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 average processing time. They also want to determine if their average 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**

Data set

MORTGAGE.MPJ

Variable	Description
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 μ (the population mean) is equal to a hypothesized value (the test mean).

The test uses the standard deviation of the sample to estimate σ (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 μ 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 μ (mu).

Statistical hypotheses

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

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

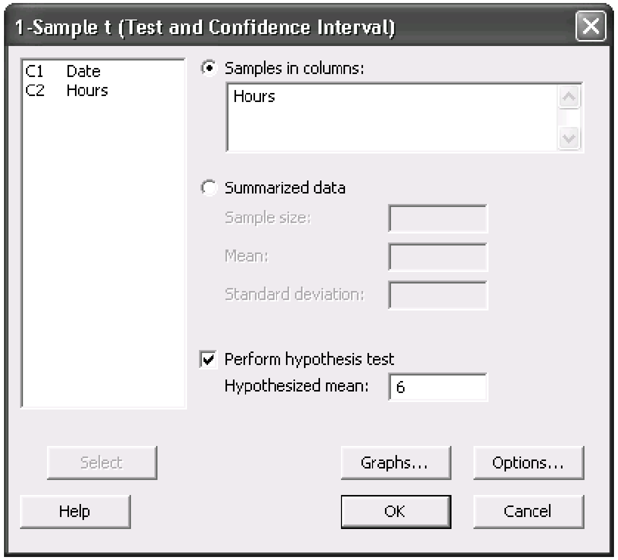
Because the analysts will not measure every loan request in the population, they will not know the true value of μ. However, an appropriate hypothesis test can help them make an informed decision. For these data, the appropriate test is a 1-sample t-test.

1-Sample t

1 Open MORTGAGE.MPJ.

2 Choose **Stat ➤ Basic Statistics ➤ 1-Sample t**.

3 Complete the dialog box as shown below.



4 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 μ 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 = (sample mean – test mean) / 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, α (alpha), before the test:

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

A typical value for α 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 α-level of 0.05 for the mortgage data, not enough evidence is available to reject H_0. P (0.081) is greater than α.

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

What's next

Check the assumption of normality.

Testing the assumption of normality

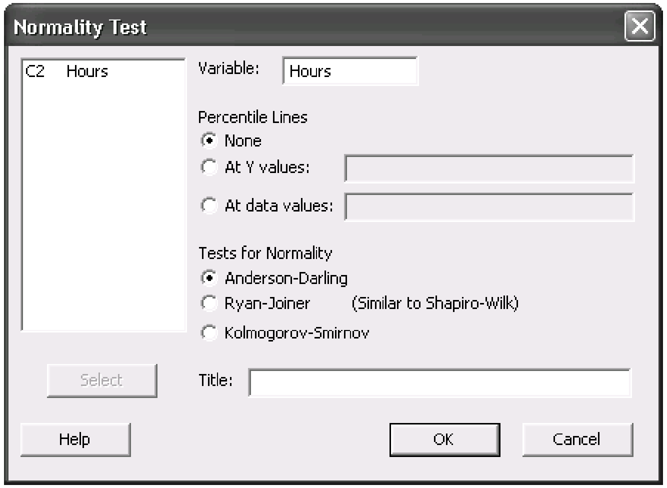
The 1-sample t-test assumes the data are sampled from a normally distributed population.

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.

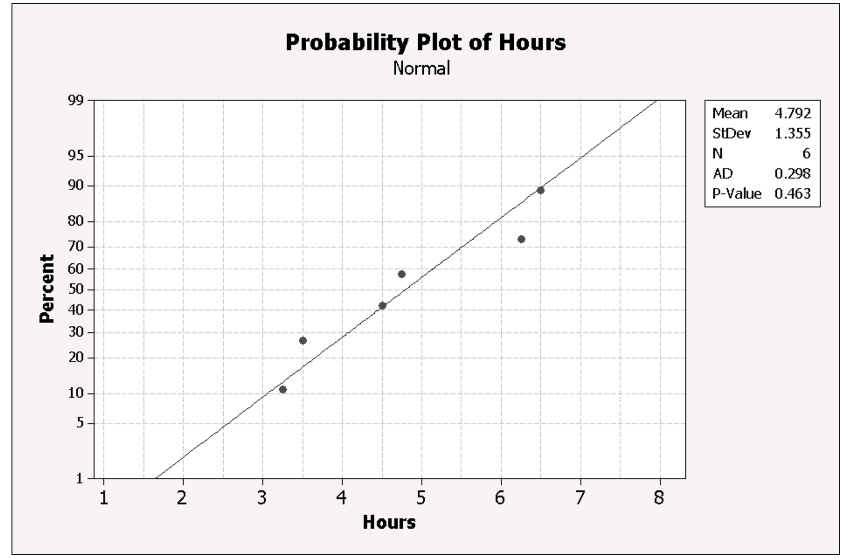


3 Click **OK**.

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 α-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

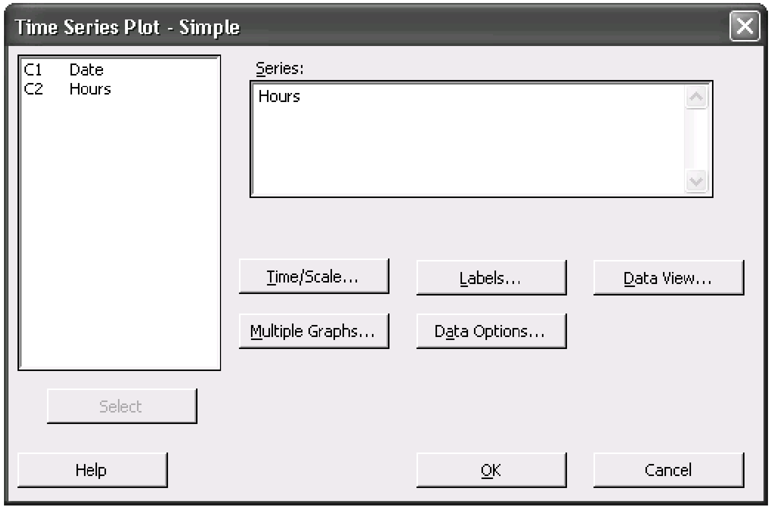
Check the data for non-random patterns over time.

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

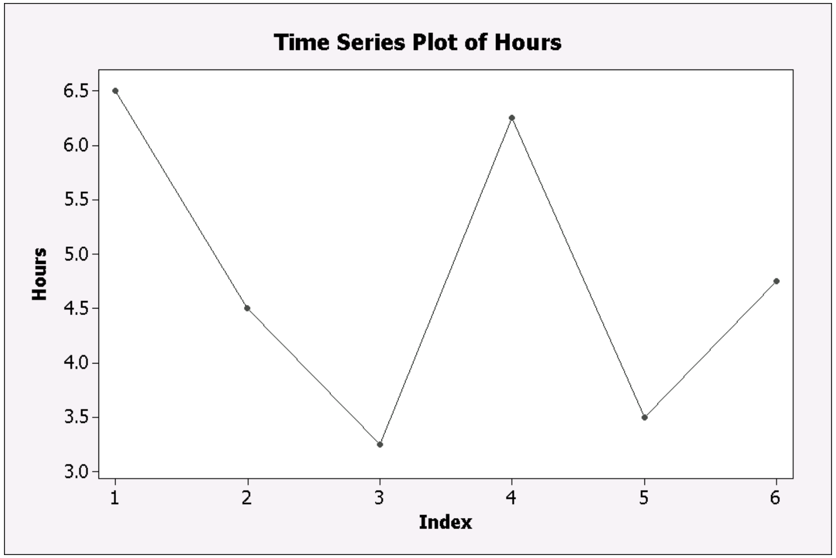
1. Choose **Graph ➤ Time Series Plot**.
2. Choose **Simple**, then click **OK**.
3. Complete the dialog box as shown below.



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 μ) that is based on sample data. For example, with a 95% confidence interval for μ, you can be 95% confident that the interval contains μ (in other words, 95 out of 100 intervals will contain μ 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 μ.

Why use a confidence interval

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

- Is μ on target?
- How much error exists in an estimate of μ?
- How low or high might μ 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 3.37 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 α 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 α-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 α-level *before* conducting the test.

- Increasing α 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 α 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 μ (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 α-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 α-level.