RESEARCH AUTOMATORS

Significance testing in Survey Automator

Survey Automator guide

Significance testing in Survey Automator

GUIDE REPORT

Survey Automator can flag statistically significant differences in banner and cross-tab tables, so you can see at a glance which results stand out rather than reflecting random sampling variation. This document explains what is tested, how, and how results are displayed.

What Gets Tested

Significance is calculated for two kinds of values in a table:

- **Proportions** — the percentage of respondents giving a particular answer.
- **Averages** — the mean of a numeric or scale value.

For each, Survey Automator compares one figure against a reference figure and asks: is this difference large enough to be unlikely to have arisen by chance?

Comparison Mode

You choose what each column is compared against:

MODE	WHAT EACH COLUMN IS COMPARED TO
Total *(default)*	The overall total (the whole sample)
Previous	The adjacent previous column

In **Total** mode, each breakdown group is tested against the full population — useful for spotting which subgroups differ from the average respondent. In **Previous** mode, columns are tested against each other in sequence — useful for tracking changes across waves or ordered groups.

How the Tests Work

All tests produce a *z-score*: a measure of how many standard errors a difference is away from zero. A larger absolute z-score means a more surprising difference. The z-score is then compared against a critical threshold tied to your chosen confidence level.

Comparing a subgroup against the total

Proportions. The subgroup's proportion is compared to the total proportion, with the standard error based on the subgroup's sample size.

Averages. The subgroup's mean is compared to the total's mean (a one-tailed test), using the total's standard deviation and the subgroup's sample size to form the standard error.

Comparing two groups against each other

Proportions. The two groups' proportions are compared using a pooled proportion (combining both groups to estimate the shared variance), with the standard error reflecting both sample sizes.

Averages. The two means are compared using the standard error built from each group's own standard deviation and sample size. This test only runs when **both groups have more than 30 respondents** — below that, no significance is calculated, because the test is not reliable on small samples.

Confidence Levels

You select a confidence level, which sets the threshold a z-score must clear to count as significant. Higher confidence demands a larger difference.

CONFIDENCE LEVEL	THRESHOLD (Z)
90%	1.282
95%	1.645
99%	2.326

These are one-tailed thresholds, so the test distinguishes "significantly higher" from "significantly lower" rather than just "different."

How Results Are Displayed

Each tested value gets one of three outcomes:

OUTCOME	MEANING	DEFAULT COLOR
▲ **Higher**	Significantly above the reference	Blue
▼ **Lower**	Significantly below the reference	Red
(none)	Not significant — no marker	—

The colors are themeable, so they can be matched to client branding. The same significance marker carries through consistently to every export format — both Excel and PowerPoint reports show the identical highlighting.

Column-to-column comparison (a/b/c letters)

In addition to the higher/lower arrows, columns can be compared pairwise, with each column labelled a letter (a, b, c, …). A letter placed on a value means *that value is significantly **higher** than the lettered column* — e.g. a "c" on column A reads "A is significantly higher than C." Like the arrows, this is a directional claim: it states which column is higher, not merely that two columns differ. Because every column is shown in the table, the reader can already see which value is larger; the letter adds the reliability claim — that the visible gap is trustworthy, not sampling noise.

Methodology notes

This section explains the reasoning behind two methodological choices.

Why the tests are one-tailed

The confidence thresholds above (1.282 / 1.645 / 2.326) are *one-tailed*. This is the correct choice **because every significance mark in the product is directional** — an up-arrow claims "higher," a down-arrow claims "lower," and an a/b/c letter claims "higher than that column." None of them claim the direction-agnostic "these differ, either way."

A one-tailed test is the right test for a one-directional claim. The mark commits to a direction, so the test only has to defend that one direction, and the whole error budget (5% at the 95% level) sits in that single tail. The everyday reading of the label and the technical reality therefore agree: when a user reads a 95% up-arrow as "about a 5% chance this is wrongly flagged as higher," that is essentially correct.

A common objection is that testing *both* directions should require a two-tailed threshold (which would be 1.645 / 1.960 / 2.576, producing fewer flags). That objection applies only to a *direction-agnostic* mark — one that means "different, unspecified direction." Survey Automator runs two separate one-directional checks ("is it higher?" and "is it lower?"), and any given cell can only trigger one of them. Each displayed flag is therefore a valid one-tailed result at its stated level. Testing both directions is not the same as any single flag claiming both directions.

When two-tailed would be correct: only if a mark were ever shown that means "differs from the reference, direction unspecified" (for example, a bare "significant" indicator with no arrow or letter). If such a mark is ever introduced, *that* mark should use two-tailed thresholds. As long as every significance mark carries a direction, one-tailed is correct throughout.

What the "vs Total" comparison actually answers

In **Total** mode, each subgroup is compared against the overall total. It is worth being precise about what this asks, because there are two readings and they are not the same:

- *"Does this group differ from the reported Total?"* — The Total is the report's headline benchmark and includes the group itself. The test answers this question directly.
- *"Is this group independently distinctive from everyone else?"* — this is a subtly different question. Because the subgroup is *part of* the total it is compared against, the two figures are not statistically independent, which makes differences look marginally more significant than a comparison against everyone *outside* the group would.

The current test answers the first question — comparison against the visible total, which has the advantage that the arithmetic matches what is on screen (the test compares against the same Total number the reader can see). It does *not* claim to answer the second. A reader who interprets a vs-Total flag as "this group stands apart from all others" is drawing a slightly stronger conclusion than the test supports. The honest framing is that a vs-Total flag means "differs from the overall benchmark," and it is a lenient, exploratory comparison — a starting point, not an independent-difference verdict.

Multiple comparisons in a/b/c tables

Pairwise column comparison runs many more tests than the arrows. Comparing each column against a single reference (Total or Previous) is a handful of tests per row; comparing every column against every other is many more — a 5-column table runs 10 pairwise comparisons per row. Each individual letter is a correct one-tailed test at its stated level, but because so many tests run across a table, some letters will appear by chance alone. This is the standard multiple-comparisons effect and is inherent to a/b/c notation across all tools that offer it; it is independent of the one- vs two-tailed choice. The honest reading is that the letters are *not* family-wise corrected, so a table full of letters should be read as the sum of many independent tests, not as one collectively-certain picture.

Correlation Tables

Correlation tables use a different, dedicated significance test (a t-test) suited to correlation coefficients. This is separate from the proportion and average testing used in standard banner and cross-tab tables.

The latest version of this guide: https://www.researchautomators.com/guides/significance-testing/