Chapter 205

One-Sample T-Test

Introduction

This procedure provides several reports for making inference about a population mean based on a single sample. These reports include confidence intervals of the mean or median, the t-test, the z-test, and non-parametric tests including the randomization test, the quantile (sign) test, and the Wilcoxon Signed-Rank test. Tests of assumptions and distribution plots are also available in this procedure.

Another procedure that produces a large amount of summary information about a single sample is the Descriptive Statistics procedure. While it is not as focused on hypothesis testing, it contains many additional descriptive statistics, including minimum, maximum, range, counts, trimmed means, sums, mode, variance, Skewness, Kurtosis, coefficient of variation, coefficient of dispersion, percentiles, additional normality tests, and a stem-and-leaf plot.

Research Questions

For the one-sample situation, the typical concern in research is examining a measure of central tendency (location) for the population of interest. The best-known measures of location are the mean and median. For a one-sample situation, we might want to know if the average waiting time in a doctor's office is greater than one hour, if the average refund on a 1040 tax return is different from $500, if the average assessment for similar residential properties is less than $120,000, or if the average growth of roses is 4 inches or more after two weeks of treatment with a certain fertilizer.

One early concern should be whether the data are normally distributed. If normality can safely be assumed, then the one-sample t-test is the best choice for assessing whether the measure of central tendency, the mean, is different from a hypothesized value. On the other hand, if normality is not valid, one of the nonparametric tests, such as the Wilcoxon Signed Rank test or the quantile test, can be applied.

Technical Details

The technical details and formulas for the methods of this procedure are presented in line with the Example 1 output. The output and technical details are presented in the following order:

- Descriptive Statistics
- Confidence Interval for μ with σ Unknown
- Confidence Interval for μ with σ Known
- Confidence Interval for the Median
- Bootstrap Confidence Intervals
- Confidence Interval of σ
- T-Test and associated power report

- Z-Test

- Randomization Test

- Quantile (Sign) Test

- Wilcoxon Signed-Rank Test

- Tests of Assumptions

- Graphs

Null and Alternative Hypotheses

The basic null hypothesis is that the population mean is equal to a hypothesized value,

$$H_0: \mu = Hypothesized\ Value$$

with three common alternative hypotheses,

$$H_a: \mu \neq Hypothesized\ Value\ ,$$

$$H_a: \mu < Hypothesized\ Value\ ,\ or$$

$$H_a: \mu > Hypothesized\ Value\ ,$$

one of which is chosen according to the nature of the experiment or study.

Assumptions

This section describes the assumptions that are made when you use one of these tests. The key assumption relates to normality or nonnormality of the data. One of the reasons for the popularity of the t-test is its robustness in the face of assumption violation. However, if an assumption is not met even approximately, the significance levels and the power of the t-test are invalidated. Unfortunately, in practice it often happens that more than one assumption is not met. Hence, take the steps to check the assumptions before you make important decisions based on these tests. There are reports in this procedure that permit you to examine the assumptions, both visually and through assumptions tests.

One-Sample T-Test Assumptions

The assumptions of the one-sample t-test are:

1. The data are continuous (not discrete).

2. The data follow the normal probability distribution.

3. The sample is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Wilcoxon Signed-Rank Test Assumptions

The assumptions of the Wilcoxon signed-rank test are as follows:

1. The data are continuous (not discrete).

2. The distribution of the data is symmetric.

3. The data are mutually independent.

4. The data all have the same median.

5. The measurement scale is at least interval.

Quantile Test Assumptions

The assumptions of the quantile (sign) test are:

1. A random sample has been taken resulting in observations that are independent and identically distributed.

2. The measurement scale is at least ordinal.

Data Structure

For this procedure, the data are entered as a single column and specified as a response variable. Multiple columns can be analyzed individually during the same run if multiple response variables are specified.

Weight Dataset

Weight
159
155
157
125
103
122
101
82
228
199
195
110
191
151
119
119
112
87
190
87

Example 1 – Running a One-Sample Analysis

This section presents an example of how to run a one-sample analysis. The data are the Weight data shown above and found in the **Weight** dataset. The data can be found under the column labeled *Weight*. A researcher wishes to test whether the mean weight of the population of interest is different from 130.

Setup

To run this example, complete the following steps:

1 Open the Weight example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Weight** and click **OK**.

2 Specify the One-Sample T-Test procedure options

- Find and open the **One-Sample T-Test** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable(s) ....**Weight**

Reports Tab

Descriptive Statistics......................................**Checked**
Confidence Level **95**
Limits ....**Two-Sided**
Confidence Interval of μ with σ Unknown ....**Checked**
Confidence Interval of μ with σ Known ....**Checked**
  σ **40**
Confidence Interval of the Median ....**Checked**
Bootstrap Confidence Intervals........................**Checked**
  Sub-Options..**Default Values**
Confidence Interval of σ..................................**Checked**
Alpha **0.05**
H0 μ =..**130**
Ha ....**Two-Sided and One-Sided**
                                                              (Usually, a single alternative hypothesis is chosen, but all three
                                                               alternatives are shown in this example to exhibit all the reporting
                                                               options.)
T-Test ....**Checked**
Power Report for T-Test ....**Checked**
Z-Test ....**Checked**
  σ **40**
Randomization Test..**Checked**
  Monte Carlo Samples **10000**

Quantile (Sign) Test..**Checked**
  Quantile Test Proportion...............................**0.5**
Wilcoxon Signed-Rank Test.............................**Checked**
  Sub-Options..**Default Values**
Tests of Assumptions ....**Checked**
  Assumptions Alpha.......................................**0.05**
Random Seed...**5213861** (for reproducibility)

Plots Tab

All Plots...**Checked**

3   Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Descriptive Statistics

Descriptive Statistics

Variable	N	Mean	Standard Deviation	Standard Error	Median
Weight	20	139.6	43.1221	9.642395	123.5

Variable

The name of the variable for which the descriptive statistics are listed here.

N

This is the number of non-missing values.

Mean

This is the average of the data values.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Standard Deviation

The sample standard deviation is the square root of the variance. It is a measure of dispersion based on squared distances from the mean for the variables listed.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

Standard Error

This is the estimated standard deviation of the distribution of sample means for an infinite population.

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}}$$

Median

The median is the 50th percentile of the data, using the AveXp(n+1) method. The details of this method are described in the Descriptive Statistics chapter under Percentile Type.

Two-Sided Confidence Interval for the Mean (μ) with the Standard Deviation (σ) Unknown

Two-Sided Confidence Interval for the Mean (μ) with the Standard Deviation (σ) Unknown

Variable	N	Mean	Standard Deviation	Standard Error	T*	DF	95% Confidence Interval Limits for the Mean (μ)	
							Lower	Upper
Weight	20	139.6	43.1221	9.642395	2.0930	19	119.4182	159.7818

Variable

This is the spreadsheet column variable.

N

This is the number of non-missing values.

Mean

This is the average of the data values.

Standard Deviation

The sample deviation is the square root of the variance. It is a measure of dispersion based on squared distances from the mean for the variables listed.

Standard Error

This is the estimated standard deviation of the distribution of sample means for an infinite population.

T*

This is the t-value used to construct the confidence limits. It is based on the degrees of freedom and the confidence level.

DF

The degrees of freedom are used to determine the T distribution from which T* is generated:

$$df = n - 1$$

95% Confidence Interval Limits for the Mean (μ) (Lower and Upper)

These are the confidence limits of the confidence interval for μ. The confidence interval formula is

$$\overline{x} \pm T_{df}^* SE_{\overline{x}}$$

Two-Sided Confidence Interval for the Mean (μ) with the Standard Deviation (σ) Known

Two-Sided Confidence Interval for the Mean (μ) with the Standard Deviation (σ) Known

Variable	N	Mean	σ	Standard Error	Z*	95% Confidence Interval Limits for the Mean (μ)	
						Lower	Upper
Weight	20	139.6	40	8.944272	1.9600	122.0695	157.1304

Variable

This is the spreadsheet column variable.

N

This is the number of non-missing values.

Mean

This is the average of the data values.

σ (Known Standard Deviation)

This is the known pre-specified value for the population standard deviation.

Standard Error

This is the standard deviation of the distribution of sample means.

$$SE_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Z*

This is the z-value used to construct the confidence limits. It is based on the standard normal distribution according to the specified confidence level.

95% Confidence Interval Limits for the Mean (μ) (Lower and Upper)

These are the confidence limits of the confidence interval for μ. The confidence interval formula is

$$\overline{x} \pm Z^* SE_{\overline{x}}$$

Two-Sided Confidence Interval for the Median

Two-Sided Confidence Interval for the Median

Variable	N	Median	95% Confidence Interval Limits for the Median	
			Lower	Upper
Weight	20	123.5	110	159

Variable

This is the spreadsheet column variable.

N

This is the number of non-missing values.

Median

The median is the 50th percentile of the data, using the AveXp(n+1) method. The details of this method are described in the Descriptive Statistics chapter under Percentile Type.

95% Confidence Interval Limits for the Median (Lower and Upper)

These are the lower and upper confidence limits of the median. These limits are exact and make no distributional assumptions other than a continuous distribution. No limits are reported if the algorithm for this interval is not able to find a solution. This may occur if the number of unique values is small.

Bootstrap Confidence Intervals for the Mean

Bootstrapping was developed to provide standard errors and confidence intervals in situations in which the standard assumptions are not valid. In these nonstandard situations, bootstrapping is a viable alternative to the corrective action suggested earlier. The method is simple in concept, but it requires extensive computation time.

The bootstrap is simple to describe. You assume that your sample is actually the population, and you draw *B* samples (*B* is over 1000) of N from the original dataset, with replacement. *With replacement* means that each observation may be selected more than once. For each bootstrap sample, the mean is computed and stored.

Suppose that you want the standard error and a confidence interval of the mean. The bootstrap sampling process has provided *B* estimates of the mean. The standard deviation of these *B* means is the bootstrap estimate of the standard error of the mean. The bootstrap confidence interval is found by arranging the *B* values in sorted order and selecting the appropriate percentiles from the list. For example, a 90% bootstrap

confidence interval for the difference is given by fifth and ninety-fifth percentiles of the bootstrap mean values.

The main assumption made when using the bootstrap method is that your sample approximates the population fairly well. Because of this assumption, bootstrapping does not work well for small samples in which there is little likelihood that the sample is representative of the population. Bootstrapping should only be used in medium to large samples.

Bootstrap Confidence Intervals for the Mean
_____

Estimation Results			Bootstrap Confidence Interval Limits		
Parameter	**Estimate**	**|**	**Confidence Level**	**Lower**	**Upper**
Mean					
Original Value	139.6	|	90%	124.6	154.75
Bootstrap Mean	139.4281	|	95%	121.4013	157.0988
Bias (BM - OV)	-0.1718833	|	99%	115.2028	162.2995
Bias Corrected	139.7719				
Standard Error	9.246735				

_____

Confidence Limit Type = Reflection, Number of Samples = 3000.
User-Entered Random Seed = 5213861

Bootstrap Histograms Section
_____
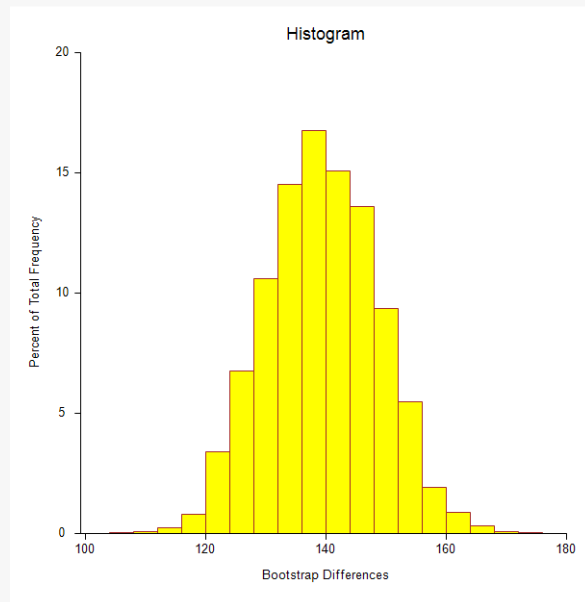
This report provides bootstrap confidence intervals of the mean. Note that since these results are based on 3000 random bootstrap samples, they will differ slightly from the results you obtain when you run this report.

Original Value

This is the parameter estimate obtained from the complete sample without bootstrapping.

Bootstrap Mean

This is the average of the parameter estimates of the bootstrap samples.

Bias (BM - OV)

This is an estimate of the bias in the original estimate. It is computed by subtracting the original value from the bootstrap mean.

Bias Corrected

This is an estimated of the parameter that has been corrected for its bias. The correction is made by subtracting the estimated bias from the original parameter estimate.

Standard Error

This is the bootstrap method's estimate of the standard error of the parameter estimate. It is simply the standard deviation of the parameter estimate computed from the bootstrap estimates.

Confidence Level

This is the confidence coefficient of the bootstrap confidence interval given to the right.

Bootstrap Confidence Interval Limits (Lower and Upper)

These are the limits of the bootstrap confidence interval with the confidence coefficient given to the left. These limits are computed using the confidence interval method (percentile or reflection) designated on the Bootstrap panel.

Note that to be accurate, these intervals must be based on over a thousand bootstrap samples and the original sample must be representative of the population.

Bootstrap Histogram

The histogram shows the distribution of the bootstrap parameter estimates.

Two-Sided Confidence Interval for the Standard Deviation (σ)

Two-Sided Confidence Interval for the Standard Deviation (σ)

Variable	N	Standard Deviation	95% Confidence Interval Limits for the Standard Deviation (σ)	
			Lower	Upper
Weight	20	43.1221	32.79395	62.98292

Variable

This is the spreadsheet column variable.

N

This is the number of non-missing values.

Standard Deviation

The sample deviation is the square root of the variance. It is a measure of dispersion based on squared distances from the mean for the variables listed.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

95% Confidence Interval Limits for the Standard Deviation (σ) (Lower and Upper)

This report gives the limits of the confidence interval for the population standard deviation. The validity of these intervals is very dependent on the assumption that the data are sampled from a normal distribution.

The formula for the confidence interval of the standard deviation is given by

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}$$

where α is one minus the confidence level.

One-Sample T-Test

This section presents the results of the traditional one-sample T-test. Here, reports for all three alternative hypotheses are shown, but a researcher would typically choose one of the three before generating the output. All three tests are shown here for the purpose of exhibiting all the output options available.

One-Sample T-Test

Alternative Hypothesis	Mean	Standard Error	T-Statistic	DF	P-Value	Reject H0 at α = 0.05?
μ ≠ 130	139.6	9.642395	0.9956	19	0.33195	No
μ < 130	139.6	9.642395	0.9956	19	0.83402	No
μ > 130	139.6	9.642395	0.9956	19	0.16598	No

Alternative Hypothesis

The (unreported) null hypothesis is

$$H_0: \mu = Hypothesized\ Value$$

and the alternative hypotheses,

$$H_a: \mu \neq Hypothesized\ Value,$$

$$H_a: \mu < Hypothesized\ Value,\ \text{or}$$

$$H_a: \mu > Hypothesized\ Value.$$

In practice, the alternative hypothesis should be chosen in advance.

Mean

This is the average of the data values.

Standard Error

This is the estimated standard deviation of the distribution of sample means for an infinite population.

$$SE_{\overline{x}} = \frac{s}{\sqrt{n}}$$

T-Statistic

The T-Statistic is the value used to produce the *p*-value (Prob Level) based on the *T* distribution. The formula for the T-Statistic is:

$$T - Statistic = \frac{\overline{x} - Hypothesized\ Value}{SE_{\overline{x}}}$$

DF

The degrees of freedom define the T *distribution* upon which the probability values are based. The formula for the degrees of freedom is:

$$df = n - 1$$

P-Value

The *p*-value, also known as the probability level or significance level, is the probability that the test statistic will take a value at least as extreme as the observed value, assuming that the null hypothesis is true. If the *p*-value is less than the prescribed α, in this case 0.05, the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, there is not sufficient evidence to reject the null hypothesis.

Reject H0 at α = 0.05?

This column indicates whether or not the null hypothesis is rejected, in favor of the alternative hypothesis, based on the *p*-value and chosen α. A test in which the null hypothesis is rejected is sometimes called *significant*.

Power for the One-Sample T-Test

The power report gives the power of a test where it is assumed that the population mean and standard deviation are equal to the sample mean and standard deviation. Powers are given for alpha values of 0.05 and 0.01. For a much more comprehensive and flexible investigation of power or sample size, we recommend you use the **PASS** software program.

Power for the One-Sample T-Test

Alternative Hypothesis	N	μ	σ	Power (α = 0.05)	Power (α = 0.01)
μ ≠ 130	20	139.6	43.1221	0.15711	0.04833
μ < 130	20	139.6	43.1221	0.00458	0.00057
μ > 130	20	139.6	43.1221	0.24681	0.08081

This section assumes that the population mean (μ) and standard deviation (σ) are equal to the sample values.

Alternative Hypothesis

This value identifies the test direction of the test reported in this row. In practice, you would select the alternative hypothesis prior to your analysis and have only one row showing here.

N

N is the assumed sample size.

μ

This is the assumed population mean on which the power calculation is based.

σ

This is the assumed population standard deviation on which the power calculation is based.

Power (α = 0.05) and Power (α = 0.01)

Power is the probability of rejecting the hypothesis that the mean is equal to the hypothesized value when they are in fact not equal. Power is one minus the probability of a type II error (β). The power of the test depends on the sample size, the magnitude of the standard deviation, the alpha level, and the true difference between the population mean and the hypothesized value.

The power value calculated here assumes that the population standard deviation is equal to the sample standard deviation and that the difference between the population mean and the hypothesized value is exactly equal to the difference between the sample mean and the hypothesized value.

High power is desirable. High power means that there is a high probability of rejecting the null hypothesis when the null hypothesis is false.

Some ways to increase the power of a test include the following:

1. Increase the alpha level. Perhaps you could test at α = 0.05 instead of α = 0.01.

2. Increase the sample size.

3. Decrease the magnitude of the standard deviations. Perhaps the study can be redesigned so that measurements are more precise and extraneous sources of variation are removed.

One-Sample Z-Test

This section presents the results of the traditional one-sample Z-test, which can be used when the population standard deviation is known. Because the population standard deviation is rarely known, this test is not commonly used in practice. Here, reports for all three alternative hypotheses are shown, but a researcher would typically choose one of the three before generating the output. All three tests are shown here for the purpose of exhibiting all the output options available.

One-Sample Z-Test

Alternative Hypothesis	Mean	σ	Standard Error	Z-Statistic	P-Value	Reject H0 at α = 0.05?
μ ≠ 130	139.6	40	8.944272	1.0733	0.28313	No
μ < 130	139.6	40	8.944272	1.0733	0.85844	No
μ > 130	139.6	40	8.944272	1.0733	0.14157	No

Alternative Hypothesis

The (unreported) null hypothesis is

$$H_0: \mu = Hypothesized\ Value$$

and the alternative hypotheses,

$$H_a: \mu \neq Hypothesized\ Value,$$

$$H_a: \mu < Hypothesized\ Value, \text{ or}$$

$$H_a: \mu > Hypothesized\ Value.$$

In practice, the alternative hypothesis should be chosen in advance.

Mean

This is the average of the data values.

σ (Known Standard Deviation)

This is the known pre-specified value for the population standard deviation.

Standard Error

This is the standard deviation of the distribution of sample means.

$$SE_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Z-Statistic

The Z-Statistic is the value used to produce the *p*-value (Prob Level) based on the standard normal *Z* distribution. The formula for the Z-Statistic is:

$$Z - Statistic = \frac{\overline{x} - Hypothesized\ Value}{SE_{\overline{x}}}$$

P-Value

The *p*-value, also known as the probability level or significance level, is the probability that the test statistic will take a value at least as extreme as the observed value, assuming that the null hypothesis is true. If the *p*-value is less than the prescribed *α*, in this case 0.05, the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, there is not sufficient evidence to reject the null hypothesis.

Reject H0 at α = 0.05?

This column indicates whether or not the null hypothesis is rejected, in favor of the alternative hypothesis, based on the *p*-value and chosen *α*. A test in which the null hypothesis is rejected is sometimes called *significant*.

Randomization Test

A randomization test is conducted by first determining the signs of all the values relative to the null hypothesized mean – that is, the signs of the values after subtracting the null hypothesized mean. Then all possible permutations of the signs are enumerated. Each of these permutations is assigned to absolute values of the original subtracted values in their original order, and the corresponding t-statistic is calculated. The original t-statistic, based on the original data, is compared to each of the permutation t-statistics, and the total count of permutation t-statistics more extreme than the original t-statistic is determined. Dividing this count by the number of permutations tried gives the significance level of the test.

For even moderate sample sizes, the total number of permutations is in the trillions, so a Monte Carlo approach is used in which the permutations are found by random selection rather than complete enumeration. Edgington suggests that at least 1,000 permutations by selected. We suggest that this be increased to 10,000.

Randomization Test (Two-Sided)
───

Alternative Hypothesis:	The distribution center is not 130
Number of Monte Carlo samples:	10000
User-Entered Random Seed:	5213861

───

Test	P-Value	Reject H0 at α = 0.05?
Randomization Test	0.32470	No

───

P-Value

The *p*-value, also known as the probability level or significance level, is the probability that the test statistic will take a value at least as extreme as the observed value, assuming that the null hypothesis is true. If the *p*-value is less than the prescribed *α*, in this case 0.05, the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, there is not sufficient evidence to reject the null hypothesis.

Reject H0 at α = 0.05?

This column indicates whether or not the null hypothesis is rejected, in favor of the alternative hypothesis, based on the *p*-value and chosen *α*. A test in which the null hypothesis is rejected is sometimes called *significant*.

Quantile (Sign) Test

The quantile (sign) test is one of the older nonparametric procedures. Each of the values in the sample is compared to the null hypothesized quantile value and determined whether it is larger or smaller. Values equal to the null hypothesized value are removed. The binomial distribution based on the Quantile Test Proportion is used to determine the probabilities in each direction of obtaining such a sample or one more extreme if the null quantile is the true one. These probabilities are the *p*-values for the one-sided test of each direction. The two-sided *p*-value is twice the value of the smaller of the two one-sided *p*-values.

When the quantile of interest is the median (the quantile test proportion is 0.5), a quantile test is called the *sign test*.

While the quantile (sign) test is simple, there are more powerful nonparametric alternatives, such as the Wilcoxon signed-rank test. However, if the shape of the underlying distribution of a variable is the double exponential distribution, the sign test may be the better choice.

Quantile (Sign) Test

Null Quantile (Q0)	Quantile Proportion	Number of Values		P-Value		
		Lower	Higher	H1: Q ≠ Q0	H1: Q < Q0	H1: Q > Q0
130	0.5	11	9	0.82380	0.41190	0.74828

This Quantile test is equivalent to the Sign test if the Quantile Proportion is 0.5.

Null Quantile (Q0)

Under the null hypothesis, the proportion of all values below the null quantile is the quantile proportion. For the sign test, the null quantile is the null median.

Quantile Proportion

Under the null hypothesis, the quantile proportion is the proportion of all values below the null quantile. For the sign test, this proportion is 0.5.

Number of Values (Lower and Higher)

This is the actual number of values that are below and above the null quantile.

P-Value for H1: Q ≠ Q0

This is the two-sided probability that the true quantile is equal to the stated null quantile (Q0), for the quantile proportion stated and given the observed values. A small *p*-value indicates that the true quantile for the stated quantile proportion is different from the null quantile.

P-Value for H1: Q < Q0

This is the one-sided probability that the true quantile is greater than or equal to the stated null quantile (Q0), for the quantile proportion stated and given the observed values. A small *p*-value indicates that the true quantile for the stated quantile proportion is less than the null quantile.

P-Value for H1: Q > Q0

This is the one-sided probability that the true quantile is less than or equal to the stated null quantile (Q0), for the quantile proportion stated and given the observed values. A small *p*-value indicates that the true quantile for the stated quantile proportion is greater than the null quantile.

Wilcoxon Signed-Rank Test

This nonparametric test makes use of the sign and the magnitude of the rank of the differences (original data minus the hypothesized value). It is typically the best nonparametric alternative to the one-sample t-test.

Wilcoxon Signed-Rank Test
───

Test Details
───

Sum of Ranks (W)	Mean of W	Standard Deviation of W	Number of Zeros	Number of Sets of Ties	Multiplicity Factor
126	105	26.77686	0	4	24

Test Results
───

Test Type	Alternative Hypothesis	Z-Value	P-Value	Reject H0 at α = 0.05?
Exact*	Median ≠ 130			
Exact*	Median < 130			
Exact*	Median > 130			
Normal Approximation	Median ≠ 130	0.7843	0.43289	No
Normal Approximation	Median < 130	0.7843	0.78356	No
Normal Approximation	Median > 130	0.7843	0.21644	No
Normal Approx. with C.C.	Median ≠ 130	0.7656	0.44392	No
Normal Approx. with C.C.	Median < 130	0.8029	0.78899	No
Normal Approx. with C.C.	Median > 130	0.7656	0.22196	No

* The Exact Test is provided only when there are no ties.
───

Sum of Ranks (W)

The basic statistic for this test is the sum of the positive ranks, $\sum R_+$ (The sum of the positive ranks is chosen arbitrarily. The sum of the negative ranks could equally be used). This statistic is called W.

$$W = \sum R_+$$

Mean of W

This is the mean of the sampling distribution of the sum of ranks for a sample of n items.

$$\mu_W = \frac{n(n+1) - d_0(d_0 + 1)}{4}$$

where d_0 is the number of zero differences.

Standard Deviation of W

This is the standard deviation of the sampling distribution of the sum of ranks. Here t_i represents the number of times the ith value occurs.

$$s_W = \sqrt{\frac{n(n+1)(2n+1) - d_0(d_0+1)(2d_0+1)}{24} - \frac{\sum t_i^3 - \sum t_i}{48}}$$

where d_0 is the number zero differences, t_i is the number of absolute differences that are tied for a given non-zero rank, and the sum is over all sets of tied ranks.

Number of Zeros

This is the number of times that the difference between the observed value (or difference) and the hypothesized value is zero. The zeros are used in computing ranks but are not considered positive ranks or negative ranks.

Number of Sets of Ties

The treatment of ties is to assign an average rank for the particular set of ties. This is the number of sets of ties that occur in the data, including ties at zero.

Multiplicity Factor

This is the correction factor that appeared in the standard deviation of the sum of ranks when there were ties.

Alternative Hypothesis

For the Wilcoxon signed-rank test, the null and alternative hypotheses relate to the median. In the two-tail test for the median difference (assuming a hypothesized value of 0), the null hypothesis would be H_0: Median = 0 with the alternative being H_a: Median ≠ 0.

The left-tail alternative is represented by Median < 0 (i.e., H_a: median < 0) while the right-tail alternative is depicted by Median > 0.

Exact Probability: P-Value

This is an exact *p*-value for this statistical test, assuming no ties. The *p*-value is the probability that the test statistic will take on a value at least as extreme as the actually observed value, assuming that the null hypothesis is true. If the *p*-value is less than *α*, say 5%, the null hypothesis is rejected. If the *p*-value is greater than *α*, the null hypothesis is accepted.

Exact Probability: Reject H0 at α = 0.05?

This is the conclusion reached about the null hypothesis. It will be to either fail to reject H_0 or reject H_0 at the assigned level of significance.

Approximations with (and without) Continuity Correction: Z-Value

Given the sample size is at least ten, a normal approximation method may be used to approximate the distribution of the sum of ranks. Although this method does correct for ties, it does not have the continuity correction factor. The *z* value is as follows:

$$z = \frac{W - \mu_W}{\sigma_W}$$

If the correction factor for continuity is used, the formula becomes:

$$z = \frac{W - \mu_W \pm \frac{1}{2}}{\sigma_W}$$

Approximations with (and without) Continuity Correction: P-Value

This is the *p*-value for the normal approximation approach for the Wilcoxon signed-rank test. The *p*-value is the probability that the test statistic will take a value at least as extreme as the actually observed value, assuming that the null hypothesis is true. If the *p*-value is less than *α*, say 5%, the null hypothesis is rejected. If the *p*-value is greater than *α*, the null hypothesis is accepted.

Approximations with (and without) Continuity Correction: Reject H0 at α = 0.05?

This is the conclusion reached about the whether to reject null hypothesis. It will be either Yes or No at the given level of significance.

Tests of the Normality Assumption

Tests of the Normality Assumption

Normality Test	Test Statistic	P-Value	Reject the Assumption of Normality at α = 0.05?
Shapiro-Wilk	0.9298	0.15338	No
Skewness	1.0233	0.30614	No
Kurtosis	-0.9607	0.33670	No
Omnibus (Skewness or Kurtosis)	1.9702	0.37340	No

The main assumption when using the t-test is that the data are normally distributed. The normality assumption can be checked statistically by the Shapiro-Wilk, skewness, kurtosis, or omnibus normality tests and visually by the histogram or normal probability plot.

In the case of non-normality, one of the nonparametric tests may be considered. Sometimes a transformation, such as the natural logarithm or the square root of the original data, can change the underlying distribution from skewed to normal. To evaluate whether the underlying distribution of the variable is normal after the transformation, rerun the normal probability plot on the transformed variable. If some of the data values are negative or zero, it may be necessary to add a constant to the original data prior to the transformation. If the transformation produces a suitable result, then the one-sample t-test could be performed on the transformed data.

Normality Test (Shapiro-Wilk, Skewness, Kurtosis, and Omnibus)

These four tests allow you to test the skewness, kurtosis, and overall normality of the data. If any of them reject the hypothesis of normality, the data should not be considered normal. These tests are discussed in more detail in the Descriptive Statistics chapter.
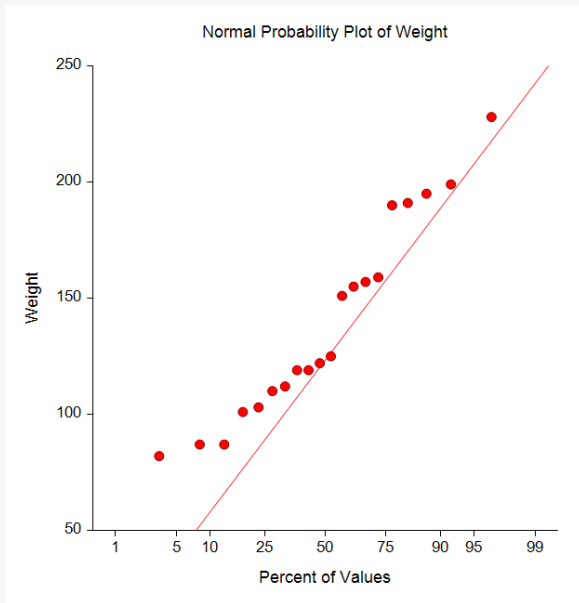
Histogram

Histogram of Weight
_____
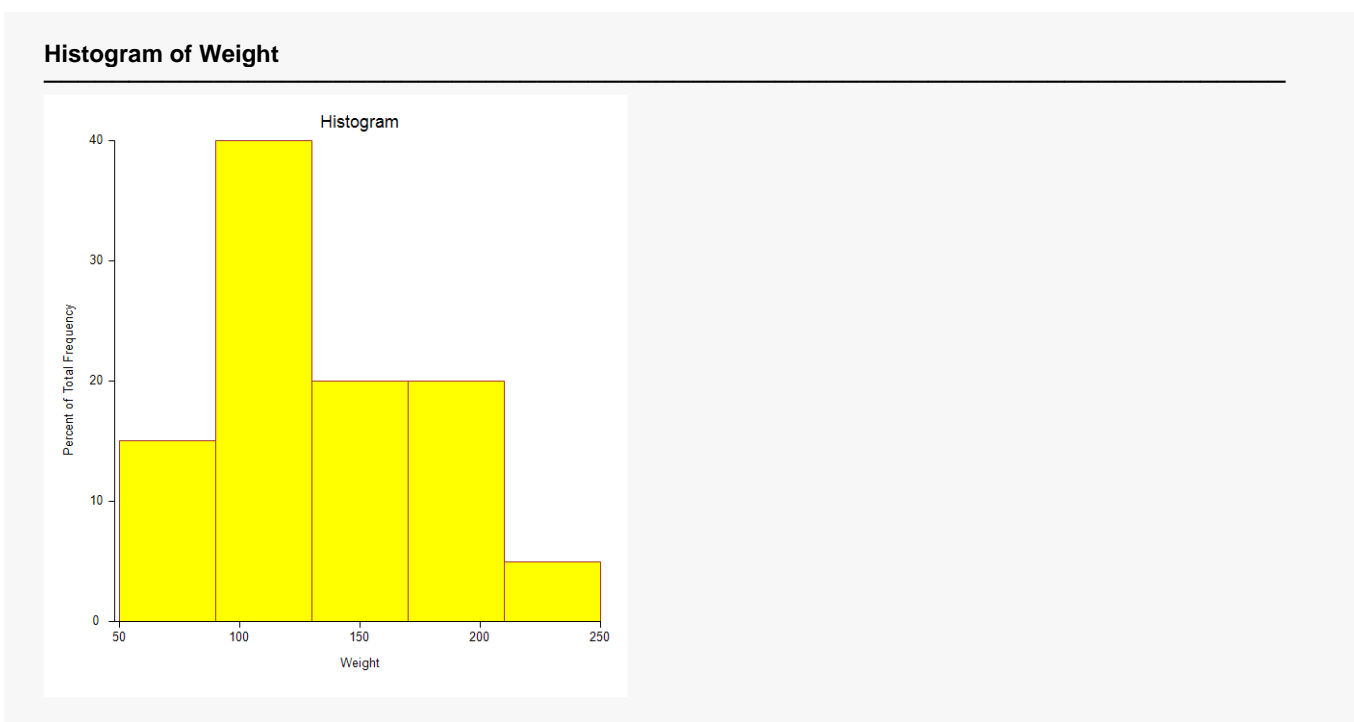
The histogram provides a general idea of the shape. The histogram appearance can be affected considerably by the bin width choices, particularly when the sample size is smaller.
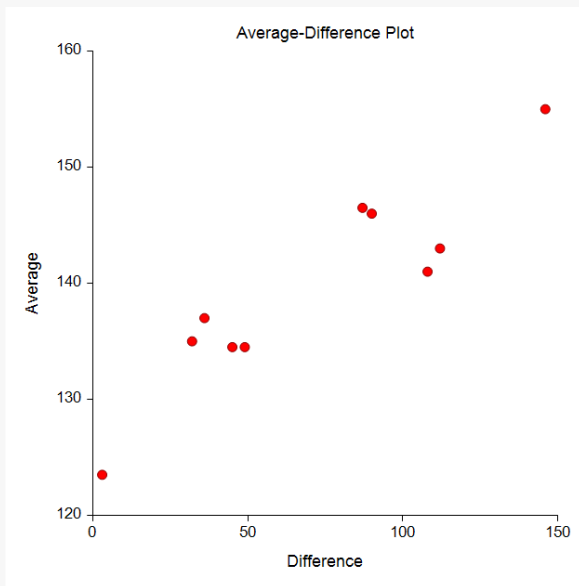
Probability Plot

Probability Plot of Weight
_____

If the observations fall along a straight line, this indicates the data follow a normal distribution.

Average vs Difference Plot

Average vs Difference Plot of Weight
_____

This average-difference plot is designed to detect a lack of symmetry in the data. This plot is constructed from the original data. Let D(i) represent the ith ordered value. Pairs of these sorted values are considered, with the pairing being done as you move toward the middle from either end. That is, consider the pairs D(1)

and D(n), D(2) and D(n-1), D(3) and D(n-2), etc. Plot the average versus the difference of each of these pairs. Your plot will have about n/2 points, depending on whether n is odd or even. If the data are symmetric, the average of each pair will be the close to the median.

Symmetry is an important assumption for some of the available tests. A perfectly symmetric set of data should show a vertical line of points hitting the horizontal axis at the value of the median. Departures from symmetry would deviate from this standard.

One-Sample T-Test Checklist

This checklist, prepared by a professional statistician, is a flowchart of the steps you should complete to conduct a valid one-sample t-test (or one of its nonparametric counterparts). You should complete these tasks in order.

Step 1 – Data Preparation

Introduction

This step involves scanning your data for anomalies, data entry errors, typos, and so on. Frequently we hear of people who completed an analysis with the right techniques but obtained strange conclusions because they had mistakenly selected the data.

Sample Size

The sample size (number of non-missing rows) has a lot of ramifications. The larger the sample size for the one-sample t-test the better. Of course, the t-test may be performed on very small samples, say 4 or 5 observations, but it is impossible to assess the validity of assumptions with such small samples. It is our statistical experience that at least 20 observations are necessary to evaluate normality properly. On the other hand, since skewness can have unpleasant effects on t-tests with small samples, particularly for one-tailed tests, larger sample sizes (30 to 50) may be necessary.

It is possible to have a sample size that is too large for a statistical significance test. When your sample size is very large, you are almost guaranteed to find statistical significance. However, the question that then arises is whether the magnitude of the difference is of practical importance.

Missing Values

The number and pattern of missing values is always an issue to consider. Usually, we assume that missing values occur at random throughout your data. If this is not true, your results will be biased since a particular segment of the population is underrepresented. If you have a lot of missing values, some researchers recommend comparing other variables with respect to missing versus non-missing. If you find large differences in other variables, you should begin to worry about whether the missing values might cause a systematic bias in your results.

Type of Data

The mathematical basis of the t-test assumes that the data are continuous. Because of the rounding that occurs when data are recorded, all data are technically discrete. The validity of assuming the continuity of the data then comes down to determining when we have too much rounding. For example, most statisticians would not worry about human-age data that was rounded to the nearest year. However, if these data were rounded to the nearest ten years or further to only three groups (young, adolescent, and adult), most statisticians would question the validity of the probability statements. Some studies have shown that the t-test is reasonably accurate when the data has only five possible values (most would call this discrete data). If your data contains less than five unique values, any probability statements made are tenuous.

Outliers

Generally, outliers cause distortion in statistical tests. You must scan your data for outliers (the box plot is an excellent tool for doing this). If you have outliers, you have to decide if they are one-time occurrences or if they would occur in another sample. If they are one-time occurrences, you can remove them and proceed. If you know they represent a certain segment of the population, you have to decide between biasing your results (by removing them) or using a nonparametric test that can deal with them. Most would choose the nonparametric test.

Step 2 – Setup and Run the Procedure

Introduction

NCSS is designed to be simple to operate, but it requires some learning. When you go to run a procedure such as this for the first time, take a few minutes to read through the chapter again and familiarize yourself with the issues involved.

Enter Variables

The NCSS panels are set with ready-to-run defaults. About all you have to do is select the appropriate variable.

Select All Plots

As a rule, you should select all diagnostic plots (histograms, etc.). They add a great deal to your analysis of the data.

Specify Alpha

Most beginners in statistics forget this important step and let the alpha value default to the standard 0.05. You should make a conscious decision as to what value of alpha is appropriate for your study. The 0.05 default came about when people had to rely on printed probability tables in which there were only two values available: 0.05 or 0.01. Now you can set the value to whatever is appropriate.

Step 3 – Check Assumptions

Introduction

Once the program output is displayed, you will be tempted to go directly to the probability of the t-test, determine if you have a significant result, and proceed to something else. However, it is very important that you proceed through the output in an orderly fashion. The first task is to determine which of the assumptions are met by your data.

Sometimes, when the data are nonnormal, a data transformation (like square roots or logs) might normalize the data. Frequently, this kind of transformation or re-expression approach works very well. However, always check the transformed variable to see if it is normally distributed.

It is not unusual in practice to find a variety of tests being run on the same basic null hypothesis. That is, the researcher who fails to reject the null hypothesis with the first test will sometimes try several others and stop when the hoped-for significance is obtained. For instance, a statistician might run the one-sample t-test on the original data, the one-sample t-test on the logarithmically transformed data, the Wilcoxon rank-sum test, and the Quantile test. An article by Gans (1984) suggests that there is no harm on the true significance level if no more than two tests are run. This is not a bad option in the case of questionable outliers. However, as a rule of thumb, it seems more honest to investigate whether the data is normal. The conclusion from that investigation should direct you to the right test.

Random Sample

The validity of this assumption depends on the method used to select the sample. If the method used ensures that each individual in the population of interest has an equal probability of being selected for this sample, you have a random sample. Unfortunately, you cannot tell if a sample is random by looking at either it or statistics from it.

Check Descriptive Statistics

You should check the Descriptive Statistics Section first to determine if the Count and the Mean are reasonable. If you have selected the wrong variable, these values will alert you.

Normality

To validate this assumption, you would first look at the plots. Outliers will show up on the box plots and the probability plots. Skewness, kurtosis, more than one mode, and a host of other problems will be obvious from the density trace on the histogram. After considering the plots, look at the Tests of Assumptions Section to get numerical confirmation of what you see in the plots. Remember that the power of these normality tests is directly related to the sample size, so when the normality assumption is accepted, double-check that your sample is large enough to give conclusive results (at least 20).

Symmetry

The nonparametric tests need the assumption of symmetry. The easiest ways to evaluate this assumption are from the density trace on the histogram or from the average-difference plot.

Step 4 – Choose the Appropriate Statistical Test

Introduction

After understanding how your data fit the assumptions of the various one-sample tests, you are ready to determine which statistical procedures will be valid. You should select one of the following three situations based on the status of the normality.

Normal Data

Use the T-Test Section for hypothesis testing and the Descriptive Statistics Section for interval estimation.

Nonnormal and Asymmetrical Data

Try a transformation, such as the natural logarithm or the square root, on the original data since these transformations frequently change the underlying distribution from skewed to normal. If some of the data values are negative or zero, add a constant to the original data prior to the transformation. If the transformed data is now normal, use the T-Test Section for hypothesis testing and the Descriptive Statistics Section for interval estimation.

Nonnormal and Symmetrical Data

Use the Wilcoxon Rank-Sum Test or the Quantile Test for hypothesis testing.

Step 5 – Interpret Findings

Introduction

You are now ready to conduct your test. Depending on the nature of your study, you should look at either of the following sections.

Hypothesis Testing

Here you decide whether to use a two-tailed or one-tailed test. The two-tailed test is the standard. If the probability level is less than your chosen alpha level, reject the null hypothesis of equality to a specified mean (or median) and conclude that the mean is different. Your next task is to look at the mean itself to determine if the size of the difference is of practical interest.

Confidence Limits

The confidence limits let you put bounds on the size of the mean (for one independent sample) or mean difference (for dependent samples). If these limits are narrow and close to your hypothesized value, you might determine that even though your results are statistically significant, there is no practical significance.

Step 6 – Record Your Results

Finally, as you finish a test, take a moment to jot down your impressions. Explain what you did, why you did it, what conclusions you reached, which outliers you deleted, areas for further investigation, and so on. Since this is a technical process, your short-term memory will not retain these details for long. These notes will be worth their weight in gold when you come back to this study a few days later!

Example of One-Sample T-Test Steps

This example will illustrate the use of one-sample tests for a single variable. A registration service for a national motel/hotel chain wants the average wait time for incoming calls on Mondays (during normal business hours, 8:00 a.m. to 5:00 p.m.) to be less than 25 seconds. A random sample of 30 calls yielded the results shown below.

Row	AnsTime	Row	AnsTime
1	15	16	8
2	12	17	12
3	25	18	30
4	11	19	12
5	20	20	25
6	10	21	26
7	16	22	16
8	26	23	29
9	21	24	22
10	23	25	12
11	32	26	12
12	34	27	12
13	17	28	30
14	16	29	15
15	16	30	39

Step 1 – Data Preparation

This is not paired data but just a single random sample of one variable. There are no missing values, and the variable is continuous.

Step 2 – Setup and Run the One-Sample T-Test

Select and run the One-Sample T-Test from the Analysis menu on the single variable, AnsTime. The alpha value has been set at 0.05. Interpretation of the results will come in the steps to follow.
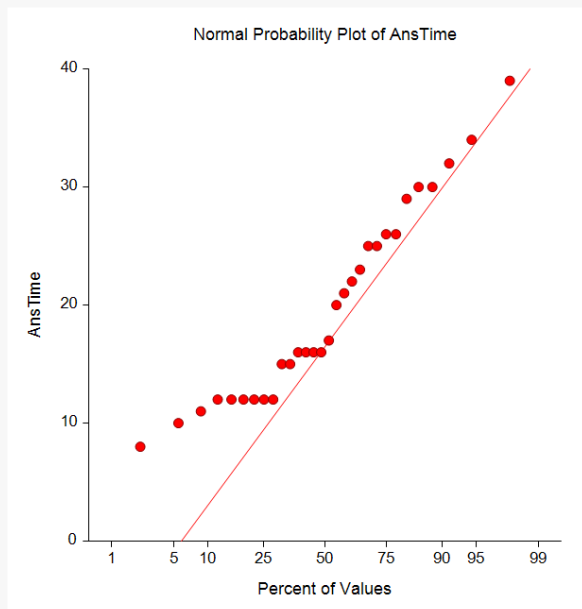
Step 3 – Check Assumptions

The major assumption to check for is normality, and you should begin with the graphic perspectives: normal probability plots, histograms, density traces, and box plots. Some of these plots are given below.
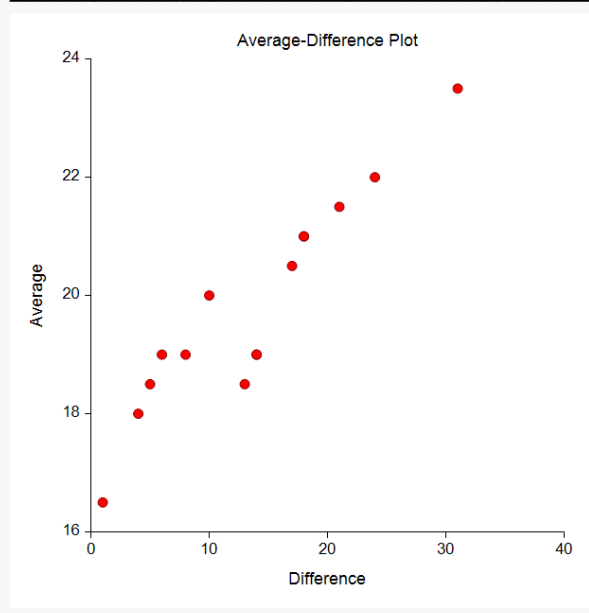
Histogram of AnsTime
_____
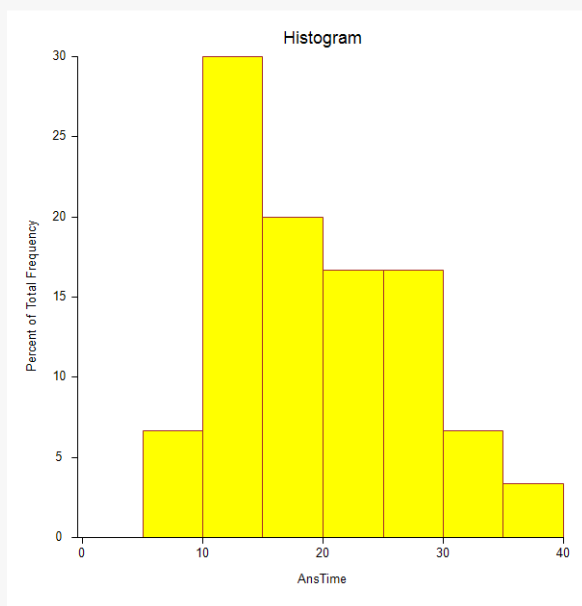
Probability Plot of AnsTime
_____

Average vs Difference Plot of AnsTime

The normal probability plot above does not look straight. It shows some skewness to the right. The histogram confirms the skewness to the right. This type of skewness to the right turns up quite often when dealing with elapsed-time data.

The skewness, kurtosis, and the omnibus normality tests in the output below have *p*-values greater than 0.05, indicating that answer time seems to be normally distributed. This conflict in conclusions between the normal probability plot and the normality tests is probably due to the fact that this sample size is not large enough to accurately assess the normality of the data.

Tests of the Normality Assumption

Normality Test	Test Statistic	P-Value	Reject the Assumption of Normality at α = 0.05?
Shapiro-Wilk	0.9313	0.05313	No
Skewness	1.4246	0.15428	No
Kurtosis	-0.7398	0.45945	No
Omnibus (Skewness or Kurtosis)	2.5766	0.27573	No

Step 4 – Choose the Appropriate Statistical Test

In Step 3, the conclusions from checking the assumptions were two-fold: (1) the data are continuous, and (2) the answer times are (based on the probability plot) non-normal. As a result of these findings, the appropriate statistical test is the Wilcoxon Signed-Rank test, which is shown in the figure. For comparison purposes, the t-test results are also shown in the output.

One-Sample T-Test

Alternative Hypothesis	Mean	Standard Error	T-Statistic	DF	P-Value	Reject H0 at α = 0.05?
μ < 25	19.8	1.496663	-3.4744	29	0.00082	Yes

Wilcoxon Signed-Rank Test

Test Details

Sum of Ranks (W)	Mean of W	Standard Deviation of W	Number of Zeros	Number of Sets of Ties	Multiplicity Factor
90	231	48.52448	2	7	378

Test Results

Test Type	Alternative Hypothesis	Z-Value	P-Value	Reject H0 at α = 0.05?
Exact*	Median < 25			
Normal Approximation	Median < 25	-2.9057	0.00183	Yes
Normal Approx. with C.C.	Median < 25	-2.8954	0.00189	Yes

* The Exact Test is provided only when there are no ties.

Step 5 – Interpret Findings

Since the nonparametric test is more appropriate here and the concern was that the average answer time was less than 25 seconds, the Median < 25 is the proper alternative hypothesis. The *p*-value for the Wilcoxon Signed-Rank test is 0.00183, which is much less than 0.05. Thus, the conclusion of the test is to reject the null hypothesis. This says that the median answer time is significantly less than 25 seconds.

It is interesting to note that the *p*-value for the left-tailed t-test is about the same. This points out the robustness of the t-test in the cases of heavy-tailed but almost symmetric distributions.

Step 6 – Record Your Results

The conclusions for this example are that the median is less than 25 seconds. Again, if you were troubled by the shape of the distribution, you could use a transformation such as the natural logarithm to make the data more normal and try the t-test. However, in this case, that seems to be more work than is needed.