



# 1

# Statistical Inference and t-Tests

## Objectives

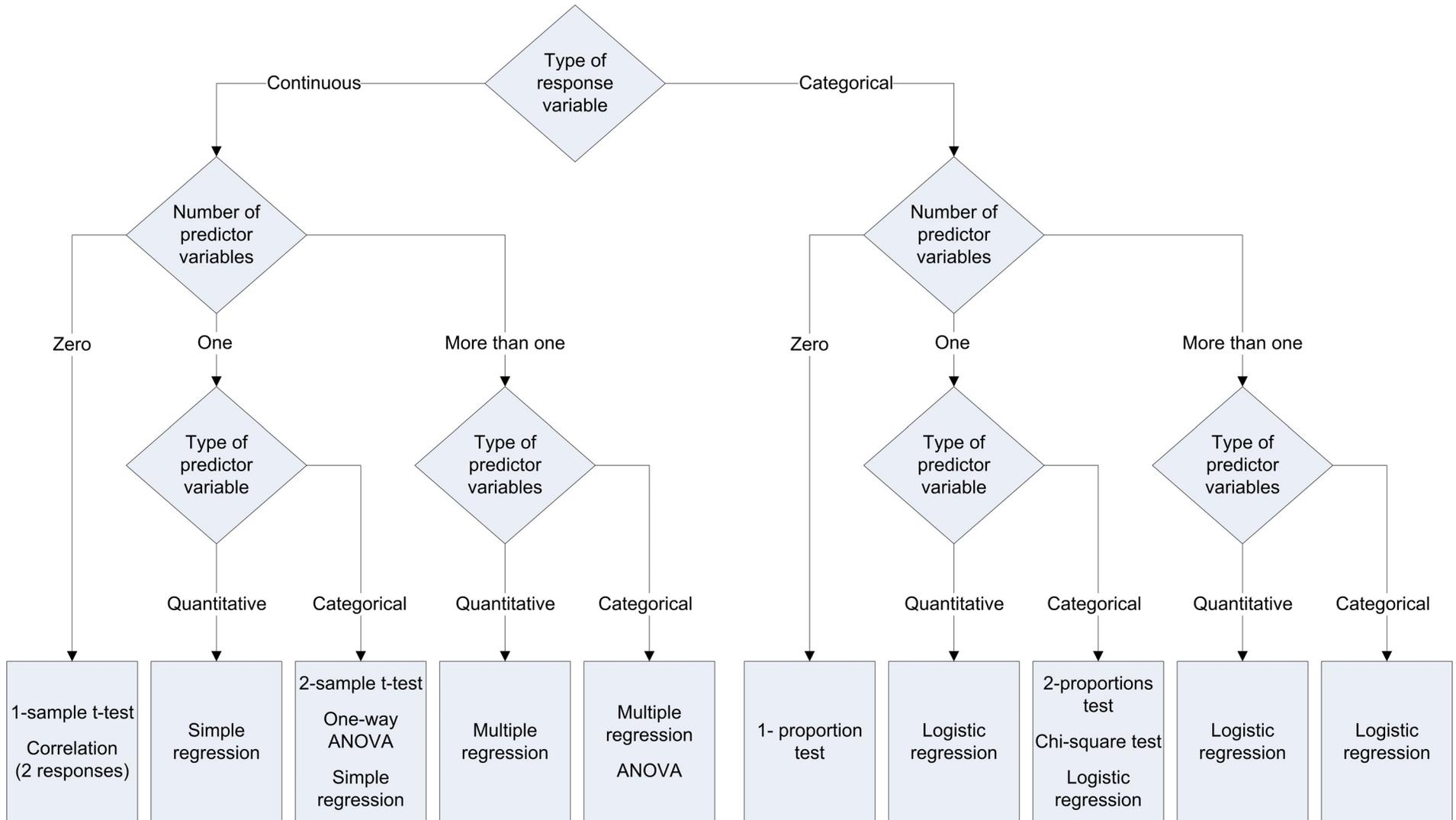
- Evaluate the difference between a sample mean and a target value using a one-sample t-test.
- Evaluate the difference between a sample mean and a target value using a confidence interval.
- Assess the power of a hypothesis test using power analysis.
- Evaluate the difference between two sample means using a two-sample t-test.
- Evaluate the differences between paired observations using a paired t-test.

# Contents

<b>Examples and Exercises</b>	<b>Purpose</b>	<b>Page</b>
<b>Choosing an Analysis</b>		
Example 1 Mortgage Process Time	Evaluate the difference between mean mortgage processing time and a target value using a one-sample t-test.	1-5
Exercise A Surgical Scheduling Time	Evaluate the difference between mean surgical time and a target value using a one-sample t-test.	1-19
<b>Power and Sample Size</b>		
Example 2 Evaluating Power	Assess the power of a hypothesis test.	1-20
<b>Two-Sample t-Test</b>		
Example 3 Customer Complaints	Evaluate the differences in the mean number of customer complaints using a two-sample t-test.	1-29
Exercise B Call Center Handling Times	Compare the difference in call center handling times using a two-sample t-test.	1-41
Exercise C Salary Comparison	Compare the difference in household salaries in two neighborhoods with a two-sample t-test.	1-42

Examples and Exercises	Purpose	Page
Paired t-Test		
Example 4 ATM Surrounds	Evaluate the difference in ATM usage before and after installation of shelters	1-43
Exercise D Car Satisfaction Ratings	Use a paired t-test to compare the difference in car satisfaction ratings one week and one year after customers purchase the car.	1-49

# Choosing an Analysis



# One-Sample t-Test

## Example 1 Mortgage Process Time

### Problem

A faster loan processing time produces higher productivity and greater customer satisfaction. A financial services institution wants to establish a baseline for their process by estimating their mean processing time. They also want to determine if their mean time differs from a competitor's claim of 6 hours.

### Data collection

A financial analyst randomly selects 6 loan applications from the past 2 weeks and manually calculates the time between loan initiation and when the customer receives the institution's decision.

### Tools

- **1-Sample t**
- **Normality test**
- **Time series plot**

<b>Variable</b>	<b>Description</b>
Date	Date of customer notification
Hours	Number of hours until customer receives notification

# Hypothesis testing

## What is a hypothesis test

A hypothesis test uses sample data to test a hypothesis about the population from which the sample was taken. The one-sample t-test is one of many procedures available for hypothesis testing in Minitab.

For example, to test whether the mean duration of a transaction is equal to the desired target, measure the duration of a sample of transactions and use its sample mean to estimate the mean for all transactions. This is an example of *statistical inference*, which is using information about a sample to make an inference about a population.

## When to use a hypothesis test

Use a hypothesis test to make inferences about one or more populations when sample data are available.

## Why use a hypothesis test

Hypothesis testing can help answer questions such as:

- Are turn-around times meeting or exceeding customer expectations?
- Is the service at one branch better than the service at another?

For example,

- On average, is a call center meeting the target time to answer customer questions?
- Is the mean billing cycle time shorter at the branch with a new billing process?

## One-sample t-test

### What is a one-sample t-test

A one-sample t-test helps determine whether  $\mu$  (the population mean) is equal to a hypothesized value (the test mean).

The test uses the standard deviation of the sample to estimate  $\sigma$  (the population standard deviation). If the difference between the sample mean and the test mean is large relative to the variability of the sample mean, then  $\mu$  is unlikely to be equal to the test mean.

### When to use a one-sample t-test

Use a one-sample t-test when continuous data are available from a single random sample.

The test assumes the population is normally distributed. However, it is fairly robust to violations of this assumption for sample sizes equal to or greater than 30, provided the observations are collected randomly and the data are continuous, unimodal, and reasonably symmetric (see [1]).

### Why use a one-sample t-test

A one-sample t-test can help answer questions such as:

- Is the mean transaction time on target?
- Does customer service meet expectations?

For example,

- On average, is a call center meeting the target time to answer customer questions?
- Is the billing cycle time for a new process shorter than the current cycle time of 20 days?

## Testing the null hypothesis

The company wants to determine whether the mean time for the approval process is statistically different from the competitor's claim of 6 hours. In statistical terms, the process mean is the population mean, or  $\mu$  (mu).

### Statistical hypotheses

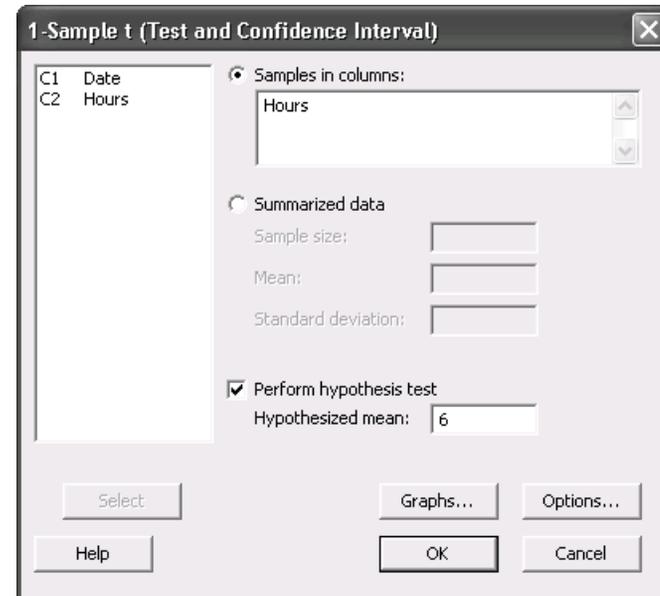
Either  $\mu$  is equal to 6 hours or it is not. You can state these alternatives with two hypotheses:

- The *null hypothesis* ( $H_0$ ):  $\mu$  is equal to 6 hours.
- The *alternative hypothesis* ( $H_1$ ):  $\mu$  is *not* equal to 6 hours.

Because the analysts will not measure every loan request in the population, they will not know the true value of  $\mu$ . However, an appropriate hypothesis test can help them make an informed decision. For these data, the appropriate test is a 1-sample t-test.

### 1-Sample t

- 1 Choose **Stat** ► **Basic Statistics** ► **1-Sample t**.
- 2 Complete the dialog box as shown below.



- 3 Click **OK**.

## Interpreting your results

### The logic of hypothesis testing

All hypothesis tests follow the same steps:

- 1 Assume  $H_0$  is true.
- 2 Determine how different the sample is from what you expected under the above assumption.
- 3 If the sample statistic is sufficiently unlikely under the assumption that  $H_0$  is true, then reject  $H_0$  in favor of  $H_1$ .

For example, the t-test results indicate that the sample mean is 4.792 hours. The test answers the question, “If  $\mu$  is equal to 6 hours, how likely is it to obtain a sample mean at least as different from 6 hours as the one you observed?” The answer is given as a probability value (P), which for this test is equal to 0.081.

### Test statistic

The t-statistic (-2.18) is calculated as:

$$t = (\text{sample mean} - \text{test mean}) / \text{SE Mean}$$

where SE Mean is the standard error of the mean (a measure of variability). As the absolute value of the t-statistic increases, the p-value becomes smaller.

### One-Sample T: Hours

Test of  $\mu = 6$  vs not = 6

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Hours	6	4.792	1.355	0.553	(3.370, 6.213)	-2.18	0.081

## Interpreting your results

### Making a decision

To make a decision, choose the significance level,  $\alpha$  (alpha), before the test:

- If P is less than or equal to  $\alpha$ , reject  $H_0$ .
- If P is greater than  $\alpha$ , fail to reject  $H_0$ . (Technically, you never *accept*  $H_0$ . You simply fail to reject it.)

A typical value for  $\alpha$  is 0.05, but you can choose higher or lower values depending on the sensitivity required for the test and the consequences of incorrectly rejecting the null hypothesis. Assuming an  $\alpha$ -level of 0.05 for the mortgage data, not enough evidence is available to reject  $H_0$ . P (0.081) is greater than  $\alpha$ .

### What's next

Check the assumption of normality.

### One-Sample T: Hours

Test of  $\mu = 6$  vs not = 6

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Hours	6	4.792	1.355	0.553	(3.370, 6.213)	-2.18	0.081

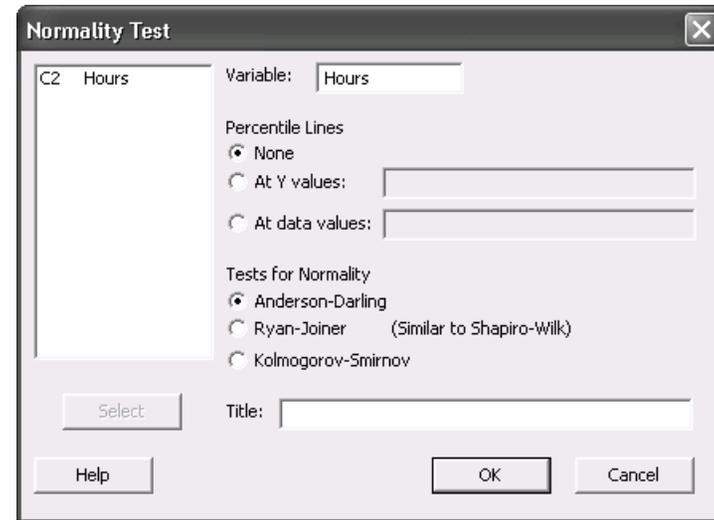
## Testing the assumption of normality

The 1-sample t-test assumes the data are sampled from a normally distributed population.

Use a normality test to determine whether the assumption of normality is valid for the data.

### Normality Test

- 1 Choose **Stat** ► **Basic Statistics** ► **Normality Test**.
- 2 Complete the dialog box as shown below.



- 3 Click **OK**.

## Interpreting your results

Use the normal probability plot to verify that the data do not deviate substantially from what is expected when sampling from a normal distribution.

- If the data come from a normal distribution, the points will roughly follow the fitted line.
- If the data do not come from a normal distribution, the points will not follow the line.

### Anderson-Darling normality test

The hypotheses for the Anderson-Darling normality test are:

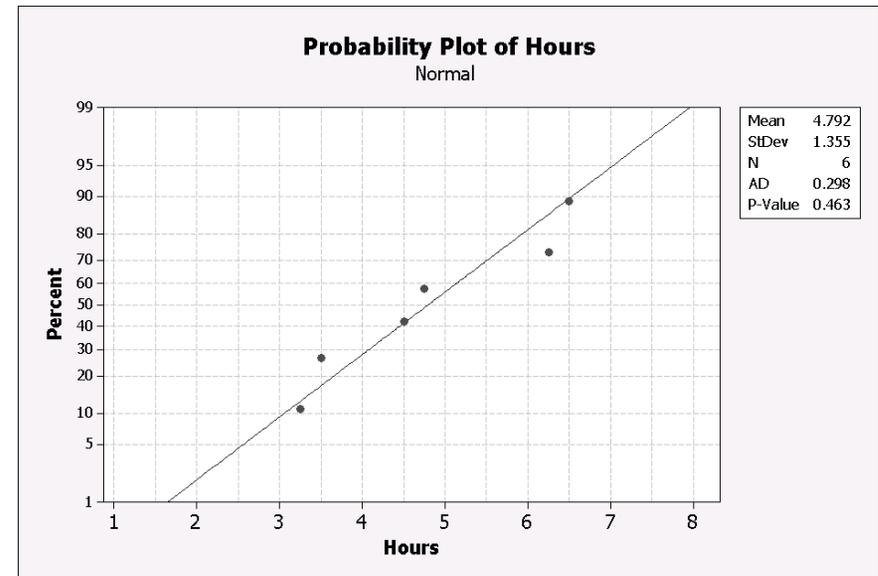
- $H_0$ : Data are from a normally distributed population
- $H_1$ : Data are not from a normally distributed population

The p-value from the Anderson-Darling test (0.463) assesses the probability that the data are from a normally distributed population. Using an  $\alpha$ -level of 0.05, there is insufficient evidence to suggest the data are not from a normally distributed population.

### Conclusion

Based on the plot and the normality test, assume that the data are from a normally distributed population.

**Note** | When data are not normally distributed, you may be able to transform them using a Box-Cox transformation or use a nonparametric procedure such as the 1-sample sign test.



### What's next

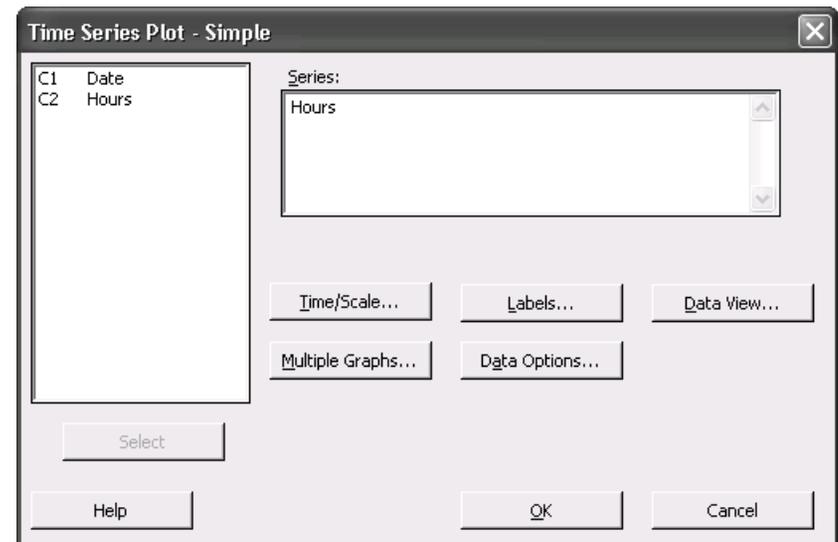
Check the data for non-random patterns over time.

## Testing the randomness assumption

Use a time series plot to look for trends or patterns in your data, which may indicate that your data are not random over time.

### Time Series Plot

- 1 Choose **Graph** ► **Time Series Plot**.
- 2 Choose **Simple**, then click **OK**.
- 3 Complete the dialog box as shown below.



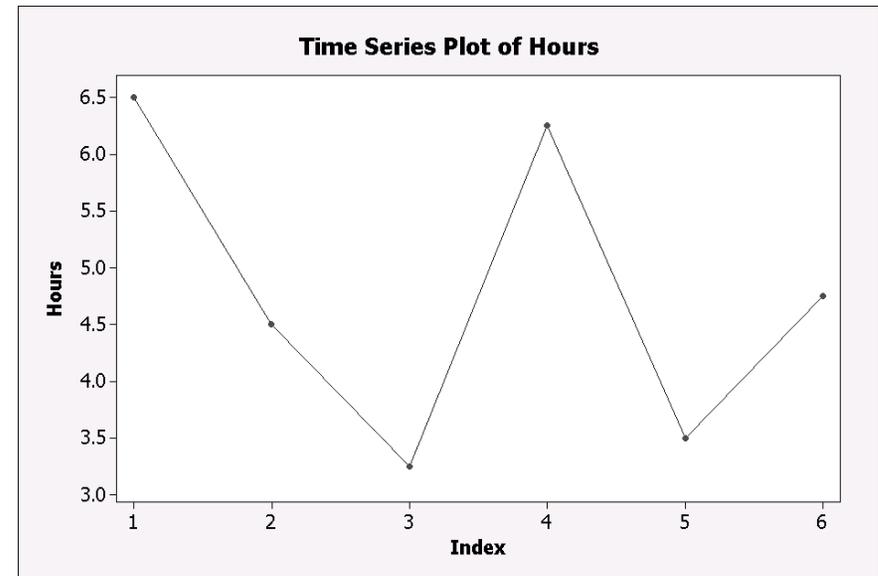
- 4 Click **OK**.

## Interpreting your results

If a trend or pattern exists in the data, we would want to understand the reasons for them. In this case, the data do not exhibit obvious trends or patterns.

### What's next

Calculate a confidence interval for the true population mean.



# Confidence intervals

## What is a confidence interval

A confidence interval is a range of likely values for a population parameter (such as  $\mu$ ) that is based on sample data. For example, with a 95% confidence interval for  $\mu$ , you can be 95% confident that the interval contains  $\mu$ . In other words, 95 out of 100 intervals will contain  $\mu$  upon repeated sampling.

## When to use a confidence interval

Use a confidence interval to make inferences about one or more populations from sample data, or to quantify the precision of your estimate of a population parameter, such as  $\mu$ .

## Why use a confidence interval

Confidence intervals can help answer many of the same questions as hypothesis testing:

- Is  $\mu$  on target?
- How much error exists in an estimate of  $\mu$ ?
- How low or high might  $\mu$  be?

For example,

- Is the mean transaction time longer than 30 seconds?
- What is the range of likely values for mean daily revenue?

## Interpreting your results

### Confidence interval

The 95% confidence interval for the average processing time is between 3.370 hours and 6.213 hours. The 95% confidence interval includes the comparison value of 6. This is equivalent to failing to reject the null hypothesis ( $H_0: \mu = 6$ ) for this t-test with an  $\alpha$  of 0.05.

### One-Sample T: Hours

Test of  $\mu = 6$  vs not = 6

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Hours	6	4.792	1.355	0.553	(3.370, 6.213)	-2.18	0.081

## Final considerations

### Summary and conclusions

According to the t-test and the sample data, you fail to reject the null hypothesis at the 0.05  $\alpha$ -level. In other words, the data do not provide sufficient evidence to conclude the mean processing time is significantly different from 6 hours.

The normality test and the time series plot indicate that the data meet the t-test's assumptions of normality and randomness.

The 95% confidence interval indicates the true value of the population mean is between 3.37 hours and 6.213 hours.

## Final considerations

### Summary and conclusions

#### Hypotheses

A hypothesis test always starts with two opposing hypotheses.

The null hypothesis ( $H_0$ ):

- Usually states that some property of a population (such as the mean) is not different from a specified value or from a benchmark.
- Is assumed to be true until sufficient evidence indicates the contrary.
- Is never proven true; you simply fail to disprove it.

The alternative hypothesis ( $H_1$ ):

- States that the null hypothesis is wrong.
- Can also specify the direction of the difference.

#### Significance level

Choose the  $\alpha$ -level *before* conducting the test.

- Increasing  $\alpha$  increases the chance of detecting a difference, but it also increases the chance of rejecting  $H_0$  when it is actually true (a Type I error).
- Decreasing  $\alpha$  decreases the chance of making a Type I error, but also decreases the chance of correctly detecting a difference.

#### Assumptions

Each hypothesis test is based on one or more assumptions about the data being analyzed. If these assumptions are not met, the conclusions may not be correct.

The assumptions for a one-sample t-test are:

- The sample must be random.
- Sample data must be continuous.
- Sample data should be normally distributed (although this assumption is less critical when the sample size is 30 or more).

The t-test procedure is fairly robust to violations of the normality assumption, provided that observations are collected randomly and the data are continuous, unimodal, and reasonably symmetric (see [1]).

#### Confidence interval

The confidence interval provides a likely range of values for  $\mu$  (or other population parameters).

You can conduct a two-tailed hypothesis test (alternative hypothesis of  $\neq$ ) using a confidence interval. For example, if the test value is not within a 95% confidence interval, you can reject  $H_0$  at the 0.05  $\alpha$ -level. Likewise, if you construct a 99% confidence interval and it does not include the test mean, you can reject  $H_0$  at the 0.01  $\alpha$ -level.

## Exercise A Surgical Scheduling Time

### Problem

Hospital administrators want to improve scheduling for tonsillectomies and to maximize the capacity of the operating rooms without rushing the surgeons. For accurate scheduling, they need to verify that the mean surgery time is 45 minutes.

### Data collection

Analysts record the operation time in minutes for each tonsillectomy over the course of 14 days.

### Instructions

- 1 Are the surgeons averaging 45 minutes per procedure? If not, is the average significantly different than 45 minutes? Does this difference have practical implications?
- 2 Investigate normality. Are the data normally distributed?
- 3 Investigate whether the data are random over time.

### Data set

Surgery.MPJ

Variable	Description
Start Time	Time and Date the operation began
End Time	Time and Date the operation was completed
Op_Time	Total operation time (minutes)

**Note** | Op\_Time is the difference between Start Time and End Time multiplied by the number of minutes in a day ( $24 \text{ hr} * 60 = 1440 \text{ min}$ ).

# Power and Sample Size

## Example 2 Evaluating Power

### Problem

The managers of a financial services institution are concerned about the results of the previous example (mortgage processing times) due to the small sample size and a wide confidence interval. They decide to conduct a power analysis to determine whether they collected enough sample data to detect a difference of 1 hour from the competitor's claim of 6 hours.

### Data collection

The managers base their power analysis on the results of the t-test in Example 1.

### Tools

- **Power and Sample Size for 1-Sample t**

### Data set

None

## Power analysis

### What is power analysis

Power is the ability of a test to detect a difference when one exists. A hypothesis test has the following possible outcomes. The outcomes and the probability of each outcome are:

Decision	Null hypothesis	
	$H_0$ is true	$H_0$ is false
Fail to reject	Correct decision probability = $1 - \alpha$	Type II error probability = $\beta$
Reject	Type I error probability = $\alpha$	Correct decision probability = $1 - \beta$ (Power)

The power of the test is the probability that you will correctly reject the null hypothesis, given that the null hypothesis is false. Use a power analysis to determine the power of an existing test, or to determine the sample size needed to assure a given power.

### When to use power analysis

Use a power analysis when designing an experiment. No data are required but, unless you are conducting a test of proportion, you will need an estimate of variability ( $\sigma$ ).

### Why use a power analysis

A power analysis can help answer questions such as:

- Is the sample size large enough?
- How large a difference can the test detect?
- Is the test powerful enough to give credibility to its conclusions?

For example, how many samples must you collect to determine whether the mean mortgage processing time differs from the competitor's by more than 1 hour?

## Determining power

The company wants to determine the power of the mortgage study. Specifically, they want to know the t-test's likelihood of detecting a difference of 1 hour from the competitor's claim of 6 hours, given a standard deviation of 1.355, an  $\alpha$ -level of 0.05, and a sample size of 6.

If the financial institution's average processing time is statistically different from the competitor's claim by an hour or more, they want the test to detect that difference, which may be practically significant to a customer.

### Values

If you enter multiple values for a parameter, Minitab performs the calculations separately for each combination of values.

### Standard deviation

Because variability in the data affects the test's power, you must provide an estimate of sigma. Use a historical estimate or the standard deviation of a sample.

For the mortgage data, the standard deviation of 1.355 comes from the t-test results.

### 1-Sample t

- 1 Choose **File** ► **New**.
- 2 Choose **Minitab Project**, then click **OK**.
- 3 Choose **Stat** ► **Power and Sample Size** ► **1-Sample t**.
- 4 Complete the dialog box as shown below.

- 5 Click **OK**.

## Interpreting your results

With 6 observations, a standard deviation of 1.355, and an  $\alpha$ -level of 0.05, the power is only 0.3124. Therefore, if  $\mu$  differs by 1 hour from the competitor's claim of 6 hours, the chance of detecting this difference with a sample size of 6 is 31.24%.

In other words, the chance that you will fail to reject the null hypothesis and incorrectly conclude that the mean is not different from 6 hours is 68% ( $1 - 0.3124$ ). Therefore, this test has insufficient power.

### Power and Sample Size

1-Sample t Test

Testing mean = null (versus not = null)

Calculating power for mean = null + difference

Alpha = 0.05 Assumed standard deviation = 1.355

Difference	Sample Size	Power
1	6	0.312412

## Interpreting your results

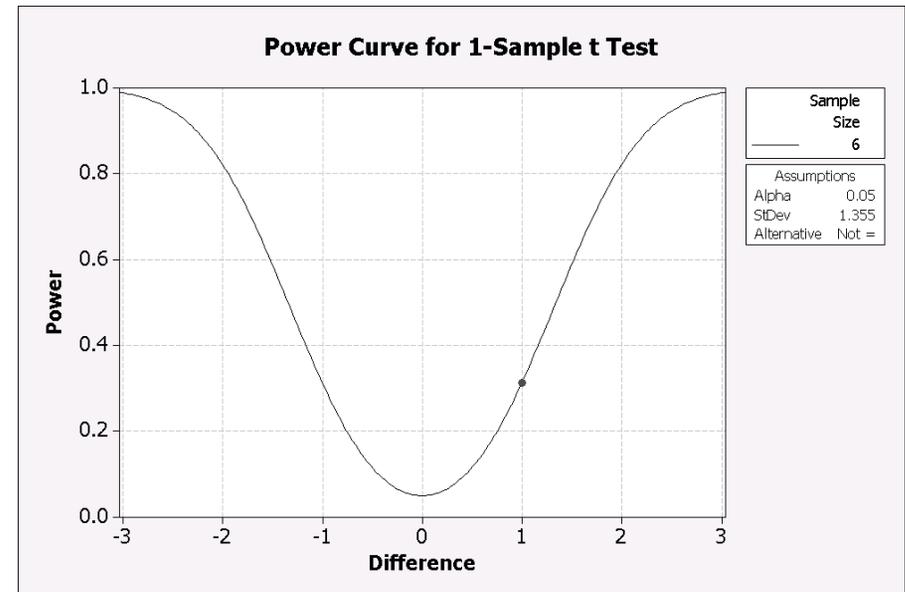
The power curve allows you to see the test's power – the probability of detecting a difference that truly exists – for a variety of differences, and optionally for a variety of sample sizes. The specific differences entered in the dialog box are indicated on the plot with red points.

In this example, Difference represents the difference between the mortgage processing time of the financial service company and its competitor's claim of 6 hours. A negative difference indicates the company's mortgage processing time is less than 6 hours. A positive difference indicates the company's mortgage processing time is greater than 6 hours. When the difference is one hour, the power to detect it with a sample size of 6 is very low.

One way to increase power is to increase sample size.

### What's next

Determine the sample size required to achieve adequate power. How many observations do you need to have an 80% chance of detecting a 1 hour difference? How many observations do you need for an 85%, 90%, or 95% chance of detecting this difference?



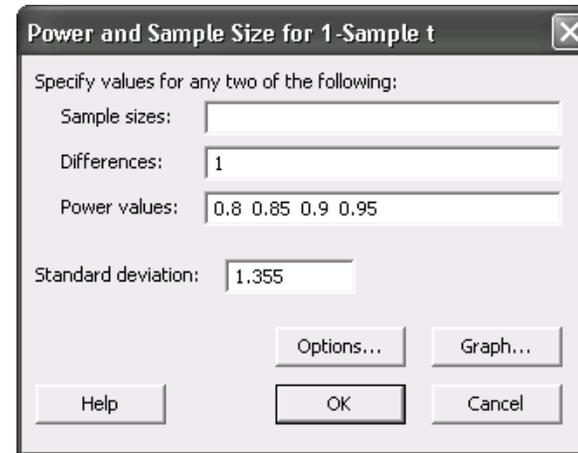
## Determining sample size

With 6 observations, the power of the test was only 0.3124. To have a better chance of detecting a difference of 1 hour, increase the power of the test to at least 0.80 by increasing the sample size.

Calculate the sample sizes required to achieve power levels of 0.80, 0.85, 0.90, and 0.95.

### 1-Sample t

- 1 Choose **Stat** ► **Power and Sample Size** ► **1-Sample t**.
- 2 Complete the dialog box as shown below.



- 3 Click **OK**.

## Interpreting your results

You must have a sample size of 17 to detect a difference of 1 hour with 80% power (Target Power).

Because sample size must be an integer, the Actual Power of the test with 17 observations (0.814922) is slightly greater than the Target Power.

Additional observations give you more power:

- 19 observations yield a power of 86%
- 22 observations yield a power of 91%
- 26 observations yield a power of 95%

As you increase the number of observations, the power of the test increases, and the test is able to detect smaller differences with greater probability.

If the power is very high (say, above 99%) the test may detect tiny shifts that do not have any practical importance, which may be inefficient from a cost perspective.

### Power and Sample Size

1-Sample t Test

Testing mean = null (versus not = null)

Calculating power for mean = null + difference

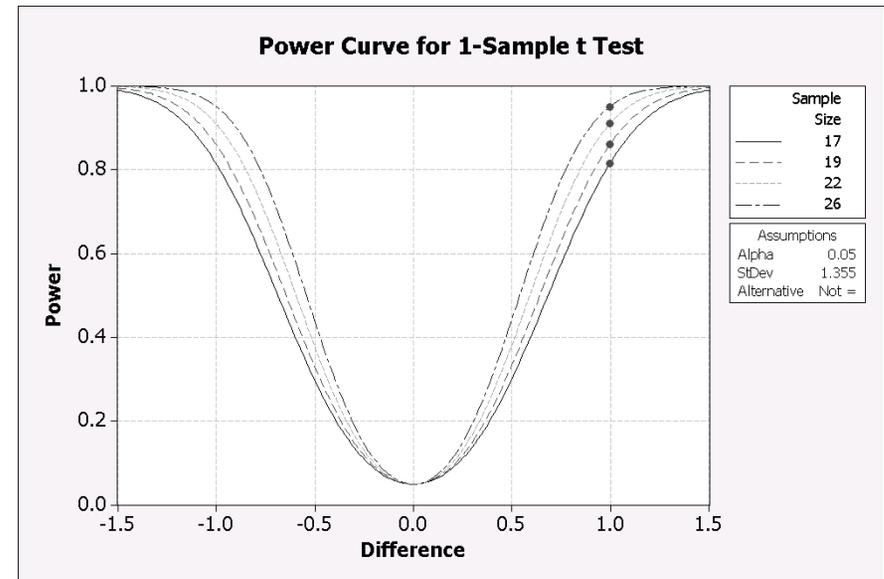
Alpha = 0.05 Assumed standard deviation = 1.355

Difference	Sample Size	Target Power	Actual Power
1	17	0.80	0.814922
1	19	0.85	0.860214
1	22	0.90	0.909789
1	26	0.95	0.951041

## Interpreting your results

The power curve displays the sample sizes required to achieve power levels of 0.80, 0.85, 0.90, and 0.95

Any of these sample sizes will provide adequate power to detect a difference of 1 hour.



## Final considerations

### Summary and conclusions

The t-test of mortgage processing times failed to detect a statistically significant difference between the company's processing time and the competitor's claim of 6 hours. However, this result could occur for two reasons: either no difference exists, or the test simply has too little power to detect a difference because of the small sample size.

Based on the number of observations (6), the difference of interest (1 hour), and the process variability (1.355), the test has a power of only 0.3124. In other words, if a difference of 1 hour truly does exist, this test would only have a 31% chance of detecting it. This power value casts doubt on the test's conclusion that no difference exists.

A larger sample size gives the test more power, enabling it to detect a difference when it truly exists.

### Additional considerations

To ensure that a test has sufficient power, conduct a power analysis prior to collecting data.

To increase the power of a test:

- Increase the sample size
- Decrease the variability that is not attributed to the effect of interest
- Increase  $\alpha$  (although this increases the chance of Type I error)

Higher power means a greater probability of detecting a difference. However, it also increases the chance of detecting small effects that may not be of practical interest. Use process knowledge to determine the smallest difference worth detecting.

# Two-Sample t-Test

## Example 3 Customer Complaints

### Problem

A company wants to compare the mean number of customer complaints on a particular product brand for their two suppliers in Louisiana and Kansas.

### Data collection

The total number of customer complaints is recorded daily at each of the two locations.

### Tools

- 2-Sample t
- Graphical Summary
- 2 Variances
- Time Series Plot

### Data set

Complaints.MPJ

Variable	Description
Date	Date that customer complaints were recorded
LA	Number of customer complaints for the Louisiana supplier
KS	Number of customer complaints for the Kansas supplier

## Two-sample t-test

### What is a two-sample t-test

A two-sample t-test helps determine whether two population means are different. The test uses the sample standard deviations to estimate  $\sigma$  for each population. If the difference between the sample means is large relative to the estimated variability of the sample means, then the population means are unlikely to be the same.

You can also use an independent two-sample t-test to evaluate whether the means of two independent populations differ by a specific amount.

### When to use a two-sample t-test

Use a two-sample t-test with continuous data from two independent random samples. Samples are independent if observations from one sample are not related to the observations from the other sample.

The test also assumes that the data come from normally distributed populations. However, it is fairly robust to violations of this assumption when the size of both samples is 30 or more (see [1]).

### Why use a two-sample t-test

A two-sample t-test can help answer questions such as:

- Are the products of two suppliers comparable?
- Is one variety of equipment better than another?

For example,

- Is the mean moisture content, which is an indicator of freshness, different for two competing baked goods vendors?
- Is the time to first contact quicker for one hospital emergency room compared to another?

## Conducting the two-sample t-test

The first step of a hypothesis test is to determine the null and alternative hypotheses. The null hypothesis usually specifies that a parameter equals a specific value. For example, the difference in the mean numbers of customer complaints in Louisiana and Kansas equals 0 ( $H_0: \mu_{LA} - \mu_{KS} = 0$ ).

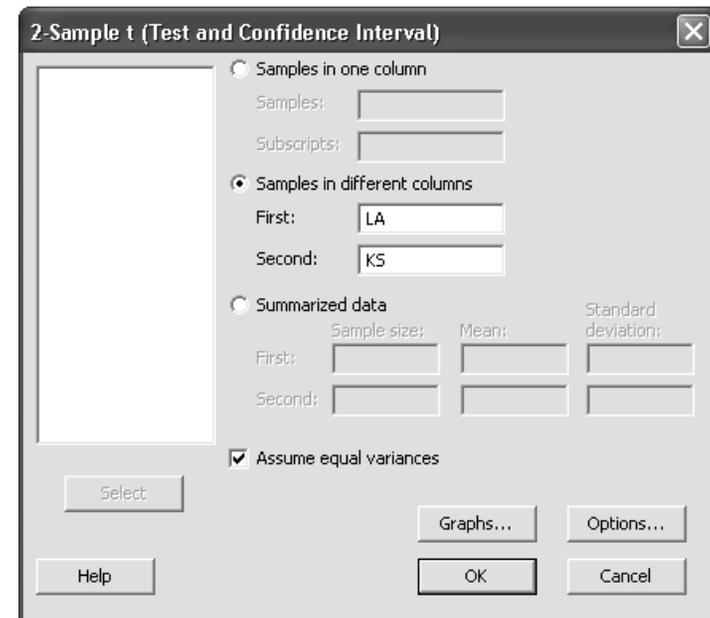
Because you do not suspect beforehand that one supplier has a greater mean number of complaints than the other, a two-tailed test is appropriate.

Therefore, the hypotheses for the test are:

- $H_0: \mu_{LA} - \mu_{KS} = 0$
- $H_1: \mu_{LA} - \mu_{KS} \neq 0$

### 2-Sample t

- 1 Open Complaints.MPJ.
- 2 Choose **Stat** ► **Basic Statistics** ► **2-Sample t**.
- 3 Complete the dialog box as shown below.



- 4 Click **OK**.

## Interpreting your results

Minitab displays the average rating (Mean) and two measures of variability – the standard deviation (StDev) and the standard error of the mean (SE Mean) – for the Louisiana and Kansas suppliers.

### Confidence intervals

The difference between the sample means (3.44) is an estimate of the difference between the population means ( $\mu_{LA} - \mu_{KS}$ ). The confidence interval for the difference is based on this estimate and the variability within the samples.

You can be 95% confident that the difference between the mean number of complaints is between 0.03 and 6.84 higher in LA than KS.

### T-value and p-value

The t-value for the test is 2.01, which is associated with a p-value of 0.048.

Thus, you can reject the null hypothesis at the  $\alpha = 0.05$  level, and conclude that there is a statistically significant difference between the mean number of complaints in LA than KS.

### What's next

Test that the two populations are normally distributed.

### Two-Sample T-Test and CI: LA, KS

Two-sample T for LA vs KS

	N	Mean	StDev	SE Mean
LA	39	50.69	7.62	1.2
KS	39	47.26	7.49	1.2

Difference = mu (LA) - mu (KS)

Estimate for difference: 3.44

95% CI for difference: (0.03, 6.84)

T-Test of difference = 0 (vs not =): T-Value = 2.01 P-Value = 0.048 DF = 76

Both use Pooled StDev = 7.5531

## Testing the normality assumption

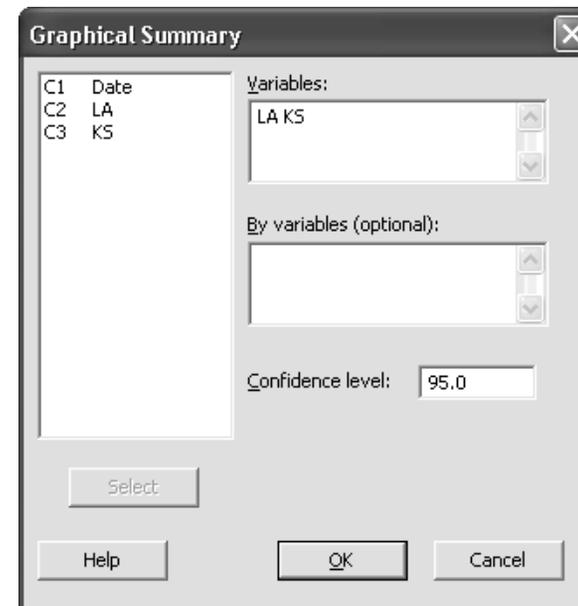
Because the t-test assumes normality for each population, use the Anderson-Darling normality test to determine whether this assumption is valid for these data.

The hypotheses for the Anderson-Darling normality test are:

- $H_0$ : Data are from a normally distributed population
- $H_1$ : Data are not from a normally distributed population

### Normality Test

- 1 Choose **Stat** ► **Basic Statistics** ► **Graphical Summary**.
- 2 Complete the dialog box as shown below.



- 3 Click **OK**.

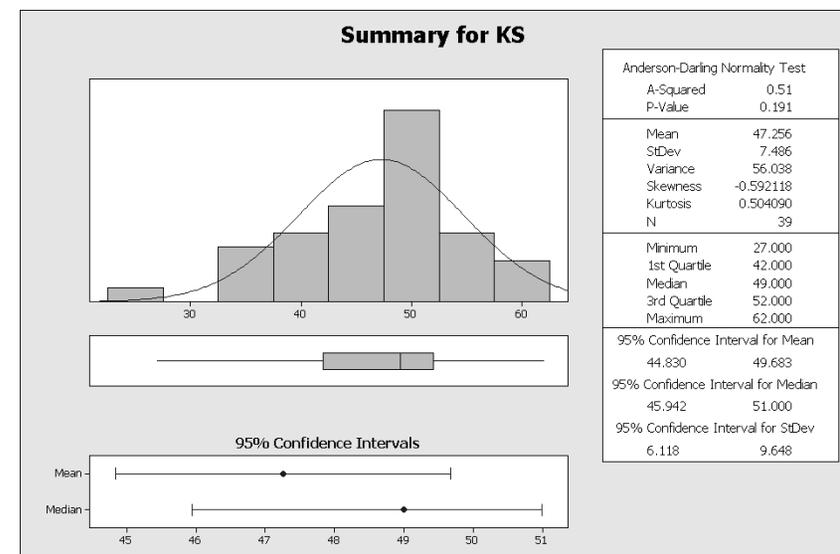
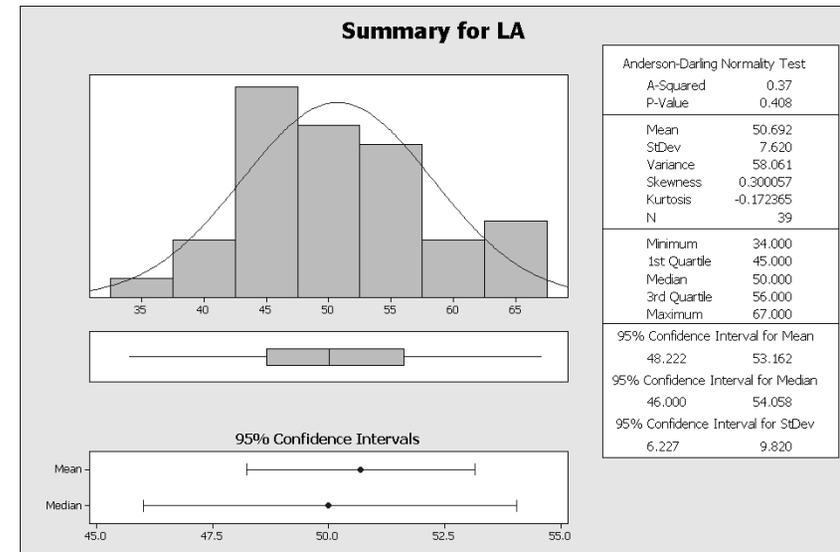
## Interpreting your results

The p-values for the Anderson-Darling normality test are 0.408 and 0.191 for LA and KS respectively. For both locations, you:

- Fail to reject the null hypothesis at the  $\alpha = 0.05$  significance level
- Conclude that there is not enough evidence to conclude the data do not come from normally distributed populations

### What's next

Compare the variability in the two populations.



## Comparing variances

The calculations for the two-sample t-test depend on whether the population standard deviations are the same or different.

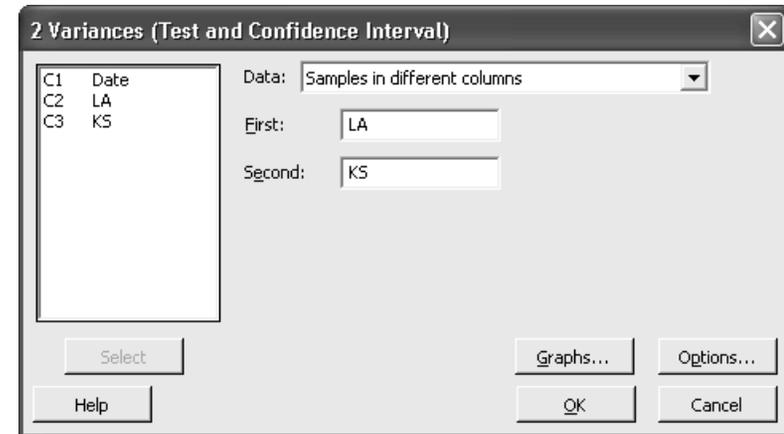
The hypotheses for the test are:

- $H_0: \sigma_1 / \sigma_2 = 1$
- $H_1: \sigma_1 / \sigma_2 \neq 1$

Evaluate the standard deviations to determine whether the ratings from one location vary more than the other.

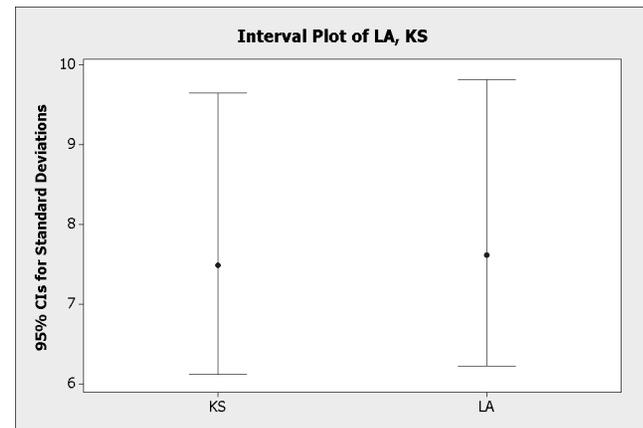
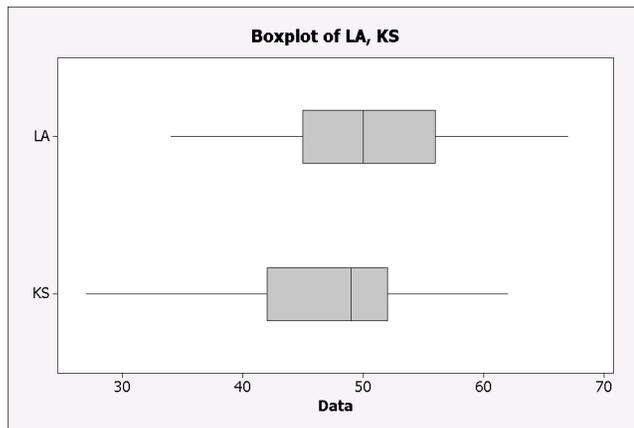
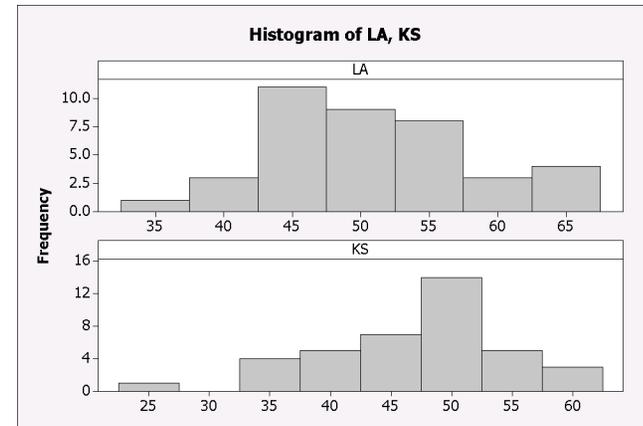
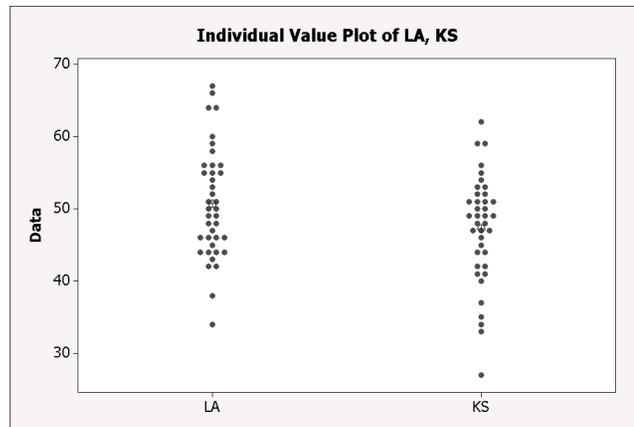
### 2 Variances

- 1 Choose **Stat** ► **Basic Statistics** ► **2 Variances**.
- 2 Complete the dialog box as shown below.



- 3 Click **OK**.

## Interpreting your results



The individual value plot, histogram, boxplot, and interval plot suggest that the variability in complaints for the LA supplier is not different than the variability in complaints for the KS supplier. Examine the Session window output to find out if the variability between suppliers is significantly different.

## Interpreting your results

### Confidence interval

Use confidence intervals to compare the standard deviation ratio for the two populations. The confidence interval for normally distributed data contains 1. Therefore, fail to reject the null hypothesis that the ratio equals 1.

### Standard deviation tests

The results include two separate tests. Which test to use depends on the data:

- If the data are continuous and normally distributed, use the F-test.
- If the data are continuous but not normally distributed, use Levene's test.

### Conclusion

The p-values for both tests are well above  $\alpha$  (0.05), so you fail to reject the null hypothesis that the ratio of the standard deviations is one. The results suggest no difference in the standard deviations of the two locations.

**Note** | Sometimes a higher risk of Type I error, such as  $\alpha = 0.1$ , is chosen to increase the power of the test.

### What's next

Verify that the data are random over time.

### Test and CI for Two Variances: LA, KS

#### Method

Null hypothesis                      Sigma(LA) / Sigma(KS) = 1  
 Alternative hypothesis            Sigma(LA) / Sigma(KS) not = 1  
 Significance level                 Alpha = 0.05

#### Statistics

Variable	N	StDev	Variance
LA	39	7.620	58.061
KS	39	7.486	56.038

Ratio of standard deviations = 1.018  
 Ratio of variances = 1.036

#### 95% Confidence Intervals

Distribution of Data	CI for StDev Ratio	CI for Variance Ratio
Normal	(0.737, 1.406)	(0.543, 1.976)
Continuous	(0.746, 1.639)	(0.557, 2.685)

#### Tests

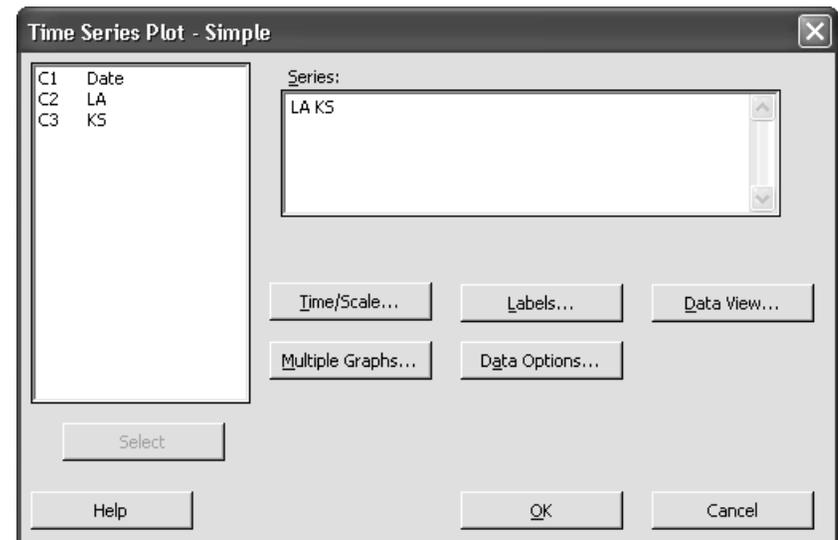
Method	DF1	DF2	Statistic	P-Value
F Test (normal)	38	38	1.04	0.914
Levene's Test (any continuous)	1	76	0.19	0.661

## Assessing the randomness assumption

Use a time series plot to look for trends or patterns in your data, which may indicate that your data are not random or are biased in some way.

### Time Series Plot

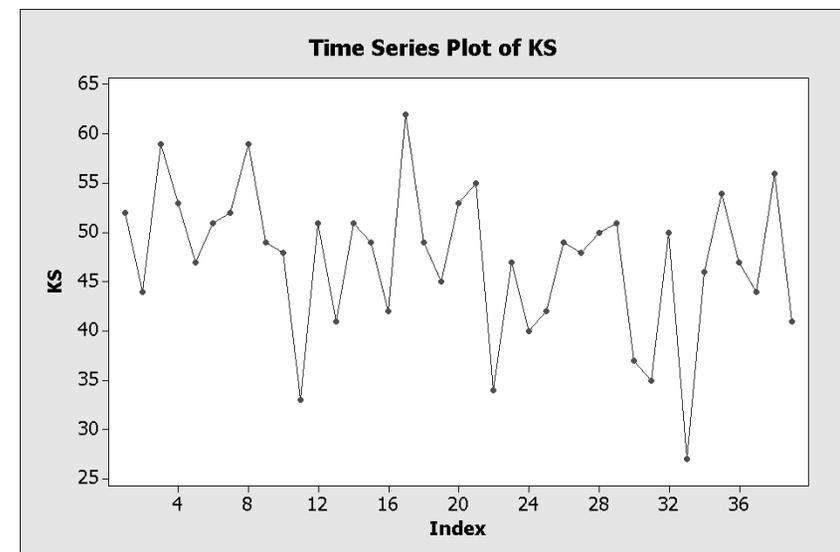
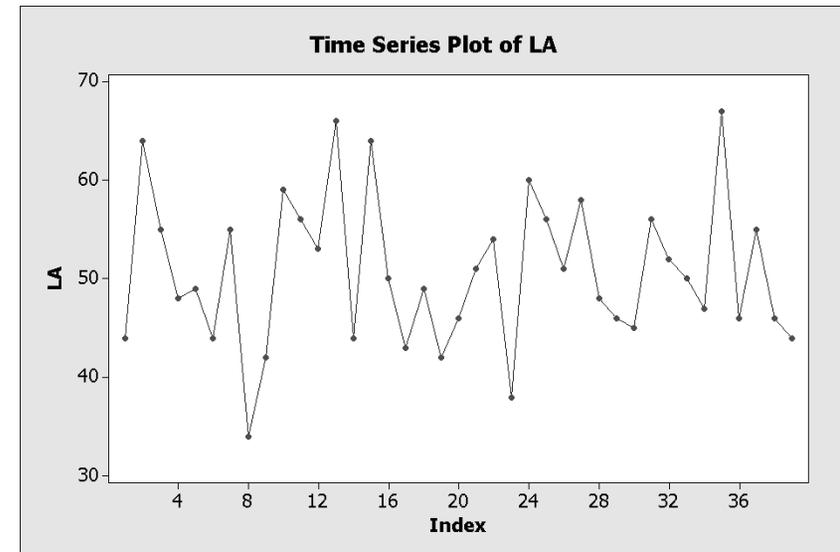
- 1 Choose **Graph** ► **Time Series Plot**.
- 2 Choose **Simple**, then click **OK**.
- 3 Complete the dialog box as shown below.



- 4 Click **OK**.

## Interpreting your results

The results shown in the time series plots indicate that the weekly number of customer complaints appear to be random for each location.



## Final considerations

### Summary and conclusions

Although the variability in the number of complaints for the two locations is similar, the average number of complaints registered is significantly higher in LA than KS. The confidence interval provides a reasonable estimate of what the mean difference might be.

Although the difference of 3.44 is statistically significant, it may not be practically important in a business sense.

When using a two-sample t-test:

- Samples must be independent and random.
- Sample data must be continuous.
- Sample data should be normally distributed.

Note that the t-test procedure is fairly robust to violations of the normality assumption, provided that the observations are collected randomly and the data are continuous, unimodal, and reasonably symmetric (see [1]).

## Exercise B Call Center Handling Times

### Problem

The Quality Assurance department at a call center wants to compare the call work times on incoming calls for two operators. Specifically, they want to determine whether the employees differ in their call time variation and call time means.

### Data collection

For each operator, 30 incoming calls are timed. The data are transformed to satisfy the normality assumption.

### Instructions

- 1 Compare the variation for the two operators using a 2-variances test. Can you conclude that one operator has more consistent call work times than the other?
- 2 Compare the mean call work times for the two operators using a 2-sample t-test.
- 3 Check the normality assumption for each operator.

### Data set

CCTimes.MPJ

Variable	Description
Date/Time	Date and time on the incoming call
HTime	Call work times in seconds
Operator	Operator (James or Laura)
LnHTime	Natural log of handling time

**Note** | HTime data are not normal. The natural log transformation has been stored in LnHTime. Use this column for your analysis.

## Exercise C Salary Comparison

### Problem

A company looking to establish a new restaurant is interested in comparing the mean household income of 2 neighborhoods. They will build the restaurant near the neighborhood with the higher mean income. They compare yearly income for households in both neighborhoods.

### Data collection

The company analyzes the mean yearly income for a random sample of households from each neighborhood. 27 samples are collected from the Toftrees neighborhood and 23 samples are collected from Pine Forest.

### Instructions

- 1 Check the assumptions of normality, equal variances, and randomness.
- 2 Compare the mean yearly income for the two neighborhoods using the appropriate test. Does the difference in means appear to be practically significant?
- 3 If you haven't already, create a graphical comparison of the neighborhoods. Is it surprising to observe potential outliers with this type of data?

### Data set

Salary.MPJ

Variable	Description
Neighborhood	Indicates whether the property is located in Toftrees or Pine Forest
Salary	The yearly income of each home

# Paired t-Test

## Example 4 ATM Surrounds

### Problem

A bank wants to increase the number of ATM transactions because it can charge higher fees to advertising firms if it can show that more customers are now using their ATMs. Using a customer survey, analysts determined that shelters around ATMs would entice more customers to use them, especially in the colder areas of the country.

### Data collection

Analysts consider monthly transactions for 100 ATMs before and after the surrounds were installed.

### Tools

- Paired t-test

### Data set

ATM.MPJ

Variable	Description
Terminal	ATM ID number
Before	Number of ATM transactions before installing the surrounds
After	Number of ATM transactions after installing the surrounds
Difference	Difference between the number of ATM transactions before and after installing the surrounds (After – Before)

## Paired t-test

### What is a paired t-test

A paired t-test helps determine whether the mean difference between paired observations is significant. Statistically, it is equivalent to performing a one-sample t-test on the differences between the paired observations. A paired t-test can also be used to evaluate whether the mean difference is equal to a specific value.

Paired observations are related in some way. Examples include:

- Cycle times recorded for the same individual before and after a training session
- Ratings of competing products from a single evaluator

### When to use a paired t-test

Use a paired t-test with a random sample of paired observations.

### Why use a paired t-test

A paired t-test can help answer questions such as:

- Does a training program improve employee effectiveness?
- Do customers prefer one method of service over another?

For example,

- Does training on diagnostic software reduce the cycle time of car repairs?
- Is the taste rating (from 0-100) for one wine different from another's?

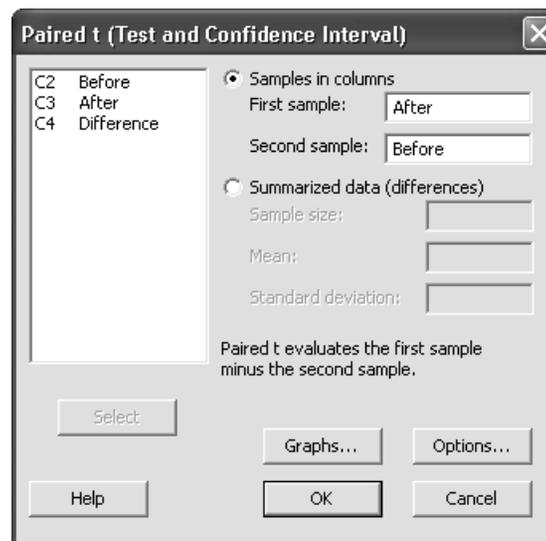
## Conducting the paired t-test

Because the data are paired (the same ATMs are assessed before and after installing the surrounds), use a paired t-test to evaluate the following hypotheses:

- $H_0$ : The mean difference between paired observations in the population is zero.
- $H_1$ : The mean difference between paired observations in the population is not zero.

### Paired t

- 1 Open ATM.MPJ.
- 2 Choose **Stat** ► **Basic Statistics** ► **Paired t**.
- 3 Complete the dialog box as shown below.



- 4 Click **Graphs**.
- 5 Choose **Histogram of differences**.
- 6 Click **OK** in each dialog box.

## Interpreting your results

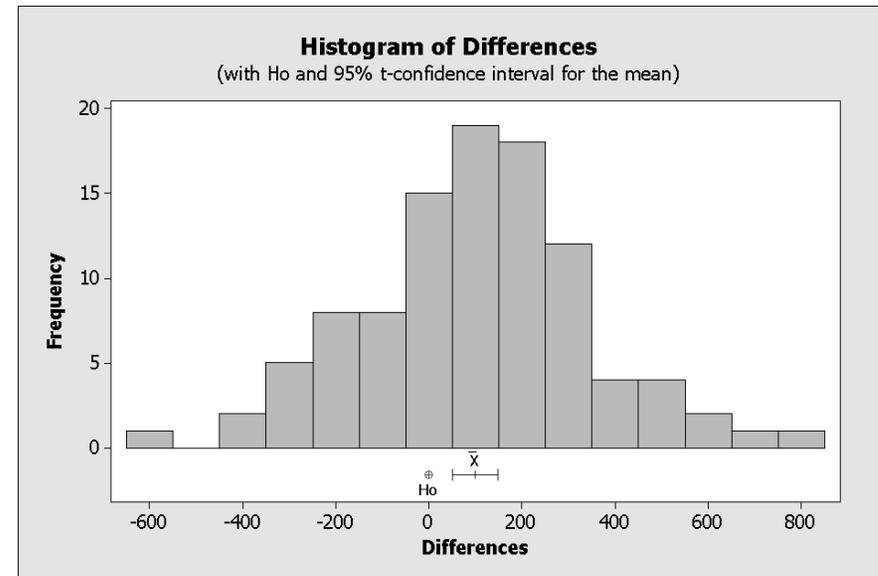
The histogram illustrates the differences between the paired observations.

The observed mean difference of 99.76 transactions is represented by  $\bar{X}$ .  $H_0$  represents the hypothesized difference (zero).

### Confidence interval

Minitab also draws the confidence interval for the population mean difference. If the null hypothesis is true, you expect  $H_0$  to be within this interval.

Because the 95% confidence interval does not include  $H_0$ , you can reject the null hypothesis and conclude that a difference exists in the number of transactions before and after the surround is installed.



## Interpreting your results

### Means

The mean number of transactions is 1372.6 after and 1272.8 before installing the surrounds. The mean difference is about 100 transactions.

The 95% confidence interval is roughly 51 to 148 transactions.

### T-value and p-value

The test gives a t-statistic of 4.06, with an associated p-value of 0.000. Thus, you can reject the null hypothesis at the 0.05  $\alpha$ -level and conclude that the number of transactions increased after the shelters were installed.

### Paired T-Test and CI: After, Before

Paired T for After - Before

	N	Mean	StDev	SE Mean
After	100	1372.6	896.2	89.6
Before	100	1272.8	873.7	87.4
Difference	100	99.8	245.5	24.5

95% CI for mean difference: (51.1, 148.5)

T-Test of mean difference = 0 (vs not = 0): T-Value = 4.06 P-Value = 0.000

## Final considerations

### Summary and conclusions

The confidence interval of the differences in the number of ATM transactions does not contain 0. Therefore, a statistically significant difference exists between the number of transactions before and after the surrounds are installed.

Will advertisers pay more for space at the ATM with an average of 100 more possible viewings per month? Whether the improvement (100 transactions) is of any practical importance is for the bank to decide.

### Additional considerations

When using a paired t-test:

- Observations must be paired
- The data must be continuous
- The differences should be normally distributed

Note that the t-test procedure is fairly robust to violations of the normality assumption, if the pairs of observations are collected randomly and the data are continuous, unimodal, and reasonably symmetric (see [1]).

Using paired observations reduces the variability caused by examining different subjects. By analyzing the differences for each ATM, the paired t-test systematically accounts for this source of variation. In this example, the standard deviation for the paired t-test is 245.5 transactions, whereas the standard deviation for the two-sample t-test would be much higher. Eliminating the variability due to the location of the ATMs from the calculations makes the test more sensitive by increasing its power.

## Exercise D Car Satisfaction Ratings

### Problem

A car company uses customer surveys to collect satisfaction ratings for each car they sell. The company asks customers a series of questions one week after they pick up the car, and they ask the same questions again one year later. The company would like to know if customers are as satisfied as they were when they first purchased the car.

### Data collection

This exercise focuses on the “overall satisfaction” question in the survey: “On a scale of 1-10, where 10 is most satisfied, how satisfied are you with your vehicle?”

108 customer surveys are available for this study. The car company considers a difference of more than 1 unit to be practically significant.

### Instructions

- 1 Use **Calc > Calculator** to create a column of difference values. Then use **Stat > Basic Statistics > Normality Test** to assess the difference values for normality. Use the Kolmogorov-Smirnov test, which is an alternative to the Anderson-Darling test, and may be more appropriate with the large sample size.
- 2 Use a paired t-test to determine whether the scores in week 52 are significantly different from the scores the

same customers gave in week 1. Does the mean difference indicate a change in satisfaction?

- 3 Is the change practically significant to the car company?

### Data set

AutoRatings.MPJ

Variable	Description
Customer ID	ID assigned to each customer
Week 1	The overall satisfaction score at week 1
Week 52	The overall satisfaction score at week 52.

