

How to Perform a One Sample t-test in Stata to Compare a Sample Mean to a Known Value

Authored by
stats writer

March 8, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform a One Sample t-test in Stata to Compare a Sample Mean to a Known Value*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134520>

Understanding the Core Concepts of the One-Sample t-test

The **One-Sample t-test** is a fundamental tool in the field of inferential statistics, designed to determine whether the average value of a single sample significantly differs from a hypothesized or known **population mean**. In practical research, we rarely have access to entire populations, so we rely on samples to draw broader conclusions. By utilizing a **One-Sample t-test**, researchers can evaluate if a specific group--such as a cohort of patients, a batch of manufactured goods, or a sample of automobiles--deviates from a standard benchmark or a historical average. This method is indispensable for validating experimental results and ensuring that observed differences are not merely the result of random chance.

Within the environment of **Stata**, this procedure is streamlined to allow for both rapid calculation and deep analytical exploration. The utility of the test extends across various disciplines, including healthcare, where it might be used to compare a sample's blood pressure against a known healthy average, or in psychology, to compare test scores against a national mean. By establishing a **null hypothesis**--which typically posits that there is no difference between the sample mean and the population mean--the test provides a mathematical framework to either support or refute this claim based on the collected data.

To execute this analysis effectively, one must understand the relationship between the **sample mean**, the **standard deviation**, and the **standard error**. The test calculates a **t-statistic**, which represents the number of standard errors the sample mean is away from the hypothesized mean. A larger t-statistic, paired with a small **p-value**, typically indicates that the difference is statistically significant. **Stata** provides a comprehensive output that includes these metrics, alongside **confidence intervals**, to offer a complete picture of the data's reliability and the strength of the findings.

Statistical Assumptions for a Valid One-Sample t-test

Before performing a **One-Sample t-test** in **Stata**, it is crucial to ensure that the data meets specific **statistical assumptions**. The most critical requirement is that the data must be **continuous**, meaning it is measured on an interval or ratio scale. For instance, in our example regarding fuel efficiency, miles per gallon (mpg) is a continuous variable. Additionally, the observations in the sample must be **independent** of one another, meaning the value of one car's mpg does not influence or depend on the value of another car in the dataset.

Another vital assumption is that the data should follow a **normal distribution**, especially when the sample size is small. While the **Central Limit Theorem** suggests that larger samples (typically $n > 30$) tend to result in a sampling distribution of the mean that is normal regardless of the population's shape, checking for **outliers** and skewness remains a best practice. Significant

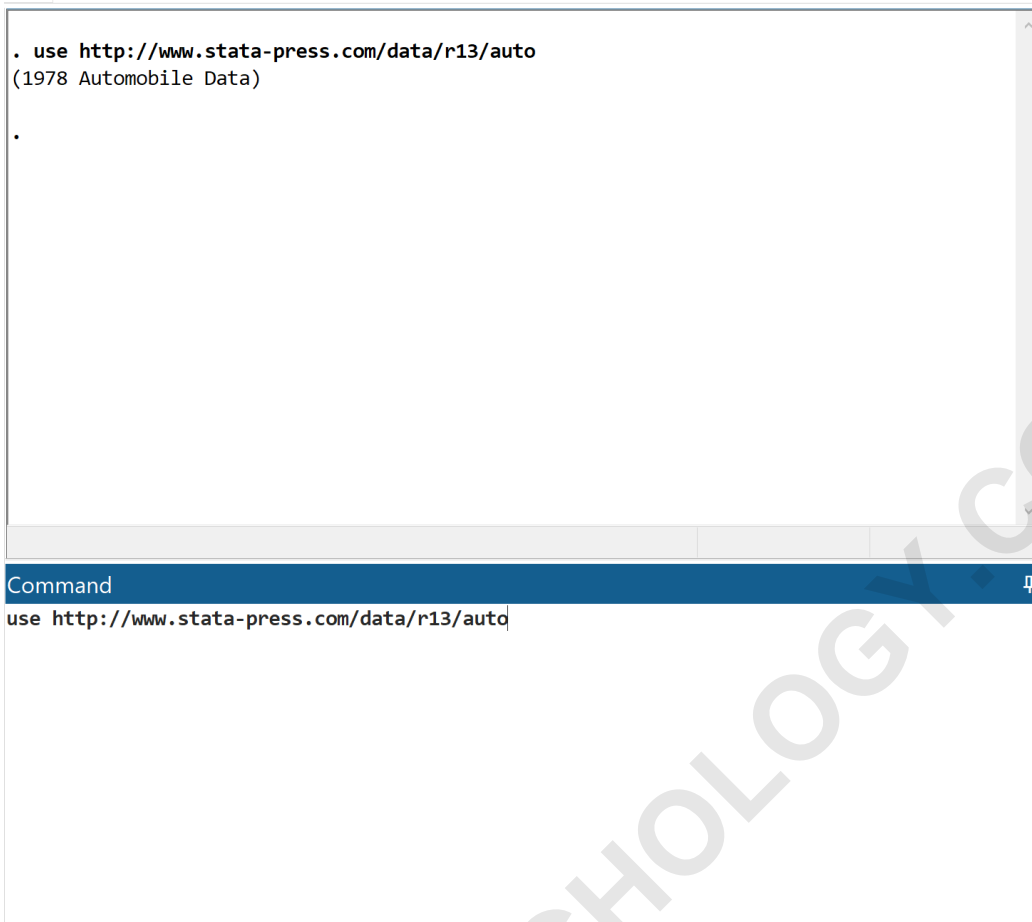
outliers can disproportionately affect the **mean** and **standard deviation**, potentially leading to misleading t-values and inaccurate conclusions regarding the **null hypothesis**.

In the context of **Stata**, users can verify these assumptions through various diagnostic tools. Histograms, Q-Q plots, and the Shapiro-Wilk test are frequently used to assess normality. If the data severely violates these assumptions, researchers might consider alternative non-parametric tests, such as the Wilcoxon signed-rank test. However, for most robust datasets in social sciences and business analytics, the **One-Sample t-test** remains the gold standard for comparing a single mean against a constant value, provided these baseline conditions are reasonably satisfied.

Step 1: Loading and Preparing the Dataset in Stata

The first practical step in conducting a **One-Sample t-test** is to import or load your data into the **Stata** workspace. For the purposes of this tutorial, we will utilize a built-in dataset commonly used for teaching statistical methods. By entering a specific command, you can pull data directly from the **Stata** servers, ensuring that you have a clean, structured environment to begin your analysis. This dataset contains various attributes of automobiles from 1978, but our primary focus will be on the fuel efficiency variable.

To load the data, type the following command into the **Stata** command box and press Enter: **use <http://www.stata-press.com/data/r13/auto>**. This action populates your software with 74 observations across several variables. It is essential to verify that the data has loaded correctly before proceeding to any mathematical calculations. Loading data via the command line is often preferred for reproducibility, as it allows other researchers to replicate your exact workflow by simply running the same script.



```
. use http://www.stata-press.com/data/r13/auto
(1978 Automobile Data)

.
```

Command

```
use http://www.stata-press.com/data/r13/auto
```

Once the data is active, you should see the variable list update on the right-hand side of the **Stata** interface. The variable "mpg" represents the miles per gallon for each car, which is the **dependent variable** we will use for our **One-Sample t-test**. Understanding your data structure at this stage prevents errors later, such as attempting to run a t-test on a categorical variable or a string of text, which would result in a syntax error or meaningless output.

Step 2: Inspecting Raw Data for Accuracy

Before jumping into the **statistical analysis**, it is a professional best practice to "get to know" your data. This involves a visual inspection of the raw values to ensure there are no obvious data entry errors or missing values that could skew the results of the **One-Sample t-test**. In **Stata**, this is most easily accomplished using the **Data Editor**. By viewing the spreadsheet-like view of the data, you can confirm that the "mpg" column contains numeric values and that the sample size matches your expectations.

To view the data, navigate to the top menu bar and select **Data > Data Editor > Data Editor (Browse)**. This opens a read-only window where you can scroll through the 74 records. While

there are many variables present, such as price, weight, and length, our focus remains strictly on the **arithmetic mean** of the "mpg" column. Seeing the data in this format helps conceptualize the variability within the sample before the software condenses these 74 points into a single **sample mean**.

	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9	Buick Riviera	10,372	16	3	3.5	17	3,880	207	43	231	2.93	Domestic
10	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic
11	Cad. Deville	11,385	14	3	4.0	20	4,330	221	44	425	2.28	Domestic
12	Cad. Eldorado	14,500	14	2	3.5	16	3,900	204	43	350	2.19	Domestic
13	Cad. Seville	15,906	21	3	3.0	13	4,290	204	45	350	2.24	Domestic
14	Chev. Chevette	3,299	29	3	2.5	9	2,110	163	34	231	2.93	Domestic
15	Chev. Impala	5,705	16	4	4.0	20	3,690	212	43	250	2.56	Domestic
16	Chev. Malibu	4,504	22	3	3.5	17	3,180	193	31	200	2.73	Domestic
17	Chev. Monte Carlo	5,104	22	2	2.0	16	3,220	200	41	200	2.73	Domestic
18	Chev. Monza	3,667	24	2	2.0	7	2,750	179	40	151	2.73	Domestic
19	Chev. Nova	3,955	19	3	3.5	13	3,430	197	43	250	2.56	Domestic
20	Dodge Colt	3,984	30	5	2.0	8	2,120	163	35	98	3.54	Domestic
21	Dodge Diplomat	4,010	18	2	4.0	17	3,600	206	46	318	2.47	Domestic
22	Dodge Magnum	5,886	16	2	4.0	17	3,600	206	46	318	2.47	Domestic
23	Dodge St. Regis	6,342	17	2	4.5	21	3,740	220	46	225	2.94	Domestic
24	Ford Fiesta	4,389	28	4	1.5	9	1,800	147	33	98	3.15	Domestic
25	Ford Mustang	4,187	21	3	2.0	10	2,650	179	43	140	3.08	Domestic
26	Linc. Continental	11,497	12	3	3.5	22	4,840	233	51	400	2.47	Domestic
27	Linc. Mark V	13,594	12	3	2.5	18	4,720	230	48	400	2.47	Domestic
28	Linc. Versailles	13,466	14	3	3.5	15	3,830	201	41	302	2.47	Domestic
29	Merc. Bobcat	3,829	22	4	3.0	9	2,580	169	39	140	2.73	Domestic
30	Merc. Cougar	5,379	14	4	3.5	16	4,060	221	48	302	2.75	Domestic
31	Merc. Marquis	6,165	15	3	3.5	23	3,720	212	44	302	2.26	Domestic
32	Merc. Monarch	4,516	18	3	3.0	15	3,370	198	41	250	2.43	Domestic
33	Merc. XR-7	6,303	14	4	3.0	16	4,130	217	45	302	2.75	Domestic

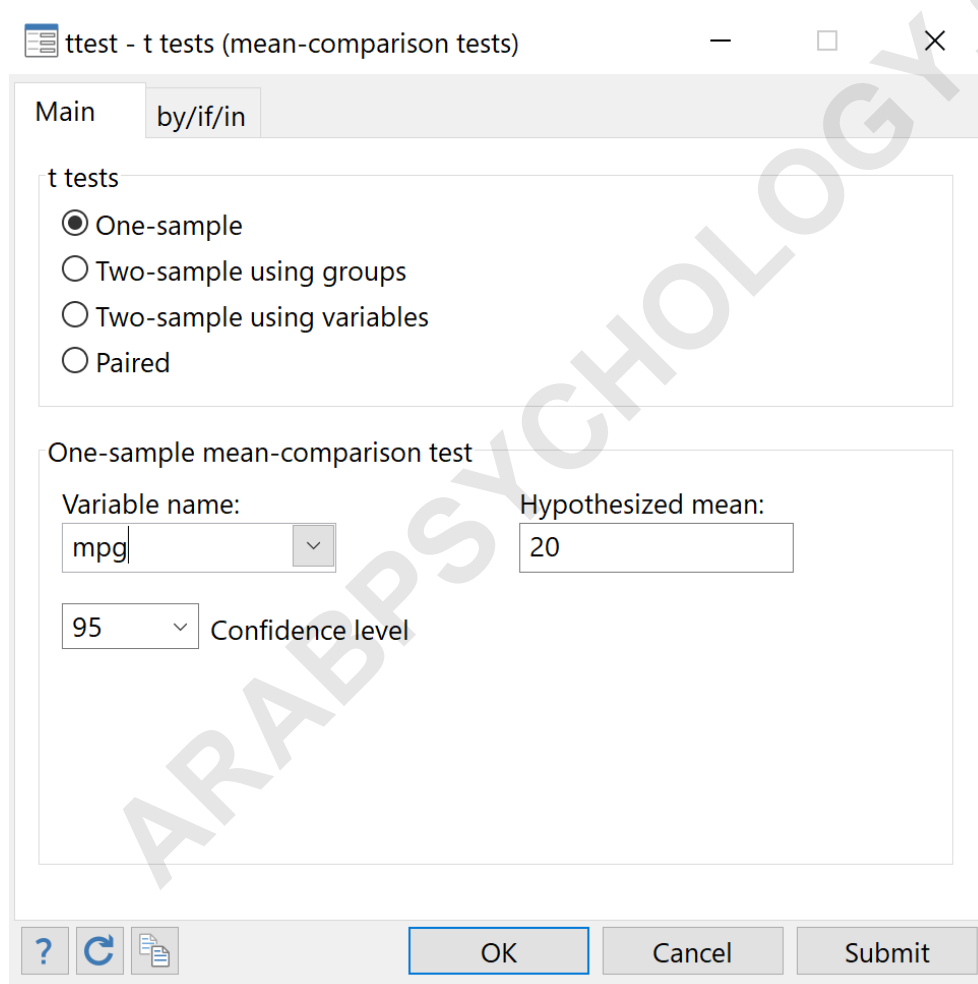
Visualizing the data also allows you to identify potential **outliers**. For instance, if one car was recorded as having 500 mpg, it would likely be a typo that would significantly inflate the **mean** and lead to a Type I error. By confirming that the values generally range within a realistic envelope for 1970s automobiles, you increase the **validity** of your subsequent findings. Once you are satisfied with the integrity of the raw data, you can close the browser and proceed to the execution of the test.

Step 3: Executing the One-Sample t-test via the GUI

The most user-friendly way to perform a **One-Sample t-test** in **Stata** is through the **Graphical User Interface (GUI)**. This method is particularly helpful for those who are new to the software or

prefer a visual workflow over typing commands. To begin, go to the top menu and follow this path: **Statistics > Summaries, tables, and tests > Classical tests of hypotheses > t test (mean-comparison test)**. This will open a dialog box where you can specify the parameters of your hypothesis.

In the dialog box, ensure that the **"One-sample"** radio button is selected. This tells **Stata** that you are comparing a single sample against a constant, rather than comparing two different groups. From the dropdown menu for **Variable name**, select "mpg". In the field labeled **Hypothesized mean**, input the value "20". This value represents our **null hypothesis**--the baseline we are testing against. We are essentially asking: "Is the average fuel efficiency of these cars significantly different from 20 miles per gallon?"



The image shows a screenshot of the Stata dialog box titled "ttest - t tests (mean-comparison tests)". The dialog has two tabs: "Main" and "by/if/in". Under "t tests", the "One-sample" radio button is selected. Below this, the "One-sample mean-comparison test" section contains a "Variable name:" dropdown menu with "mpg" selected, a "Hypothesized mean:" text input field with "20" entered, and a "Confidence level:" dropdown menu with "95" selected. At the bottom of the dialog are buttons for "?", a refresh icon, a save icon, "OK", "Cancel", and "Submit". A large watermark "ARABPSYCHOLOGY.COM" is overlaid diagonally across the dialog box.

Finally, you can adjust the **Confidence level**. The default is usually set to 95, which corresponds to a **significance level** (alpha) of 0.05. This is the standard threshold in most scientific research. After ensuring all fields are correctly filled, click **OK**. **Stata** will then process the request and display the results in the main Output window. This automated process handles all the complex **calculus** and **probability** distributions behind the scenes, providing you with a clean table of results.

This result implies that the observed **sample mean** of 21.2973 is not "different enough" from 20 to rule out random sampling variation. It is a common misconception that a non-significant result means the means are "equal." Rather, it technically means we **fail to reject** the claim that they are 20. The **statistical significance** just barely missed the 0.05 threshold, which might prompt a researcher to suggest further study with a larger sample size to increase **statistical power**.

Understanding the **two-tailed test** is vital here. We were testing if the mpg was simply "not 20," meaning it could be significantly higher or significantly lower. If our hypothesis had been directional--for example, if we only cared if the cars got *better* than 20 mpg--we would have looked at the one-tailed p-value on the right (0.0288), which *would* have been significant at the 0.05 level. Choosing the correct tail for your **hypothesis testing** is a critical step that must be decided before looking at the data to maintain **academic integrity**.

Step 6: Formal Reporting of the t-test Results

The final stage of the process is communicating your findings in a clear, professional, and standardized format. A good report should include the **t-statistic**, the **degrees of freedom**, the **p-value**, and the **confidence interval**. This allows readers to understand not just whether the result was significant, but also the **effect size** and the precision of your estimate. In academic writing, this is often formatted according to specific style guides like APA or Vancouver.

Based on our **Stata** analysis, a formal summary would look like this: "A **One-Sample t-test** was conducted to compare the fuel efficiency of 74 automobiles to a hypothesized **population mean** of 20 mpg. The results indicated that the mean fuel efficiency ($M = 21.30$, $SD = 5.79$) was not significantly different from the hypothesized value of 20 mpg, $t(73) = 1.93$, $p = .058$. The **95% confidence interval** for the true population mean ranged from 19.96 to 22.64."

This detailed reporting ensures that your **statistical analysis** is transparent. By including the **confidence interval**, you provide a range of plausible values for the mean, which is often more informative than a simple "yes/no" decision regarding **significance**. Whether you are writing a research paper, a business report, or a clinical study, following this structured approach to reporting ensures your **Stata** results are interpreted correctly by your audience, leading to better-informed decisions and more reliable scientific conclusions.

Summary of Key Terms and Procedures

To conclude this guide on performing a **One-Sample t-test** in **Stata**, let us review the essential components of the workflow. The process begins with clear data preparation and ends with nuanced interpretation. Here is a summary of the critical steps and terms you should master:

Null Hypothesis (H0): The assumption that the **sample mean** is equal to the **population mean**.

Alternative Hypothesis (Ha): The assumption that the **sample mean** is different from (or greater/less than) the population mean.

Standard Deviation: A measure of how much the individual mpg values vary from the average.

t-statistic: The ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error.

Stata Commands: While we used the GUI, the command line equivalent is **ttest mpg == 20**.

Mastering these concepts allows you to leverage **Stata** as a powerful ally in your research. While this tutorial focused on fuel efficiency, the exact same steps apply whether you are measuring the efficacy of a new drug, the accuracy of a manufacturing process, or the average income in a specific region. By consistently applying these methods, you ensure your **quantitative research** is grounded in rigorous **statistical methodology**.

Finally, always remember that **statistical significance** is not the same as practical significance. Even if a result is statistically significant, you must ask yourself if the difference is large enough to matter in the real world. **Stata** provides the numbers, but as the researcher, you provide the context and the **critical thinking** necessary to turn data into meaningful knowledge.