching method and class year. It asks, for each factor, "Out of the variation not already explained by the other factors, how much does this one explain?" Because each factor is measured against a different leftover amount, partial values from one study can add up to more than 100%. That is normal, not a bug. In a design with only one factor, eta squared and partial eta squared are the same number.

Statversity's labels for both: under .01 negligible, .01 to under .06 small, .06 to under .14 medium, and .14 or more large. In repeated-measures and mixed designs Statversity may report **generalized eta squared**, which is adjusted so that numbers from different kinds of designs are fairly comparable. Statversity applies the same labels to it.

Cramer's V (counts in categories)

When both of your variables are categories, such as teaching method (A, B, C) and outcome (pass or fail), the test is a chi-square test and the effect size is **Cramér's *V***. It runs from 0 (no link between the two variables) to 1 (a perfect link). In a 2 by 2 table it is called **phi**. Statversity labels *V* with the same cutoffs it uses for *r*, shown in the next section. The in-app Learn page for Cramér's V mentions that the ideal cutoffs shift with table size, so treat the label as a rough guide.

r (correlations and rank-based tests)

r, the correlation coefficient, runs from -1 to +1. Zero means no relationship, +1 means the two variables rise together perfectly, and -1 means one rises as the other falls. A sample correlation of $r = .45$ between hours studied and exam score says that more studying goes with higher scores, moderately strongly.

Statversity also reports an *r*-type effect size for rank-based tests such as Mann-Whitney and Wilcoxon, where it is called the rank-biserial correlation. It is read on the same scale. The labels, again ignoring the sign: under .10 negligible, .10 to under .30 small, .30 to under .50 medium, .50 or more large.

> **IN PLAIN LANGUAGE**
> A correlation of .30 does not mean "30% of the variation." Squaring it does: $.30 \times .30 = .09$, so 9% of the variation is shared. That squared number is what **R-squared** reports.

Odds ratio (yes/no outcomes)

An **odds ratio** compares the odds of an outcome in two groups. Odds are different from probability: if 8 of 10 students pass, the probability of passing is 80%, but the odds are 8 to 2, which is 4. If the comparison group passes at 5 of 10, their odds are 1. The odds ratio is $4 / 1 = 4.0$.

An odds ratio of 1 means no difference. Above 1 means the odds are higher in the first group. Below 1 means lower. Unlike *d* or *r*, Statversity does not attach a small, medium, or large label to an odds ratio, because how impressive one feels depends on how common the outcome already was. Judge it against similar studies and look at the confidence interval.

> **CAUTION**
> **Odds are not probabilities.** An odds ratio of 4 does not mean students are four times as likely to pass. When the outcome is common, the odds ratio looks much more dramatic than the real change in probability. Say "the odds were four times higher," not "four times more likely."

A note about the labels

The labels above are Cohen's rules of thumb from the 1980s, drawn from general behavioral research, not specifically from education. Statversity shows them because readers expect them, and each time it adds a reminder that they are general guides. In education studies, effects judged "small" by these cutoffs are often meaningful, because classrooms are noisy and real improvements are usually modest. The best comparison is always a published study of a similar program, not a generic chart.

Confidence intervals

A sample gives you one estimate, such as “the gap was 6 points.” If you ran the study again with different students you would get 5, or 8, or 3. A **confidence interval** (CI) gives a range of values that fits your data, instead of a single number.

Statversity reports 95% confidence intervals by default. Suppose the review game gap is 6 points with a 95% CI of [1.2, 10.8]. Read it this way: “My best estimate is 6 points, and values anywhere from about 1 to 11 points are plausible given my data.” The strict meaning is about the method: if you repeated the study many times and built an interval each time, about 95 of every 100 intervals would catch the true value.

Two practical uses:

- **Precision.** A narrow interval means a precise estimate. A wide one means you do not know much yet, usually because the sample is small.
- **A second look at significance.** For a difference, if the 95% interval does not include 0, the result is significant at the .05 level. If it does include 0, “no difference” is still a plausible value. For an odds ratio, the matching “no effect” value is 1 instead of 0. This match is exact for the interval around the difference that the test itself uses; an interval around an effect size, or a p-value adjusted for several comparisons, can occasionally land on the other side of the line.

Statversity also gives confidence intervals for effect sizes themselves. These are two-sided at the requested level, so a 95% interval has a lower and an upper end.

> **CAUTION**
>
> **A 95% interval does not mean there is a 95% chance the true value is inside this particular interval.** The 95% describes how well the method works over many repetitions. It is a subtle point, and the practical reading above (“values in this range are plausible”) is a safe way to talk about it.

Assumption checks

Most tests work properly only if your data behave in certain ways. Those requirements are called **assumptions**. For a t-test or ANOVA the usual ones are:

- **Normality.** The scores (or the leftover errors) are roughly bell-shaped.
- **Equal spread** (homogeneity of variance). The groups are spread out by similar amounts.
- **Independence.** One student’s score does not influence another’s.
- **No wild outliers.** One extreme score is not driving the result.

Statversity checks the ones it can check, and shows them to you before you read the main result. It uses tests such as Shapiro-Wilk for normality and Levene's test for equal spread, and it pairs them with pictures (histograms and Q-Q plots). Independence cannot be tested from the data. It depends on how you collected them, so Statversity reminds you instead.

How to read an assumption check

For these checks the p-value works in the opposite direction from the main test. You *want* a large p-value, because the null hypothesis here is "the assumption is fine." If Levene's test gives $p = .62$, there is no evidence the spread differs, so you carry on. If it gives $p = .01$, there is evidence the spread differs.

> **CAUTION**
>
> **Do not rely on assumption p-values alone.** With a large sample, these checks flag differences too tiny to matter. With a very small sample, they can miss serious problems. This is why Statversity shows pictures as well. Look at the histogram or Q-Q plot, and ask whether the shape is badly off, not just whether a p-value crossed .05.

What Statversity does when an assumption fails

You are not stuck when a check fails. The usual repairs are:

- **Unequal spread between two groups:** nothing needs repairing. For two independent groups Statversity always runs **Welch's t-test**, a version of the t-test that does not require equal spread, and reports the classic Student's version alongside. So when only the spread check fails, Statversity keeps the recommended test.
- **Unequal spread among three or more groups:** the matching choice is **Welch's ANOVA**, whose follow-up comparisons use the Games-Howell method, which also does not assume equal spread. The Advisor recommends it when you tell it the groups' spreads differ (it assumes similar spread unless you change that answer under **Show the details** beneath its recommendation), and a standard one-way ANOVA result also reports Welch's F. If you run a standard one-way ANOVA and its spread check fails, the **Which test should you use?** screen suggests the rank-based Kruskal-Wallis test instead.
- **Badly non-normal data, ordinal data, or strong outliers:** use the matching **nonparametric** test, which works on ranks instead of raw scores. Mann-Whitney takes the place of the independent t-test, Wilcoxon signed-rank the place of the paired t-test, and Kruskal-Wallis the place of one-way ANOVA.
- **Repeated measures with unequal differences (sphericity):** Statversity reports the Greenhouse-Geisser and Huynh-Feldt corrections next to the sphericity test (Mauchly's test). When that test fails, the headline result uses the Greenhouse-Geisser correction, which adjusts the degrees of freedom so the p-value stays honest.

> **TIP**
> A failed assumption check is not a failure on your part. It is information. The Advisor recommends the test that fits your data, and the results screen tells you which version it ran and why.

"Significant" does not mean "important"

The word "significant" in statistics means only "unlikely to be pure luck." It says nothing about whether the result is large, useful, or worth acting on.

Here is the trap. A district tests an online homework program on 40,000 students and finds the program group scored 0.4 points higher on a 100-point test, $p = .003$. Statistically significant? Yes. Important? Almost certainly not: that is a gap of about 1/30 of a standard deviation, $d \approx 0.03$, and it would cost money and time to implement. The sample was so big that even a trivial gap stood out from the noise.

The reverse also happens. A teacher tries a new lab approach with 12 students and sees a 10-point gain, $d = 0.7$, but $p = .08$ because the sample is small. That is not significant, but a medium-to-large effect deserves another look with more students, not a shrug.

> **CAUTION**
> **Never report only a p-value.** Always give a p-value, an effect size, and a confidence interval together. The first says whether it could be luck, the second says how big it is, and the third says how precisely you know.

Importance is a judgment about your context. Ask: how large is the change in real units (points, grade levels, percentage passing)? What would it cost? How does it compare with other programs? Statversity can supply the numbers, but this judgment is yours.

Sample size and power

Sample size, written n (or N for a whole study), is the number of people in the analysis. It matters because a small sample is a noisy sample. With only six students per group, one unusually strong student can swing the average.

Power is the chance your study will detect a real effect if one exists. Power depends on sample size, true effect size, alpha, and how noisy the measurements are. A common target is 80%: if the effect is real and of the size you expect, you have four chances in five of getting a significant result.

An underpowered study is like trying to read small print without glasses. The text is there, but you cannot make it out. A non-significant result from a tiny sample therefore tells you very little: "I could not tell" is different from "there is no difference."

Statversity's power tools let you:

- **Plan ahead** (a priori power analysis): "I expect a medium effect and want 80% power. How many students do I need?"
- **Check what you can detect** (sensitivity analysis): "I only have 24 students. What is the smallest effect I could reliably find?"

Statversity deliberately does not offer "observed power" calculated from your own finished results. That number is just your p-value in disguise and adds no information, and methods experts consider it misleading.

> **CAUTION**
>
> **Choose the effect size for planning from earlier research or from the smallest change you would care about.** Plugging in the effect size you just observed in a small pilot tends to promise too much, because small studies often overstate effects.

Reading the "What this means" box

On every result, Statversity puts a short plain-language summary at the top, labeled **What this means**. Read it first. It is written to be understood without any statistics background, in sentences like "Scores went up from pre to post, and the improvement was large."

The box typically does these things, in order:

1. States the finding in words, naming the groups and the outcome.
2. Says whether the difference is statistically significant, in plain language rather than only a p-value.
3. Names the size of the effect ("small," "medium-sized," "large") using the benchmark labels above. The reminder that these are general rules of thumb appears with the effect size further down the page (and, for reliability and agreement results, in the box itself).
4. Explains anything easy to misread, such as which direction a negative effect size points.

Just below the box, separate notices flag other things you should know, such as rows that were left out, items that were flipped, or an item that does not fit.

Treat the box as a translation, not a verdict. It tells you what the numbers say; it cannot tell you whether the finding matters to your question. Also note that it is generated from the same numbers shown below it, so you can always check a sentence against the table.

The APA sentence

Below the summary you will find a sentence formatted in APA 7th edition style, with a copy button. It is ready to paste into a paper. A typical one looks like this (made-up numbers):

> Section A ($M = 78.0$, $SD = 12.1$) scored significantly higher than Section B ($M = 72.0$, $SD = 11.8$), $t(57.96) = 2.04$, $p = .046$, $g = 0.52$.

Read it left to right:

- ***M*** is the mean (average) and ***SD*** is the standard deviation (typical spread), one pair per group.
- ***t*(57.96)** is the test statistic, here from a t-test, with its **degrees of freedom** in parentheses. For a t-test, degrees of freedom are roughly the number of students minus the number of groups. Welch's version, which Statversity uses for two independent groups, adjusts this number for unequal spread, which is why it can have decimals. Bigger samples give bigger values. You do not need to calculate it; you just report it.
- ***p* = .046** is the p-value.
- ***g* = 0.52** is the effect size (Hedges' *g*).

Other tests use other symbols: *F* for ANOVA, χ^2 for chi-square, *U* for Mann-Whitney, *r* for correlation. The pattern is always "statistic, degrees of freedom, p-value, effect size." Statistical symbols are italicized, and values that cannot exceed 1, such as *p* and *r*, are written without a leading zero.

> **TIP**
> Always read the sentence before you paste it. Make sure it names the right groups and outcome, and replace Statversity's variable names with the wording your readers will recognize.

The APA table

Beneath the sentence is the **APA table**: the same results arranged in a standard layout, with clear column headings, group means and standard deviations, the test statistic, the p-value, the effect size, and its confidence interval when available. Use **Copy** to paste it into Word with its formatting intact.

How to read it:

1. Find the row for your comparison (a group, a factor, or a predictor).
2. Read the descriptive columns (*n*, *M*, *SD*) to see what the groups actually looked like.
3. Read the test columns (statistic, *df*, *p*) to decide whether it could be luck.
4. Read the effect size column to see how big it is, and the interval to see how precisely it is known.
5. Check any table note below the table. Notes carry details such as which correction was used or which version of the test ran.

The sentence and the table contain the same numbers. Use the sentence for a quick mention in the text and the table when you have several results to show together.

A short checklist for any result

> **TRY IT YOURSELF**
>
> Next time you run an analysis, answer these five questions before you write anything down.
>
> 1. Did the assumption checks pass, or did Statversity switch to a different test? Which one ran?
> 2. What is the p-value, and is it below your alpha (usually .05)?
> 3. What is the effect size, and which benchmark label does it earn?
> 4. What is the confidence interval, and is it narrow or wide?
> 5. Given your sample size and your field, does a difference this size matter?

If you can answer those five questions, you are reading results the way experienced researchers do. For definitions of any term in this chapter, see the Glossary.

CHAPTER 24

Glossary

This glossary defines the terms you will meet in Statversity and in this guide, in alphabetical order. Hover over a dotted term inside the app to see a shorter version of the same definition. Where a term has a longer explanation in the guide, see the chapter “Reading a result when you are new to statistics”.

A

A priori power analysis. A power calculation done before collecting data to find how many participants you need. You supply the effect size, alpha, and target power, usually 80%.

Aligned rank transform. A way to test interactions in a factorial design without assuming normal data. It first removes the other effects from the scores, then ranks what is left.

Alpha. The cutoff you choose before running a test for how much false-alarm risk you will accept, usually .05. A p-value below alpha is called statistically significant.

ANCOVA. Analysis of covariance: an ANOVA that also adjusts for a covariate, such as pretest score, so groups are compared fairly. It reports estimated marginal means.

ANOVA. Analysis of variance: a test that compares the means of three or more groups at once. It tells you whether the groups differ somewhere, not which ones differ.

APA 7. The seventh edition of the American Psychological Association style guide. Statversity formats its results sentences and tables to follow it.

Assumption. A condition your data should roughly meet for a test to give trustworthy results, such as normality or equal spread. Statversity checks the ones it can and shows you the results.

B

Bartlett’s test. A test of whether your items are correlated with each other at all, used before factor analysis. A significant result means the items are not all unrelated, the minimum needed to continue. With large samples it is almost always significant, so read it together with KMO.

Benchmark comparison. Checking a sample’s average or median against a fixed target, like a passing score, rather than against another group. One-sample tests do this.

Best-fit line. The straight line through a scatter plot that comes closest to all the points overall. It summarizes a linear relationship and lets you predict one variable from another.

Between-groups variance. How far apart the group means are from each other and from the overall mean. ANOVA compares it with within-groups variance.

Between-subjects. A design in which each person is in only one group, such as control or treatment. It is the design-level word for independent groups.

Bonferroni correction. A strict way to guard against false alarms when running many tests: divide alpha by the number of tests. It is simple and safe but can hide real effects.

Bootstrapping. Estimating how much a result would vary by resampling your own data thousands of times. Statversity uses it for confidence intervals on indirect effects.

Box plot. A chart showing the median, the middle 50% of scores as a box, and whiskers and dots for the extremes and outliers.

Box's M test. A check, used before MANOVA and MANCOVA, that the groups share a similar pattern of variances and relationships among the outcomes. It is overly sensitive with big samples.

C

Categorical variable. A variable that sorts people into named groups, like Control or Intervention A, rather than measuring an amount. The categories are not numbers you would average.

Ceiling effect. Scores bunching up near the top of a scale, so there is little room left to show improvement. It can distort simple gain scores.

CFI and RMSEA. Two common fit indices for confirmatory factor analysis. CFI is better closer to 1 (above about .95 is common); RMSEA is better closer to 0 (below about .06 is common).

Chi-square statistic. The test statistic that compares observed counts in categories with the counts expected by chance. Bigger mismatches give a bigger value and a smaller p-value.

Chi-square test. A test for counts in categories. The test of independence asks whether two categorical variables are related; the goodness-of-fit test asks whether one variable matches an expected pattern.

Cohen's d. An effect size for the difference between two means, measured in standard deviations. By Cohen's labels, 0.20 is small, 0.50 medium, and 0.80 large.

Cohen's f. An effect size for ANOVA-family tests, built so it can be used directly in power analysis. Statversity's labels are .10 small, .25 medium, and .40 large.

Cohen's kappa. A measure of how well two raters agree on categories, after removing the agreement expected by chance. See kappa.

Confidence interval. A range of values that fits your data, usually the 95% version, instead of a single best guess. A narrow interval means a more precise estimate.

Contingency table. A grid of counts for every combination of two categorical variables, such as teaching method by pass or fail. Also called a crosstab.

Continuous. Data that can take any value along a range, like a test score or reaction time. Scale scores built from many items are usually treated as continuous.

Correlation. A number from -1 to +1 showing how strongly two variables move together. It describes a relationship but never proves one thing causes the other.

Correlation matrix. A table of the correlations between every pair of variables in a set. Because it runs many tests at once, the results are usually adjusted for multiple comparisons.

Covariate. An extra variable you statistically control for, such as pretest score, so the main comparison is cleaner.

Cramér's V. An effect size for a chi-square test on categories, from 0 (no link) to 1 (a perfect link). In a 2 by 2 table it is called phi.

D

Degrees of freedom. A bookkeeping number (df) tests use to pick the right reference distribution. Roughly, it counts how many values were free to vary. It appears in parentheses in APA sentences.

Dependent variable. The outcome you are measuring, such as a test score. Its value may depend on the independent variable.

Dichotomous variable. A variable with exactly two possible values, like pass/fail or yes/no.

Direct effect. In mediation, a predictor's effect on the outcome once the mediator is held constant. It is written c-prime.

E

Effect size. A number saying how big a difference or relationship is, on a standard scale, rather than whether it could be due to chance. Two studies can both have $p < .05$ while one effect is tiny and the other large.

Epsilon squared. A less biased cousin of eta squared that estimates the share of variation linked to group membership. It is read the same way, using the same size labels.

Estimated marginal means. Group means adjusted to what they would be if every group started at the same covariate level. ANCOVA reports these instead of raw averages.

Eta squared. An effect size for ANOVA: the share of all the variation in scores that is linked to group membership. Statversity's labels are .01 small, .06 medium, and .14 large.

Expected frequency. The count you would expect in a category or cell if there were no real pattern. A chi-square test compares it with the observed frequency.

F

F-squared. An effect size for regression and ANCOVA based on R-squared: for a whole model it is R-squared divided by (1 minus R-squared), and for one predictor it is based on the change in R-squared when that predictor is added. Statversity's labels are .02 small, .15 medium, and .35 large.

F-statistic. The test statistic from ANOVA-type tests: variation between the group means divided by variation within the groups. A large F points to a small p-value.

Factor loading. A number from -1 to 1 showing how strongly a survey item connects to an underlying factor. Higher absolute values mean a clearer indicator.

Family-wise error rate. The chance of at least one false alarm across a whole set of related tests. Corrections such as Bonferroni and Holm aim to keep it near alpha.

Fisher's exact test. A test for a table of counts, most often 2 by 2 (Statversity also handles larger tables), that stays accurate when counts are small, where the chi-square test would be unreliable.

Friedman test. A rank-based test for the same people measured three or more times. It takes the place of repeated-measures ANOVA when the data are not normal.

G

Gain score. A student's posttest score minus their pretest score, a simple measure of change.

Games-Howell. A follow-up comparison method that does not assume the groups have equal spread. Statversity uses it after Welch's ANOVA.

Generalized eta squared. A variant of eta squared adjusted so effect sizes stay comparable across between-subjects, within-subjects, and mixed designs.

Goodness of fit. How well one categorical variable's observed counts match an expected pattern, such as an even split across four answer choices.

Greenhouse-Geisser correction. An adjustment to the degrees of freedom in repeated-measures ANOVA when sphericity is violated, so the p-value stays honest.

H

Hedges' g. Cohen's d with a small-sample correction. It is read the same way, with the same labels, and is nearly identical to d when samples are large.

Histogram. A bar chart showing how many scores fall into each range of values. It is a quick way to check shape, spread, and outliers.

Homogeneity of regression slopes. The ANCOVA assumption that the covariate predicts the outcome with the same slope in every group.

Homogeneity of variance. The assumption that the groups being compared are spread out by roughly the same amount. Levene's test checks it.

I

Independence. The assumption that one person's score does not influence another's. It depends on how data were collected, so it cannot be tested from the data alone.

Independent groups. Separate, unrelated sets of people, such as a control class and a different treatment class. Each person contributes data to only one group.

Independent variable. The factor you compare or manipulate, such as group, time, or teaching method.

Indirect effect. In mediation, the part of a predictor's effect that runs through the mediator (path a times path b). Statversity tests it with a bootstrap confidence interval.

Interaction effect. When the effect of one factor depends on the level of another. An intervention might help younger students a lot but older students barely.

Intercept. The predicted outcome when every predictor in a regression equals zero. It anchors the regression line.

Internal consistency. How well the items on a scale agree with each other, one type of reliability. Cronbach's alpha and McDonald's omega estimate it.

Interquartile range. The distance between the 25th and 75th percentiles, covering the middle 50% of scores. It describes spread without being thrown off by outliers.

Interrater reliability. How consistently different raters give the same scores or classifications to the same items. ICC and kappa are common measures.

Intraclass correlation (ICC). A measure of agreement among raters on numeric or ordinal scores, from 0 (none) to 1 (perfect). Its absolute-agreement forms, including the ICC(2,1) that Statversity's results sentence reports, penalize raters who are consistently higher or lower than each other; its consistency forms do not.

Item difficulty. The share of test-takers who answered an item correctly, from 0 to 1. Despite the name, a higher number means an easier item.

Item discrimination. How well a test item separates students who did well overall from those who did not. A negative value is a red flag.

Item-total correlation. How strongly one item correlates with the total of the rest of the scale. A low or negative value suggests the item does not belong or needs reverse-scoring.

J

Johnson-Neyman range. The range of moderator values where a predictor's effect is statistically significant. Trust it only within the range of moderator values actually observed.

K

Kappa. A measure of agreement on categories, adjusted for chance. 0 means no better than chance and 1 means perfect agreement. Cohen's kappa is for two raters, Fleiss' kappa for three or more.

Kendall's tau-b. A rank-based correlation suited to ordinal data and many ties. It detects monotonic relationships.

Kendall's W. An effect size and agreement measure for rank-based tests with several conditions or raters, from 0 to 1.

KMO. The Kaiser-Meyer-Olkin measure: whether your items are correlated enough to justify factor analysis. Values above about .60 are generally acceptable.

Kruskal-Wallis test. A rank-based test that compares three or more independent groups. It takes the place of one-way ANOVA when the data are not normal.

Kurtosis. How heavy-tailed or peaked a distribution is compared with a normal bell curve. High kurtosis means more extreme scores.

L

Levene's test. A test of whether groups have similar spread. A small p-value suggests unequal spread and points toward a version of the test that does not assume equal spread, such as Welch's. Statversity uses the Brown-Forsythe version, which measures spread around the median.

Likert scale. A rating scale such as 1 (strongly disagree) to 5 (strongly agree). A single item is ordinal; several combined into a total are usually treated as continuous.

Linear relationship. A relationship that follows a straight line on a scatter plot. Pearson's correlation measures its strength.

Logit. The natural log of the odds of an outcome. Logistic regression works on this scale, then converts to odds ratios for reporting.

M

Mahalanobis distance. A distance that flags unusual data points across several variables at once, including ones that look ordinary on each variable alone.

Main effect. The overall effect of one factor, averaged across the levels of the other factors.

Mann-Whitney test. A rank-based test that compares two independent groups. It takes the place of the independent t-test when the data are not normal.

Mean. The everyday average: add up all the scores and divide by how many there are. A few extreme scores can pull it a lot.

Mean centering. Subtracting a variable's average from every score so that zero means average. Statversity does this for moderation analyses.

Median. The middle score when all values are lined up from lowest to highest. It is not dragged around by outliers.

Mediator. A variable thought to explain how or why a predictor relates to an outcome, sitting in the middle of a proposed causal chain between them. A mediation analysis can be consistent with that chain but cannot prove it.

Model fit indices. Numbers such as CFI, RMSEA, and SRMR that summarize how well a proposed factor structure matches your data. Report several together.

Moderator. A variable that changes how strong, or in which direction, another relationship runs. It shows up statistically as an interaction.

Monotonic relationship. A relationship in which one variable consistently rises (or falls) as the other does, though not necessarily in a straight line. Spearman's correlation detects it.

Multiple comparisons. Running many tests at once, which raises the chance of a false alarm somewhere. Statversity explains the problem and lets you choose a correction.

Multivariate normality. The assumption that several outcome variables together follow a bell-shaped pattern, not just each one alone.

N

Negligible. Statversity's label for an effect smaller than the lowest benchmark, so small it is hard to distinguish from nothing in practice.

Nominal. Category labels with no natural order, such as school or eye color. You can count them but not average them.

Nonparametric. A test that does not assume a normal distribution, often working on ranks. It is the backup plan for non-normal data, ordinal data, or outliers.

Normal distribution. The symmetric bell-shaped curve, with most scores near the middle and fewer toward the extremes.

Normality. The assumption that scores (or differences, or residuals) are roughly normally distributed. Statversity checks it with the Shapiro-Wilk test and Q-Q plots.

Normalized gain. A student's actual gain divided by the most they could have gained, given their starting score. It makes gains comparable across different starting points.

Null hypothesis. The boring explanation a test tries to rule out: no difference, no relationship, no effect.

O

Observed frequency. The actual count of observations that landed in a category in your data.

Odds ratio. How many times the odds of an outcome multiply for each one-unit increase in a predictor, or between two groups. An odds ratio of 1 means no effect. It is not the same as "times more likely."

Omega squared. A less biased effect size for ANOVA that estimates the share of variation linked to group membership. It is read like eta squared.

One-tailed test. A test that looks for an effect in one direction only. It is more powerful in that direction but blind to the other, so Statversity defaults to two-tailed.

Ordinal. Data with a meaningful order but no guarantee that the gaps between values are equal. A single Likert item is a classic example.

Outlier. An unusual data point far from the rest, which can distort means and some tests.

P

p-value. The probability of seeing results at least this extreme if the null hypothesis were true. It is not the chance the null is true, and it does not measure effect size.

Paired. The same people (or matched pairs) measured twice, such as pretest and posttest. Also called dependent data.

Pairwise comparison. A comparison between exactly two groups at a time. Post hoc tests run every relevant pair.

Parallel analysis. A method that compares your factors with those from random data to decide how many are worth keeping.

Parametric. A test that assumes the data follow a specific pattern, usually roughly normal, with similar spread across groups. The t-test and ANOVA are examples.

Partial correlation. The correlation between two variables after a third variable's shared influence is removed.

Partial eta squared. An ANOVA effect size for designs with several factors: how much of the variation not already explained by the other factors a given factor explains. Values from different factors can add to more than 100%.

Path diagram. A picture of a factor model: circles for factors, boxes for items, and arrows showing loadings.

Phi. The effect size for a 2 by 2 table of counts. It is Cramér's V for the two-by-two case and is read like a correlation.

Pillai's trace. Statversity's default test statistic for MANOVA and MANCOVA, chosen because it holds up when assumptions are somewhat violated.

Post hoc. Follow-up comparisons run after an overall test to find exactly which groups differ, with corrections to control false alarms.

Post hoc (observed) power. Power calculated from your own finished results. It is circular and misleading, so Statversity does not offer it.

Power. The chance your test will detect a real effect if one exists. It rises with bigger samples, bigger true effects, and less noisy measurement.

Power analysis. A calculation linking sample size, effect size, alpha, and power. Give any three and it solves for the fourth.

Proportional odds assumption. The assumption, in ordinal regression, that a predictor's effect is the same across every boundary between adjacent categories.

Q

Q-Q plot. A chart that compares your data's values with what a perfect normal distribution would give. Points near the diagonal line mean the data are close to normal.

R

r. The correlation coefficient, from -1 to +1. Also the name of Statversity's effect size for rank-based tests, read on the same scale: .10 small, .30 medium, .50 large.

R-squared. The proportion of variation in an outcome that is explained by, or lines up with, the predictor or predictors.

Rank. A score's position in order (1st, 2nd, 3rd) instead of its raw value.

Rank-biserial correlation. An effect size for rank-based tests such as Mann-Whitney and Wilcoxon, read on the same scale as r.

Rank transform. Replacing raw scores with their ranks before analysis, which makes a test less sensitive to outliers and skew.

Regression coefficient. The slope for one predictor: how much the predicted outcome changes for each one-unit increase in that predictor, holding the others constant.

Reliability. Whether a measure gives consistent results, such as a scale whose items agree with each other. Reliability is about consistency, not accuracy.

Repeated measures. The same people measured several times, such as pretest, posttest, and follow-up. It needs tests built for correlated data.

Reverse-scoring. Flipping the numbers on negatively worded survey items so that a high score always means the same thing. On a 1-5 scale the new score is 6 minus the answer.

S

Sample size. How many people (or rows of data) went into your analysis, written n. Bigger samples give more precise estimates and more power.

Scale score. A single combined score made by adding or averaging several related survey items.

Scree plot. A chart of factor sizes from largest to smallest, used to decide how many factors to keep by finding the elbow where the line flattens.

Sensitivity analysis. A power analysis with a fixed sample size that finds the smallest effect you could reliably detect.

Shapiro-Wilk test. Statversity's main test for whether data look normally distributed. It becomes oversensitive with large samples, so Statversity also shows a Q-Q plot.

Simple slope. A predictor's effect on the outcome at one specific value of the moderator, such as one standard deviation above its mean.

Skewness. How lopsided a distribution is. A long tail of high scores is right-skewed; a long tail of low scores is left-skewed.

Sobel test. An older significance test for an indirect effect, shown only for teaching comparison. Statversity relies on the bootstrap instead.

Spearman-Brown correction. A formula that adjusts a split-half estimate upward to reflect the full-length scale.

Sphericity. The assumption, for repeated measures with three or more time points, that the differences between all pairs of time points have similar variance. Mauchly's test checks it.

Split-half reliability. A reliability check that splits a scale in two and correlates the halves. The Spearman-Brown correction then adjusts for the shorter length.

Standard deviation. A typical distance between individual scores and the mean. A bigger SD means more spread out.

Standard error. How much a sample statistic, such as a mean, would bounce around if you repeated the study. It shrinks as the sample grows.

Statistically significant. A result whose p-value is below alpha, meaning it is unlikely to be pure luck. It says nothing about whether the result is large or important.

Sum of squares. A total of squared distances from a mean. ANOVA splits it into a between-groups piece and a within-groups piece.

T

t-statistic. The test statistic from a t-test, roughly the size of a difference compared with its noise.

t-test. A test that compares means: two independent groups, the same people twice (paired), or one group against a fixed value (one-sample).

Test statistic. The single number a test calculates from your data, such as t, F, or chi-square, which is then converted to a p-value.

Ties. Two or more scores that are exactly equal. They complicate ranking-based tests, which share the ranks and often apply a small correction.

Total effect. In mediation, a predictor's overall effect on the outcome ignoring the mediator. It equals the direct effect plus the indirect effect.

Two-tailed test. A test that checks for a difference in either direction, higher or lower. It is Statversity's default and the usual choice for papers.

Type I error. A false alarm: concluding there is an effect when there really is not. Alpha sets how much of this risk you accept.

Type III sums of squares. The way Statversity splits up variation in factorial ANOVA and ANCOVA, matching SPSS. Each effect is judged after accounting for all the others.

V

Validity. Whether a measure actually captures the concept it claims to measure. A measure can be reliable without being valid.

Variance. The average squared distance of scores from the mean; the square of the standard deviation.

Variance inflation factor. A number (VIF) showing how much a predictor overlaps with the other predictors. A value above about 5 to 10 signals a multicollinearity problem.

W

Welch's ANOVA. A version of one-way ANOVA that does not require equal spread across groups. Follow-ups use Games-Howell.

Welch's t-test. A version of the independent t-test that does not require equal spread in the two groups. It is the safer choice when Levene's test flags unequal spread, and Statversity runs it by default for every independent-samples t-test (Student's version is reported alongside).

Wilcoxon signed-rank test. A rank-based test for paired data. It takes the place of the paired t-test when the data are not normal.

Within-groups variance. How much individual scores vary inside each group, ignoring group differences. It is the baseline noise in ANOVA.

Within-subjects. A design factor in which the same people experience every condition, such as every time point. It is the design-level word for repeated measures.

Z

z-score. How many standard deviations a score is above or below the mean. A z-score of 2 is two standard deviations above.

CHAPTER 25

Troubleshooting and FAQ

This chapter lists the problems people most often run into, grouped by the stage of the work where they show up. Each entry has three parts: what you see (in the words people usually use for it), why it happens in one sentence, and the exact steps to fix it. Scan the headings for your symptom.

> **TIP**
> Before anything else, check three things. Is Statversity's "engine ready" card showing? Is the project saved (no unsaved dot next to the project name)? Did the problem start right after a change you can undo? Every edit you make to your data is one labeled entry in the **Changes** list, and you can undo it.

Installing and opening

"macOS says Statversity cannot be verified, or is damaged and should be moved to the Trash."

Why: Statversity is signed and checked by Apple's notarization service, so a fresh download from statversity.com only shows the ordinary "downloaded from the internet, Apple checked it for malicious software" question with an **Open** button. A "cannot be verified" or "damaged" message instead means the copy you have is old (from before the app was signed) or was altered on the way to you, for example by an email gateway or a cloud-sync tool.

Fix:

1. Click **Cancel** or **Done** on the warning (not **Move to Trash** if the app is already in Applications; you will replace it in a moment).
2. Download the installer again from **statversity.com**.
3. Open the new disk image and drag Statversity into **Applications**, choosing **Replace** if asked.
4. Open Statversity from **Applications**. Click **Open** on the "downloaded from the internet" question, and it starts.

"Windows says Windows protected your PC."

Why: The Windows installer is not yet signed with a Windows code-signing certificate, so SmartScreen warns about it the first time.

Fix: Click **More info**, then **Run anyway**.

“Statversity opens but never shows the engine ready card,” or “it says the engine had a problem.”

Why: The statistics engine (the program inside Statversity that does the calculations) did not start or stopped.

Fix:

1. Wait about ten seconds; the first launch is slower than later ones.
2. Quit Statversity with `Cmd` + `Q` and open it again.
3. If it happens again, open the log at `~/Library/Logs/edu.statversity.desktop/engine.log` (on Windows, `%LOCALAPPDATA%\edu.statversity.desktop\logs\engine.log`) and send it to whoever gave you Statversity.

“Statversity says a newer version is available.”

Why: Statversity checks statversity.com for the newest version when it opens, and yours is older. The Home screen shows **Statversity 1.4.0 is available.** (the number will match the new version), or the **Software updates** dialog says so.

Fix: Nothing is urgent. It is safe to keep working in the version you have, and your project files are not affected. When you are ready, click **See what’s new**, then **Download from statversity.com**. Download the new installer, quit Statversity, and drag the new copy over the old one in **Applications** (on Windows, run the new installer). Click **Not now** to hide the banner for that version. Chapter 2 has the full steps.

“Statversity couldn’t check for updates.”

Why: The **Software updates** dialog could not reach statversity.com. You may be offline, or a firewall on a university or work network may be blocking the request. Nothing is wrong with your data or your project.

Fix: Make sure you are connected to the internet and click **Try again**. If you are on a campus network that blocks the request, try again on another network, or check later by hand: in the **Statversity** menu choose **Check for Updates…** (on a Mac), or click **Check for updates…** at the bottom of the Home screen. You can also look for the newest installer on statversity.com in your web browser. Statversity works the same whether or not the check succeeds.

“I don’t want Statversity to check for updates.”

Why: By default, Statversity asks statversity.com which version is newest each time it opens. That is the only thing it asks the internet for, and it never sends your data or your project.

Fix: Open the **Software updates** dialog (the **Statversity** menu, then **Check for Updates…**, or **Check for updates…** on the Home screen) and untick **Check for updates automatically when Statversity opens**. Statversity then never checks unless you open the dialog yourself.

Importing files

"Statversity guessed the wrong kind of column" (an ID treated as text, a rating treated as a category, and so on).

Why: Statversity guesses from the values it sees, and short columns or unusual codes can fool it.

Fix:

1. On the **Check how we read them** step, read each guess. It tells you when labels, answer choices, or missing codes came from the file itself.
2. Use **Back** to change what you can, or finish the import and correct the column on the **Variables** tab.
3. The Variable Interview ("What is each question?") lets you set the role of every question, and you can reopen it later to change a single answer.

"My Qualtrics file shows weird extra rows at the top, or the column names look like questions."

Why: A Qualtrics export has up to three header rows: short names like Q4 in row 1, the full question wording in row 2, and a row of internal import codes in row 3 (older exports have only two).

Fix:

1. Choose the file exactly as Qualtrics exported it. Statversity detects the layout, uses row 1 as the column name and row 2 as the question text, and discards row 3.
2. Check the **Headers** line on **Check how we read them**.
3. If you opened the file in Excel and deleted or re-sorted rows before importing, export it from Qualtrics again and import the untouched file.

"The date, status, and IP address columns from Qualtrics are missing."

Why: Statversity hides Qualtrics' own system columns (StartDate, EndDate, Status, IPAddress, and similar) and removes them on the clean-up step by default.

Fix: On the clean-up step, untick **Survey system columns** if you want to keep them. Kept date columns can then be used, for example, to work out how long each response took.

“Rows are missing after import: I had 1,262 responses and Statversity shows fewer.”

Why: The clean-up step leaves out rows you did not mean to analyze, such as survey previews, spam, or unfinished responses, and shows how many it left out.

Fix: On the clean-up step, read each filter and its count. Untick the filter you do not want (for example **Leave out unfinished responses**). The review step shows the final count before you confirm.

“My Excel file only shows one sheet of data.”

Why: An Excel workbook can hold several sheets, and Statversity reads one at a time.

Fix: After you choose the `.xlsx` file, pick the sheet you want in the sheet picker. To use data from another sheet too, import it as an additional file and combine the two.

“The text looks garbled” (strange symbols in place of accented letters, or everything in one column).

Why: The file’s text encoding or separator was not what Statversity expected; Statversity auto-detects common cases, including UTF-8 with a byte-order mark and the UTF-16 files Qualtrics sometimes produces.

Fix: Look at the **Text encoding** line on **Check how we read them**. If it is still wrong, open the file in Excel, use **Save As** and choose **CSV UTF-8**, then import that copy.

“I have the data in Excel or Google Sheets and don’t want to save a file.”

Why: You do not need to save a file first.

Fix: Copy the table, then on **Choose files** click **Paste from clipboard**. It imports as “Pasted data”.

“It says there are repeated IDs.”

Why: The same participant ID appears on more than one row, which can mean a student submitted twice or that two students got the same ID.

Fix:

1. Click **Review** next to “N repeated IDs, M identical rows” (or click **Find duplicates** on the Data tab).
2. Click a group to see its rows highlighted in the grid and decide whether they are the same person.
3. For each group choose **Keep first** (delete the later rows), **Remove all**, or **Mark** (adds a Duplicate column, 0 = No and 1 = Yes, and deletes nothing).
4. Statversity never removes duplicates on its own. If you change your mind, undo it in **Changes**.

“My dates came in as text, or the day and month are swapped.”

Why: A column is read as dates only when nearly all its values (95 percent, and at least three) fit one date layout, and a date such as 3/4/2025 can mean March 4 or April 3.

Fix:

1. On **Check how we read them**, find the **Dates** box. If every date in the column is ambiguous, Statversity reads month/day/year and says so.
2. Use the **Month first** or **Day first** switch, or **Keep as text** if the column is not really dates.
3. For a column already imported, open the **Variables** tab and click **Read as dates…** next to it. The link appears only on text columns that look like dates.

“I see ‘uses the answer codes 1, 5, 6, which skip values’ on my rating questions.”

Why: Nobody picked the in-between choices in your data, so Statversity cannot be sure the scale is what you think it is.

Fix: This is a heads-up, not an error. Add your Qualtrics survey file (`.qsf`) on import so Statversity can read the real answer choices from it. Without one, read the note, confirm the codes are right, and continue.

“Statversity won’t let me continue after choosing files for a new time point or new responses.”

Why: Each file needs to be labeled with its time point, and **Continue** stays off until it is.

Fix: In the **Time point** box for each file, pick or type the label (Statversity proposes one from the file name, such as Pre or Post). When every file has nothing new, Continue is off and the screen says “Nothing to add”.

Setting up variables

“The interview has over a hundred steps.”

Why: Statversity asks about each question so that it can explain and score everything correctly.

Fix:

1. When you mark one question as the survey scale, accept the offer to tag every question with the same answer choices in one go.
2. Click any step in the left-hand list to jump straight to it, including steps you have not reached.
3. On the per-question steps, use **Jump to question…** to go straight to the one you want.
4. Choose **Skip for now** on anything you are unsure about. You can reopen the interview later.

“Statversity wants correct answers for questions that aren’t a test.”

Why: The question was guessed as a **Test question**, for example a yes/no behavior question.

Fix: On **What is each question?**, change the role to the right one (for example **Survey (Likert) item** or **Background / demographic**). The scoring step then stops asking for correct answers for it.

“Gender, age, or program doesn’t show up where I want it” (or “Q2 is a group when it should be a demographic”).

Why: Each question has one role, and facts about a person are normally filed as background, not as the groups being compared.

Fix: Set the question’s role to **Background / demographic** in the interview or on the **Variables** tab. If you want to compare groups on it, tick it under **Groups to compare**. Number columns such as age are not offered as groups, and typed “Other - Text” answers are not offered either.

“A hours-studied or GPA column has no role that fits.”

Why: It is neither a rating, a test, nor a demographic category.

Fix: Set the role to **Other score or measurement**, which is meant for any number you measured, such as hours studied or GPA. It can be used as an outcome.

“My alpha is negative” (or “the scale loads strangely”).

Why: Some items are worded the opposite way (“I find statistics boring” next to “I enjoy statistics”), so their scores need flipping.

Fix: In the scale step of the interview, tick the reverse-worded items (the **Worded the opposite way** option). Statversity then flips them for scale scores, Cronbach’s alpha, and factor analysis.

“A person’s scale or test score is blank.”

Why: Scores use only the questions each person answered, and a person who answered none of them gets a blank.

Fix: Open the Data tab and look at that person’s answers. If they skipped every item, a blank is correct. Statversity never counts an unanswered question as wrong.

“I typed a value into the grid and it was refused.”

Why: The column cannot hold that kind of value, for example text in a number column.

Fix: Read the short message under the cell, enter a valid value, or press `Esc` to cancel. Leaving a cell blank saves it as missing. Calculated variables and the time-point column cannot be typed over.

Running an analysis

“Statversity warns that an assumption is not met” (normality, equal variances, expected counts).

Why: A test has conditions that should hold for its result to be trustworthy, and Statversity checked yours.

Fix: Read the plain-language verdict, which says why it passed or failed. Statversity shows the plots next to the test, because with large samples a formal normality test flags tiny, harmless differences. You make the final choice: keep the test, or take the more cautious alternative Statversity suggests (for example a nonparametric test). Chapter 23 explains assumptions in detail.

“It says there is not enough data” or “no one has an answer.”

Why: Each analysis uses only people with answers on the variables involved, and paired or repeated tests need a complete set of answers from the same person.

Fix:

1. Open the **Data** tab and check the variables you chose for blanks. After import, Statversity shows a missing-data summary for each variable.
2. Check the roles: a group variable that nobody answered in one time point cannot split that time point.
3. Check that a filter is not hiding most of your rows.
4. Check that -99 or 999 codes are defined as missing codes. Otherwise they count as real numbers.

“Some groups weren’t tested,” or “This is 3 separate tests.”

Why: When you run an analysis separately for each time point or each group, a group with too few responses is skipped.

Fix: Read the banner and the small-groups warning on the result. They name which groups ran and which did not. To include a small group, combine it with another through a new calculated variable, or collect more responses.

“My result pooled Pre, Post, and Follow-up together.”

Why: The data holds several time points, and the analysis did not use the time-point variable.

Fix: Read the notice on the result. Either apply a filter (for example **Filter rows**, then Time point is Pre) or choose to run the analysis separately for each time point or compare across them.

“Chi-square says expected counts are too small.”

Why: Chi-square needs enough people in each cell of the table.

Fix: Statversity switches to Fisher’s exact test when counts are small. If you stayed on chi-square, use the one-click **Run Fisher’s exact test instead** button.

“I got a significant ANOVA but don’t know which groups differ.”

Why: ANOVA and Kruskal-Wallis only tell you that at least one group is different.

Fix: Run the follow-up (post-hoc) comparisons from the result. They produce a table of every pair of groups.

“Statversity says some runs are out of date.”

Why: The data or a saved filter changed after the runs happened. At the time of writing, Statversity can also flag runs after you only add new calculated variables, even though their own data did not change.

Fix: If you changed the data that a run uses, use **Run again** on it. If you only added a new column, you can leave the runs as they are.

Results and export

“The effect size is negative, but the text says Post is higher.”

Why: The sign depends on which group is subtracted from which.

Fix: Read the sentence under **What this means**, which states the direction in words. Report the direction with the words, not the sign alone.

“My table looks wrong when I paste it into Word.”

Why: Pasting plain text loses the formatting.

Fix: Use the **Copy** button on the APA table. It copies the table as formatted content (rich HTML) so Word keeps its formatting.

“My saved chart is not in the report.”

Why: A chart built in Chart Builder only appears in a report if it has been saved to the project and is ticked in the export box.

Fix:

1. In Chart Builder, click **Save chart**. Statversity warns you if you try to leave with an unsaved chart.

2. Open **Export**, choose the report, and tick your chart under **Project charts**.

"My exported SPSS file lost its missing-value codes."

Why: **Leave missing-value codes blank** was ticked, so a value such as 99 is written empty.

Fix: In the data export, untick **Leave missing-value codes blank** to keep up to three missing codes per variable in SPSS. Stata cannot hold missing codes, so they are always written blank there.

Projects and saving

"Statversity says Backup failed. Save your project soon."

Why: Statversity's automatic backup could not write the project. In the one case recorded so far, the app contained an out-of-date statistics engine that rejected the project data.

Fix:

1. Click **Save** and look at the error text in the header.
2. If Save also fails, use **Save As** to write a copy under a new name.
3. Install the latest Statversity build (see the installing chapter) and open the copy. If it keeps happening, send the `engine.log` described earlier.

"I closed Statversity and lost my work," or "Statversity offers to Recover unsaved work."

Why: Statversity closed before your last changes were saved, and an automatic backup is waiting.

Fix: At the next launch, click **Recover unsaved work**, then immediately save the project with **Save**. To avoid this, answer the quit prompt: when you quit with unsaved work, Statversity asks whether to save before quitting. Choose **Save** (or **Don't Save**, which also deletes the backup).

"I double-clicked a .stv file and Statversity opened empty (or did not open at all)."

Why: The Mac connects `.stv` files to Statversity after Statversity has been launched at least once. Older `.statly` files still open.

Fix: Open Statversity once from **Applications**, then double-click the project again. If Statversity is already open, it asks about unsaved changes first. Or open the file from inside Statversity's project menu.

"I changed something by mistake."

Why: Every edit to your data is recorded.

Fix: Open **Changes** and undo the entry (every row edit, added row, deleted rows, combine job, and duplicate mark is one labeled entry that can be undone and redone).

“Can I share my project with a classmate?”

Why: A `.stv` file holds your data, your variable settings, your results, and your Test Log in one file.

Fix: Send the single `.stv` file. Remember it contains the data, so if the data includes names, emails, or IP addresses, remove them first (Statversity offers to drop such columns on import) or get permission from your ethics board.

“I imported a new export and lost my variable setup.”

Why: Replace this project’s data rebuilds everything from the new file. The old data and all of its variable setup (types, labels, scale groups, survey flow, tags) are discarded, and this cannot be undone with **Changes**.

Fix: Next time, choose **Add new responses** in the **Import into which project?** dialog. It adds the new rows and keeps your setup. If you saved the project before the import, the saved `.stv` file still has the old setup: open it again without saving over it (use **Save As…** to keep both).