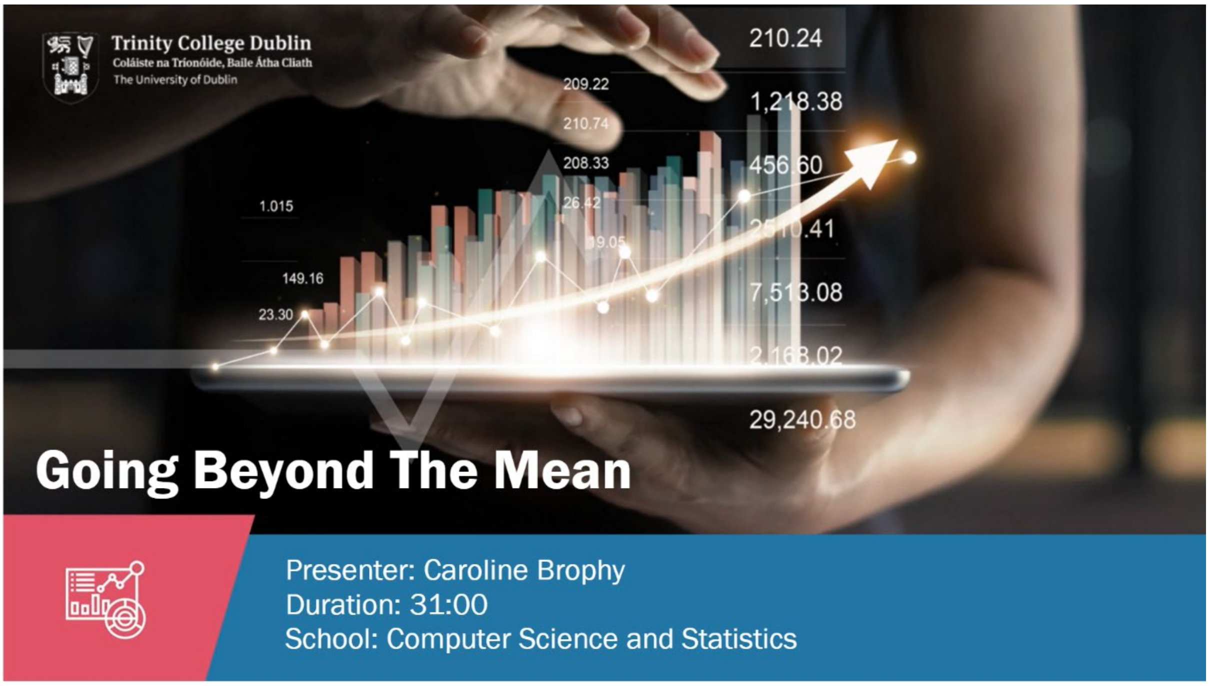
Going Beyond the Mean

Contents

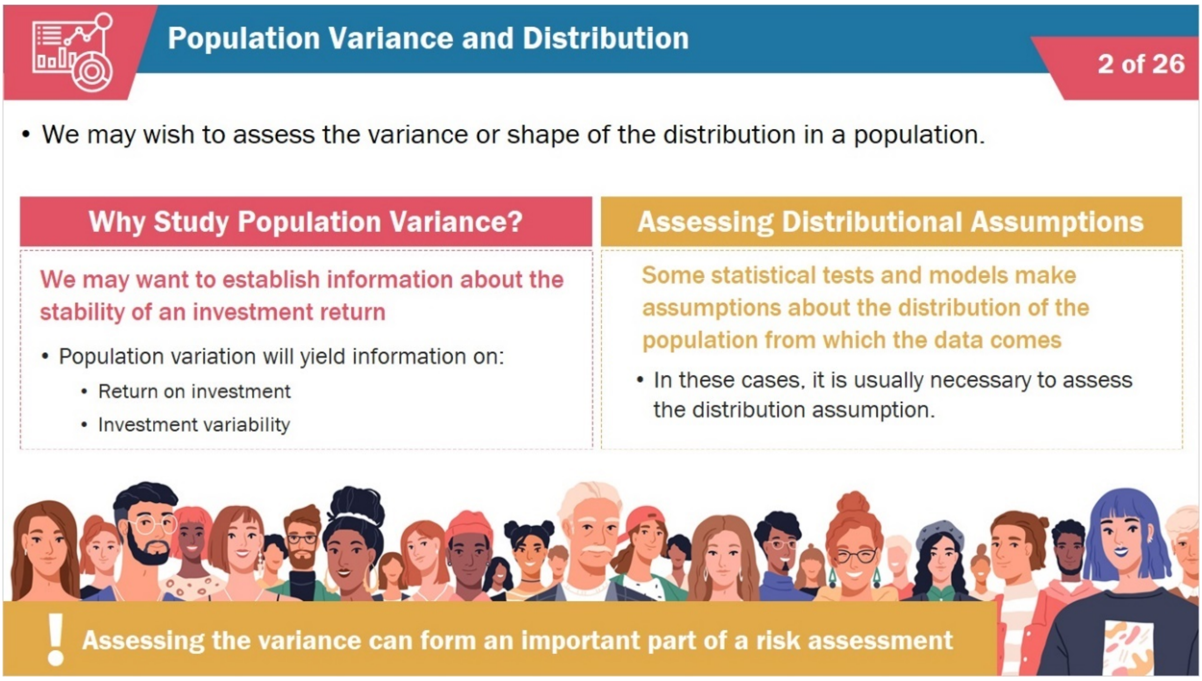
Slide 1: Introduction



Welcome to this session titled ‘Going Beyond the Mean’. My name is Caroline Brophy.

In this session we will explore statistical methods for testing single variances and comparing two variances; we will explore graphical methods for assessing if continuous data is normally distributed; and we will examine a goodness-of-fit test for categorical data.

Slide 2: Population Variance and Distribution



In previous sessions, we have learned how to conduct hypothesis tests about the mean of a population or a population proportion.

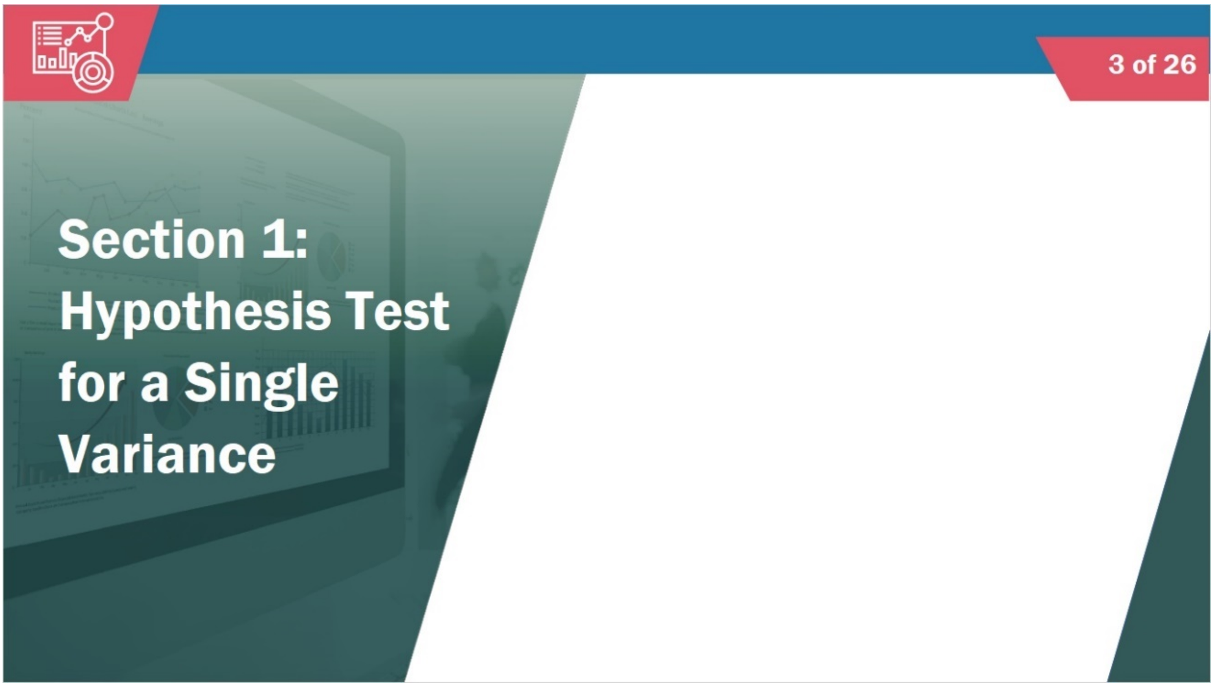
However, we may also be interested in other aspects of a population, for example the population variance or the distribution of the population.

When might we be interested in studying a population variance? When we are trying to establish information about the stability of an investment return. Consider an example where two financial investment options are being compared. As an investor, you will of course be interested in the average return on your investment, but you will also be interested in the variability. This will give you information about the stability of your investment return.

Suppose two investments yield similar returns on average, but one seemed to be more variable than the other. As an investor, you would want to know about this variability because the more variable option might give you a higher return, but of course it comes with more risk of a lower return than the stable investment option. Assessing the variance could be an important part of the risk assessment.

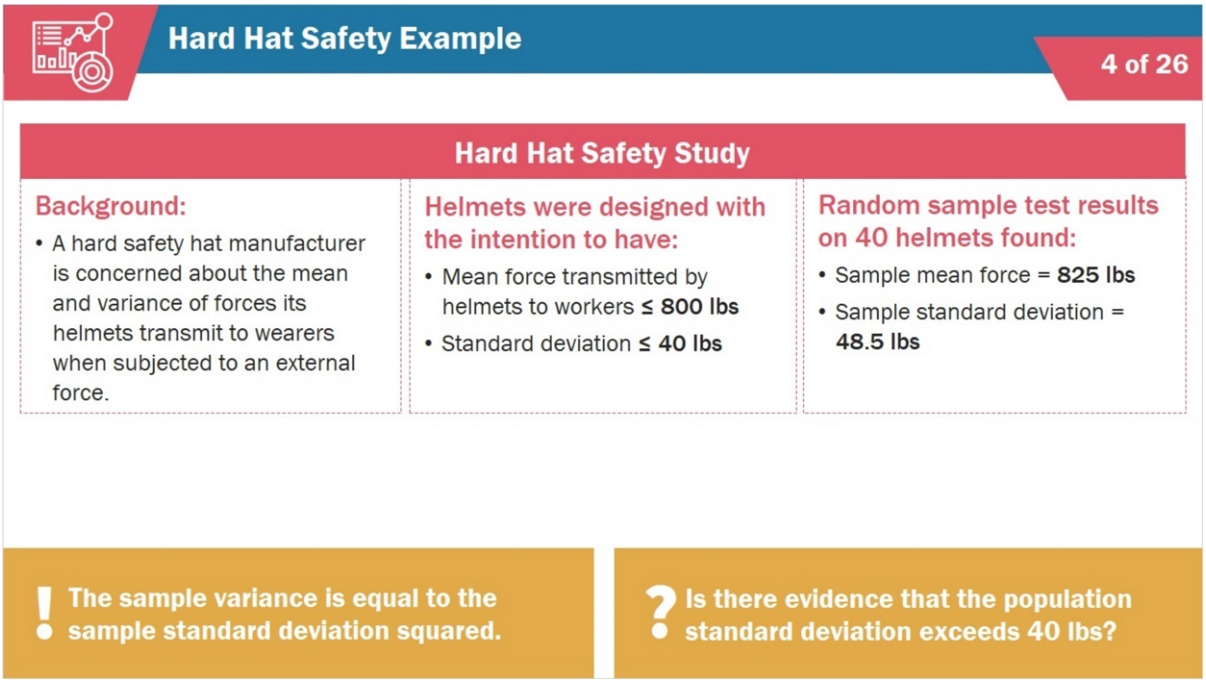
In previous sessions, we have talked about assumptions for statistical tests that we have performed. Many statistical tests and models make assumptions about the distribution of the population from which the data comes. In such cases, it can be very useful to have tools to assess or test the distribution assumption. We will explore such tools in this session.

Slide 3: Section 1: Hypothesis Test for a Single Variance



We will first explore how to do a hypothesis test for a single variance parameter.

Slide 4: Hard Hat Safety Example



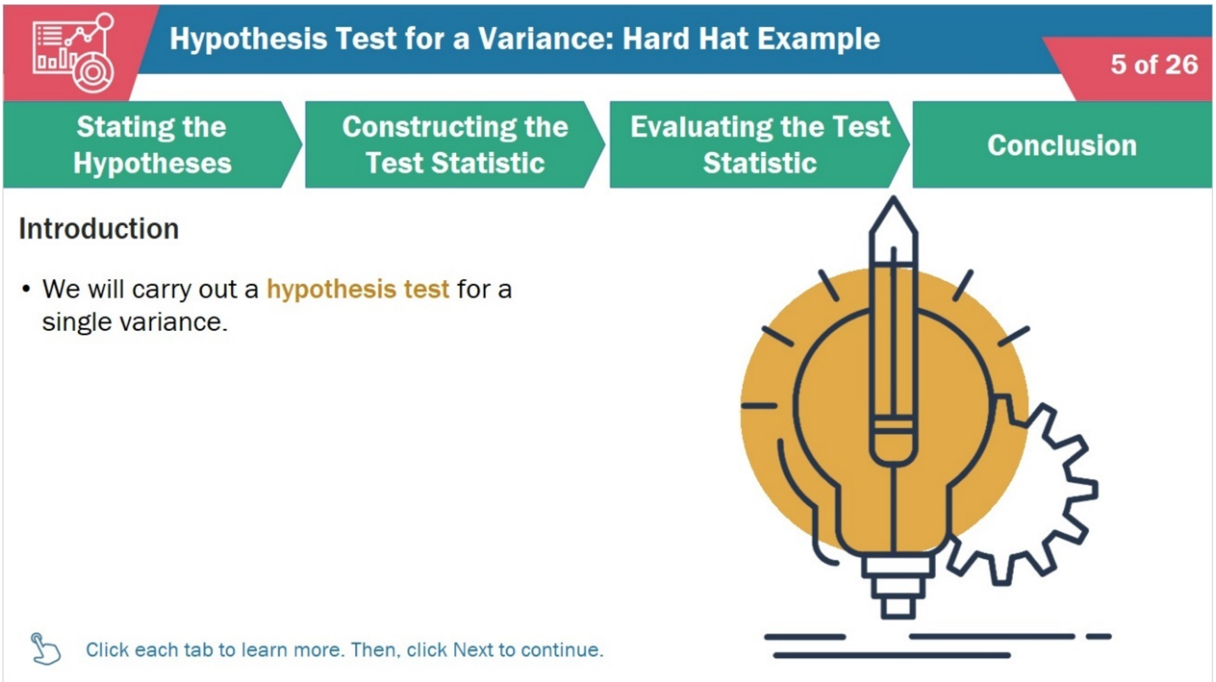
Let's start with a motivating example about the safety of hard hats.

The study was carried out as follows.

A manufacturer of hard safety hats for construction workers is concerned about the mean and the variance of the forces its helmet transmits to wearers when subjected to an external force. The helmets are designed so that the mean force transmitted by the helmets to the workers is 800 pounds (or less) with a standard deviation to be less than 40 pounds. Tests were run on a random sample of 40 helmets, and the sample mean force and sample standard deviation were found to be 825 pounds and 48.5 pounds, respectively.

We will focus on the variance parameter here. Remember that the sample variance is equal to the sample standard deviation squared. Is there evidence that the population standard deviation exceeds 40 pounds? That is the same as asking if there is evidence that the population variance exceeds 1600, which is 40 squared.

Slide 5: Hypothesis Test for a Variance: Hard Hat Example

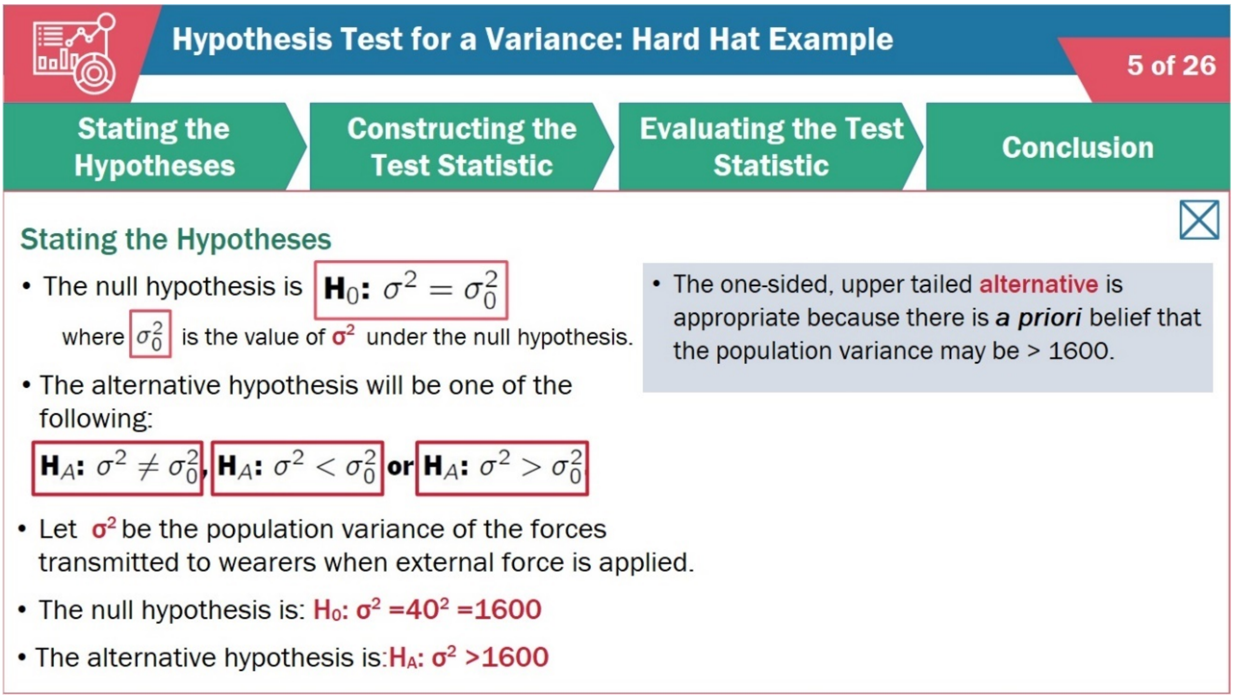


We will carry out a hypothesis test for a single variance.

By now, you will be familiar with the general steps required to perform a hypothesis test. We specify the hypotheses, construct the test statistic, evaluate the test statistic against the reference or null distribution and make a conclusion.

Click each tab to learn more. When you are ready, click next to continue.

Tab 1: Stating the Hypotheses



When testing a single variance, the null hypothesis is that sigma squared is equal to sigma 0 squared. Where sigma 0 squared is the value of sigma squared under the null hypothesis. Remember that when we say, “under the null hypothesis”, we mean assuming that the null hypothesis is true.

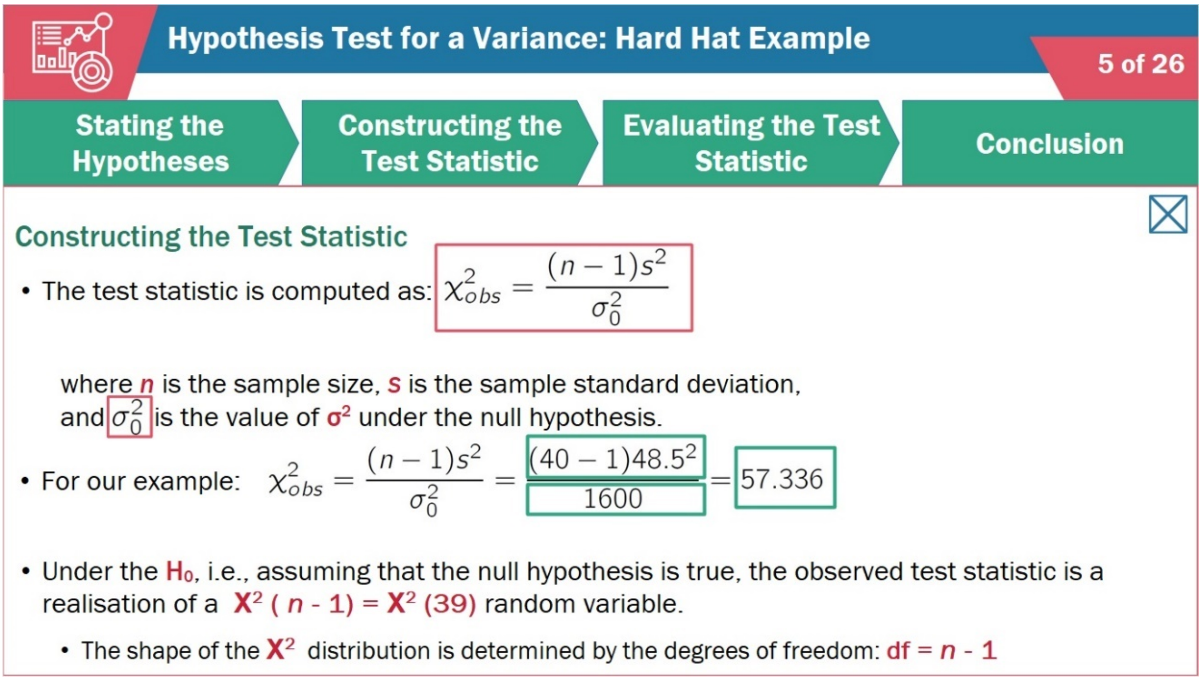
The alternative hypothesis can be one-sided or two-sided.

- The alternative hypothesis that sigma squared is not equal to sigma 0 squared is the two-sided alternative.
- The alternative hypothesis that sigma squared is less than sigma 0 squared is the one-sided alternative, or the lower-tailed alternative since we are testing “less than”.
- The alternative hypothesis that sigma squared is greater than sigma 0 squared is the one-sided alternative, or the upper-tailed alternative since we are testing “greater than”.

For our example, let sigma squared be the population variance of the forces transmitted to wearers when external force is applied.

Our null hypothesis is that sigma squared is equal to 1600, versus the alternative hypothesis that sigma squared is greater than 1600. We choose the one-sided upper tailed alternative because that is what we were asked about in the question: there was *a priori* belief that the population variance may be greater than 1600. Remember that we do a two-sided test by default, but in this example the one-sided alternative is appropriate.

Tab 2: Constructing the Test Statistic



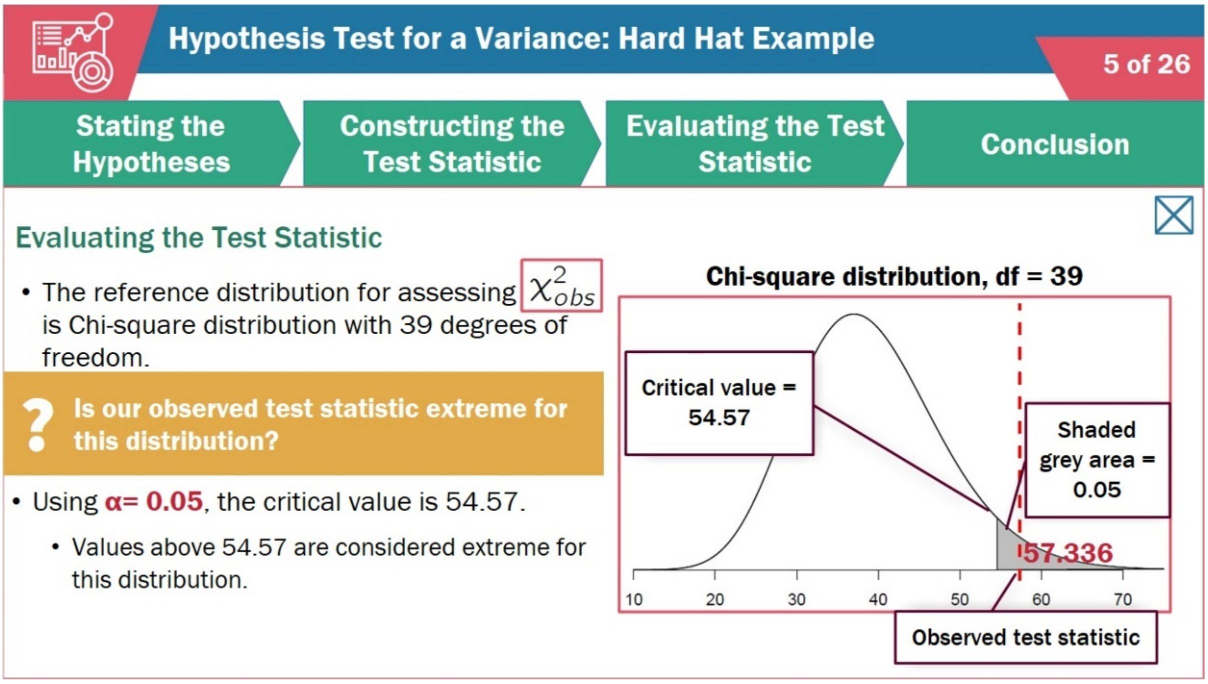
The formula for our test statistic is n – 1 multiplied by s squared, divided by sigma 0 squared, where n is the sample size, s is the sample standard deviation and sigma 0 squared is the value of sigma squared under the null hypothesis.

In our example, n, the sample size is 40, s, the sample standard deviation is 48.5 and under the null hypothesis, sigma squared is equal to 1600. We plug in our values and find that our observed test statistic is equal to 57.336.

If the null hypothesis is true, the observed test statistic is a realisation of a chi-square random variable with 39 degrees of freedom. The degrees of freedom are found by subtracting 1 from the sample size, here this is 40 – 1 = 39.

The shape of the chi-square distribution is determined by the degrees of freedom.

Tab 3: Evaluating the Test Statistic



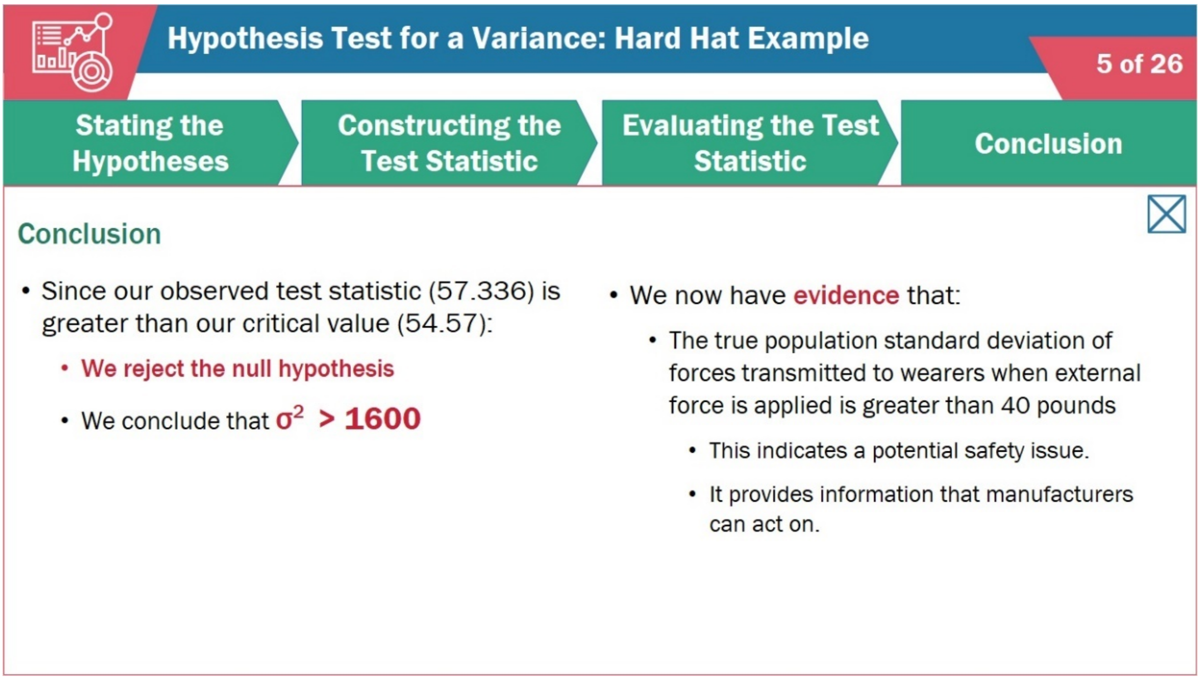
The reference distribution for assessing the test statistic is the Chi-square distribution with n – 1 = 39 degrees of freedom. We want to know if our observed test statistic is extreme for this reference distribution, or not. We will use alpha = 0.05.

You can see the shape of the Chi-square distribution with 39 degrees of freedom in the graph. The tail on the right-hand side is little bit longer than the tail on the left-hand side.

The shaded grey area on the graph marks the area under the curve greater than 54.57; this shaded grey area is equal to 5% of the total area under the curve. Values above 54.57 are assumed to be extreme for this distribution, when using alpha = 0.05.

Our observed test statistic, which was equal to 57.336, is marked on the graph by a red line. Clearly 57.336 is greater than 54.57 and so our test statistic is extreme for this distribution. In this case, we reject the null hypothesis and accept the alternative hypothesis.

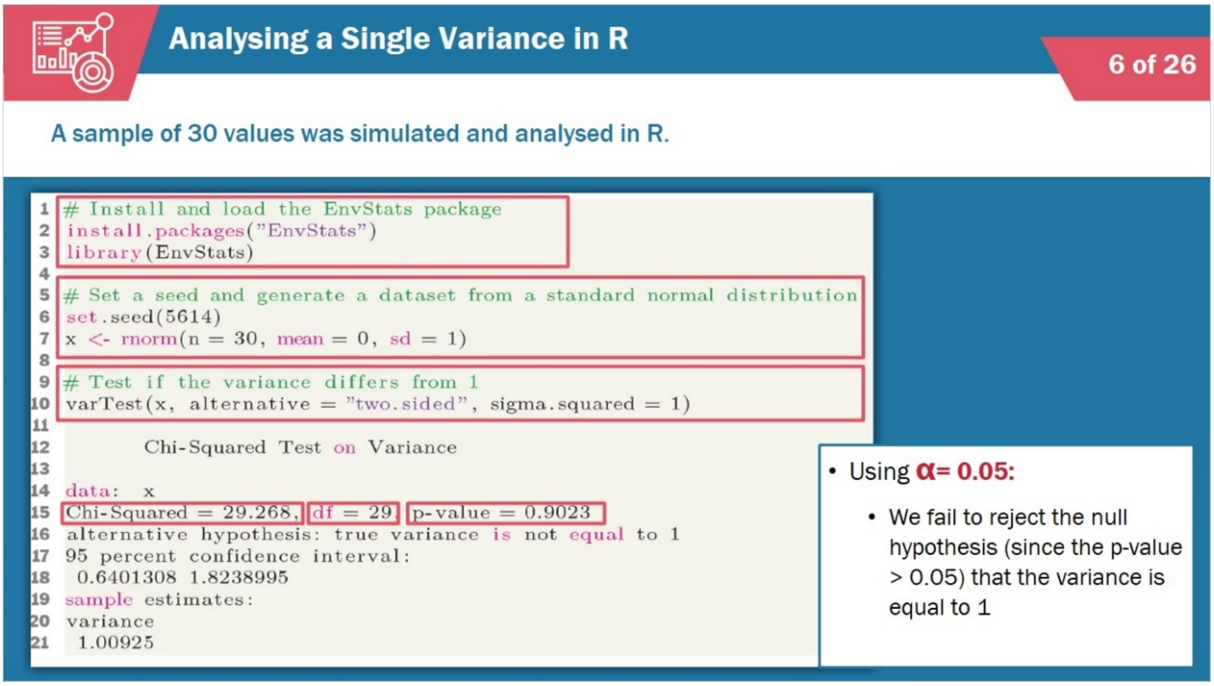
Tab 4: Conclusion



Our observed test statistic of 57.336 is greater than our critical value 54.57, therefore we reject the null hypothesis and conclude that sigma squared is greater than 1600.

Going back to the question we asked when we introduced our motivating example, we can now say that the true population standard deviation of forces transmitted to wearers when external force is applied is greater than 40 pounds. This indicates a potential safety issue with the hard hats and will be valuable information that the manufacturers can act on.

Slide 6: Analysing a Single Variance in R



R software can be used to carry out a hypothesis test for a single variance. We will have a look at code for doing this on a dataset of size 30 simulated from a standard normal distribution.

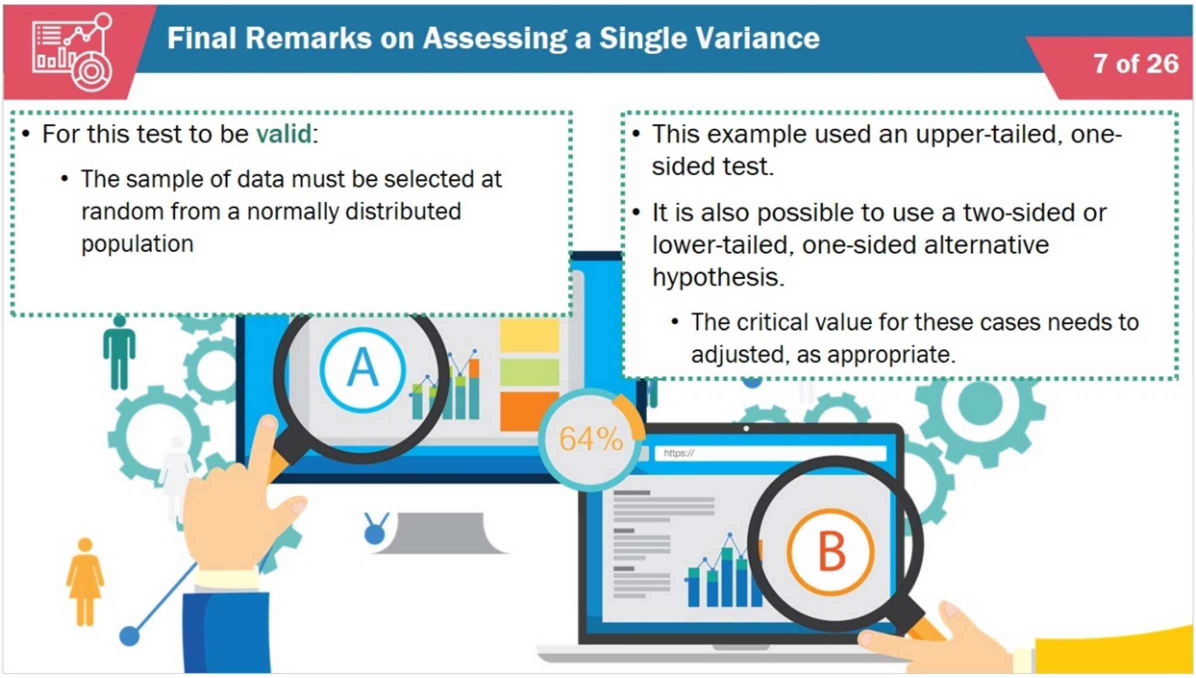
First, install and load the EnvStats package.

Then, generate the data using the 'rnorm' function. From the code here, a vector called 'x' of size 30 from a standard normal distribution will be generated. Recall that the standard normal distribution has mean 0 and variance 1.

We can then use the VarTest function to carry out the hypothesis test. We specify the vector containing the data, which is x, whether we want a two-sided test or not, and the value of sigma squared under the null hypothesis.

The output shows a test statistic of 29.268, degrees of freedom equal to 29 (which is the sample size of 30 – 1) and the p-value of 0.9023. We fail to reject the null hypothesis and have no evidence that the variance differs from 1. Of course, this result was expected since we simulated from a standard normal distribution which has a variance of 1.

Slide 7: Final Remarks on Assessing a Single Variance

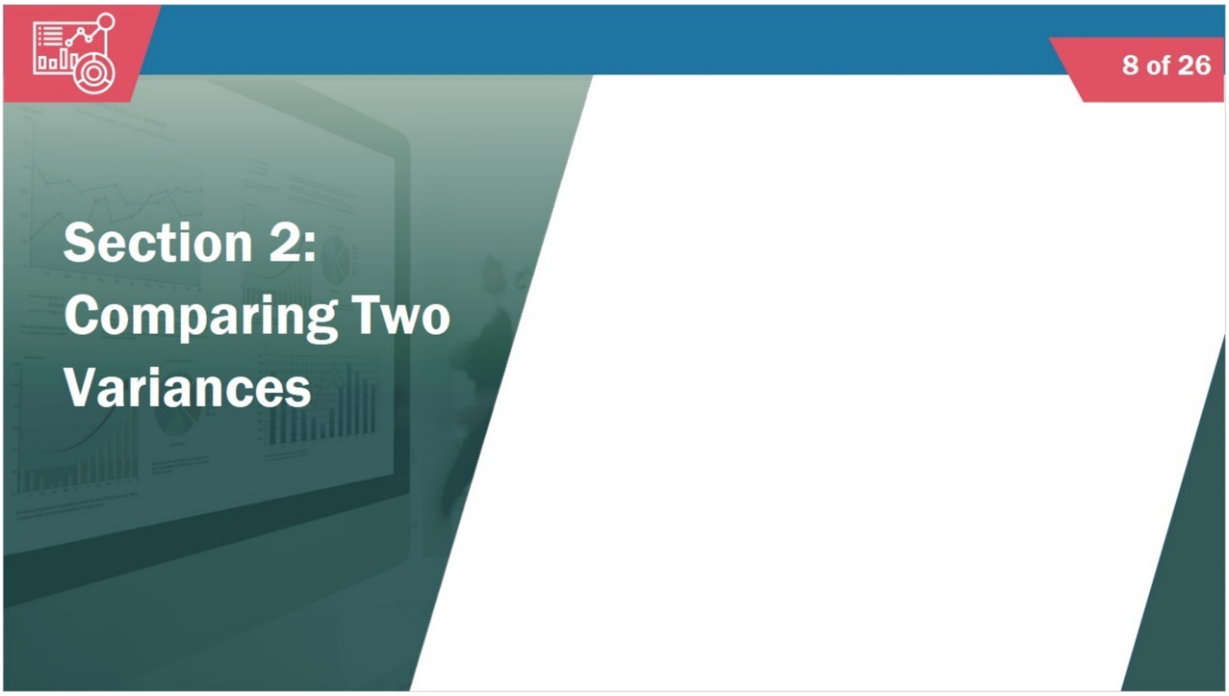


There are two final points to note on this section.

The first point is that for the test we have seen here for a single variance, the sample of data must be selected at random from a normally distributed population. If this is not satisfied, the test may not be valid. It is advisable to check a sample of data to see if it is reasonable to assume normality, before carrying out this test. We will see some ways to do this later in this session.

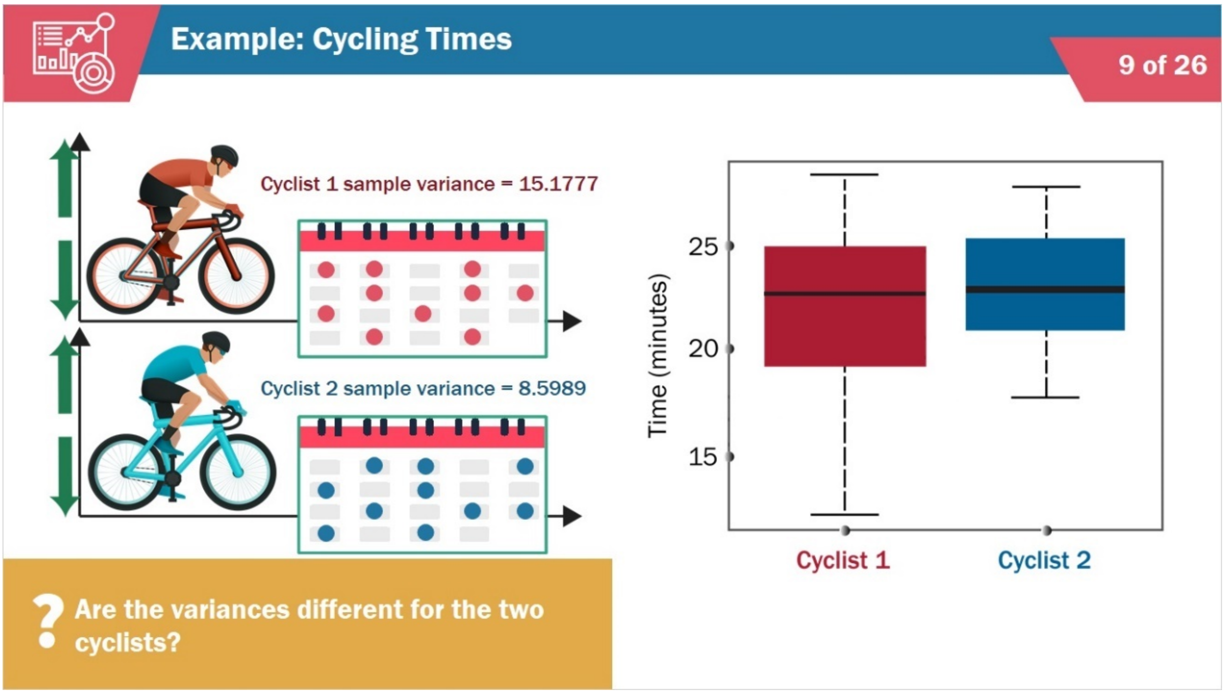
The second point is that we did an upper tailed test in this example, however, in general it is also possible to use a two-sided or lower-tailed one-sided alternative hypothesis. In these cases, the critical value needs to be adjusted as appropriate.

Slide 8: Section 2: Comparing Two Variances



We will now move on to comparing two variances.

Slide 9: **Example: Cycling Times**



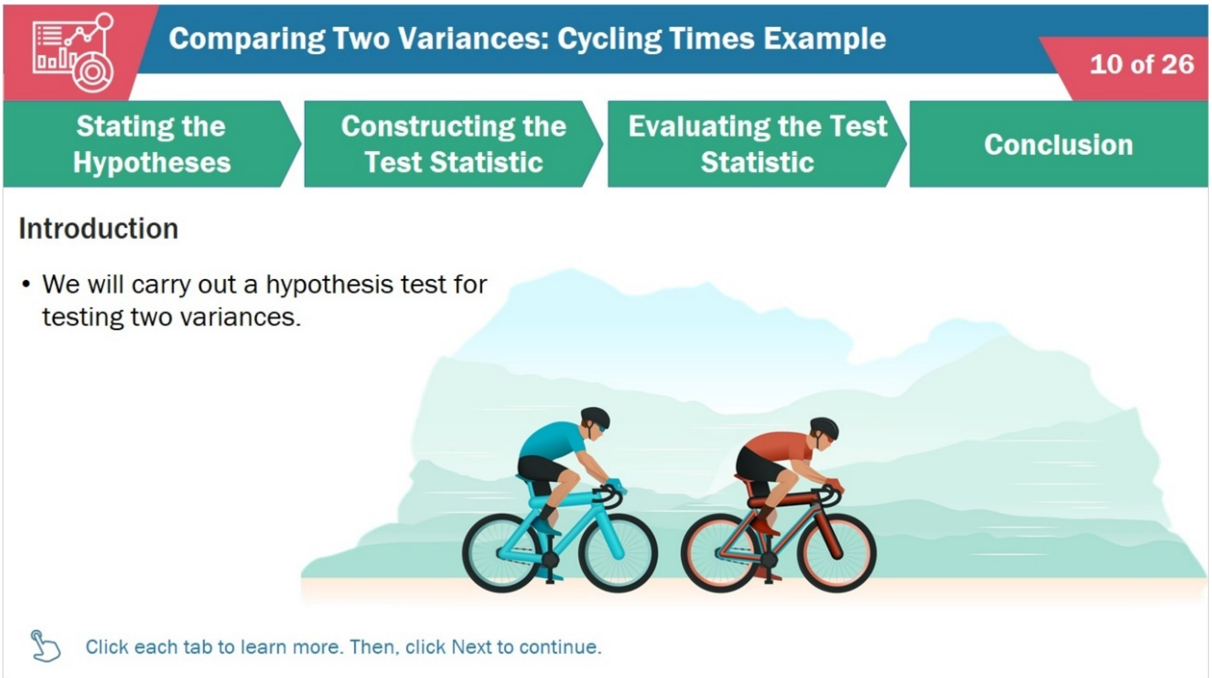
We will start with a motivating example.

A study was carried out as follows.

Two cyclists regularly cycle the same route. They each recorded their journey time on 25 randomly selected days. The first cyclist had a sample variance of 15.1777 and the second cyclist had a sample variance of 8.5989. The two samples of data are displayed in the following boxplots.

In the boxplots, time in minutes is on the y-axis and there is one boxplot for the data from each cyclist. It seems as if the variance of cyclist 1 is slightly larger than cyclist 2. But of course, this difference could just be by chance. We will now formally assess whether the two variances are the same or not.

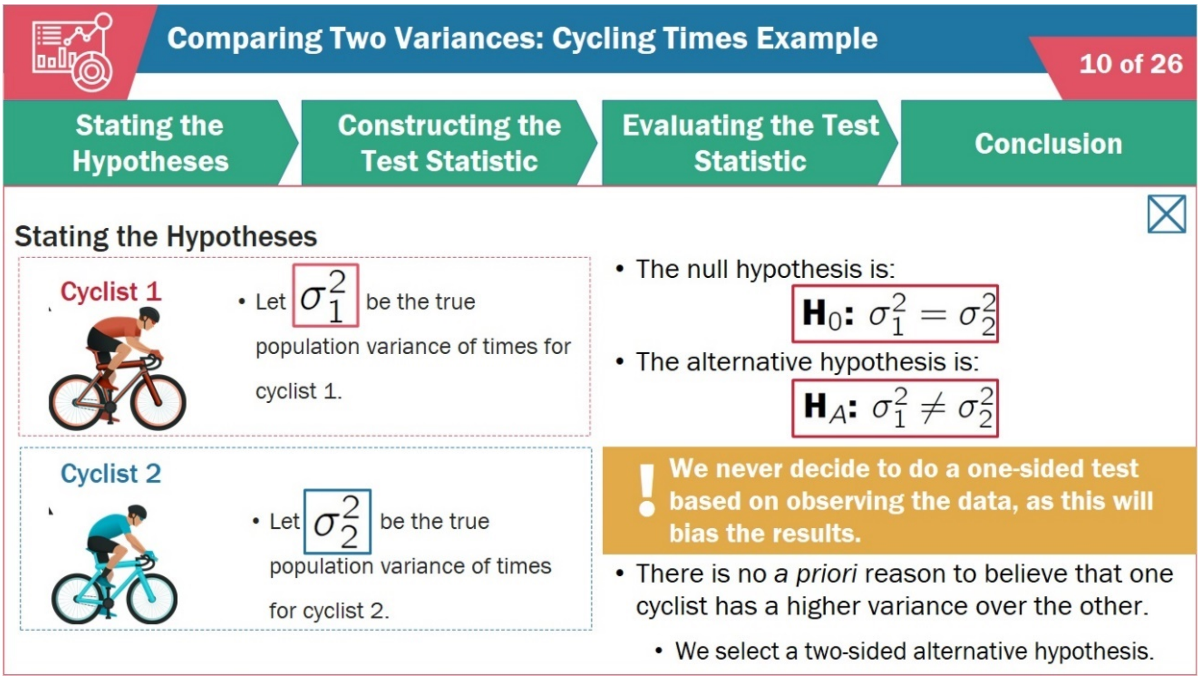
Slide 10: Comparing Two Variances: Cycling Times Example



We will once again follow the same general steps for performing a hypothesis test, this time for testing two variances. We specify the hypotheses, then construct the test statistic, then evaluate the test statistic against the null distribution and make a conclusion.

Click each tab to learn more. When you are ready, click next to continue.

Tab 1: Stating the Hypotheses



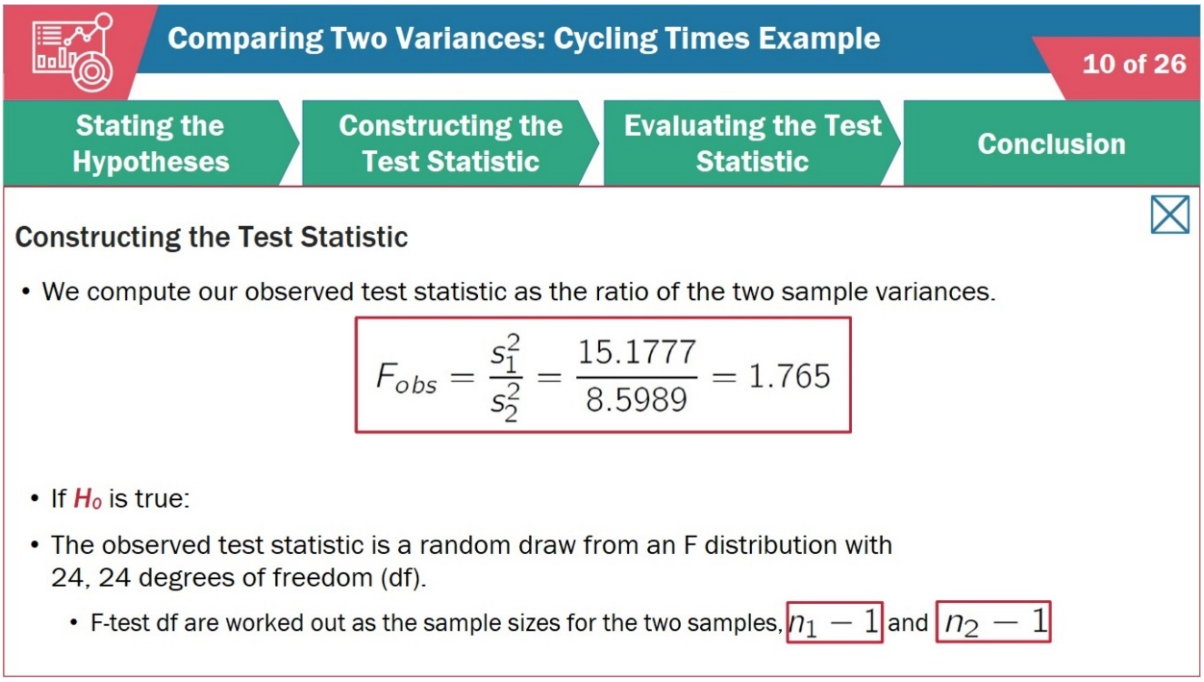
Let sigma 1 squared be the true population variance of times for cyclist 1 and let sigma 2 squared be the true population variance of times for cyclist 2.

The null hypothesis is that sigma 1 squared is equal to sigma 2 squared.

The alternative hypothesis is that sigma 1 squared is not equal to sigma 2 squared.

This is a good example of where it might be tempting to go for a one-sided alternative, since we saw in the boxplots that the values in the sample cyclist 1 were more spread out than the values in the sample for cyclist 2, but that would not be appropriate. Remember, we don't ever decide on the direction of the alternative hypothesis based on the data, to do so would actually bias your results. Here, we had no *a priori* belief that one cyclist had more variable times than the other, so a two-sided test is most appropriate.

Tab 2: Constructing the Test Statistic

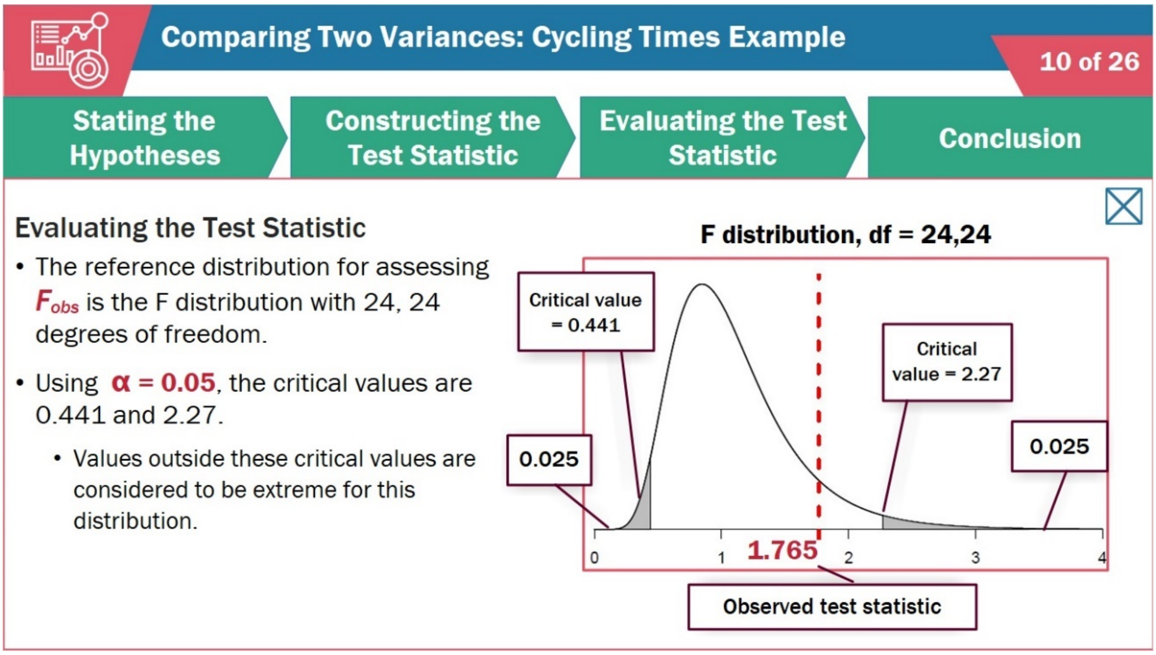


We compute our observed test statistic as the ratio of the two sample variances. The two sample variances are denoted s 1 squared and s 2 squared respectively. Plugging in the values, we find that our test statistic is equal to 15.1777 divided by 8.5989 which is equal to 1.765.

If the null hypothesis is true, the observed test statistic is a random draw from an F distribution with 24 and 24 degrees of freedom.

The degrees of freedom for the F test are worked out as the sample sizes for the two samples each with 1 subtracted.

Tab 3: Evaluating the Test Statistic

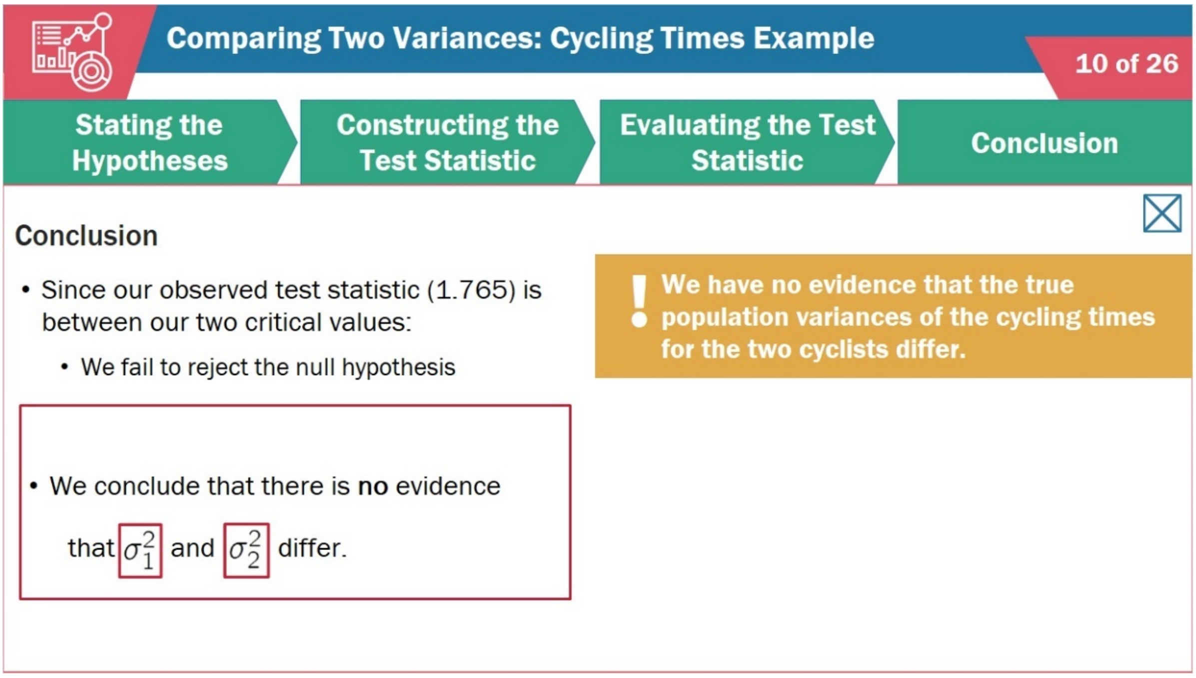


The reference distribution for assessing the F observed test statistic in this example is the F distribution with 24 and 24 degrees of freedom. The shape of the F distribution is determined by its degrees of freedom; you can see the shape of the F distribution with 24 and 24 degrees of freedom in the graph. We use alpha equal to 0.05.

Since we are doing a two-sided test, we have two critical values. Using alpha = 0.05, these are 0.441 and 2.27. The shaded grey area under the curve to the left of 0.441 equals 0.025 and the shaded grey area to the right of 2.27 also equals 0.025. The total area shaded in grey equals 5% of the total area under the curve. Values that lie below 0.441 or above 2.27 are considered to be extreme for this distribution, when using alpha equal to 0.05.

Our observed test statistic is marked in red and equals 1.765. We can see that our test statistic lies between our two critical values and so is not considered to be extreme for this distribution. We fail to reject the null hypothesis.

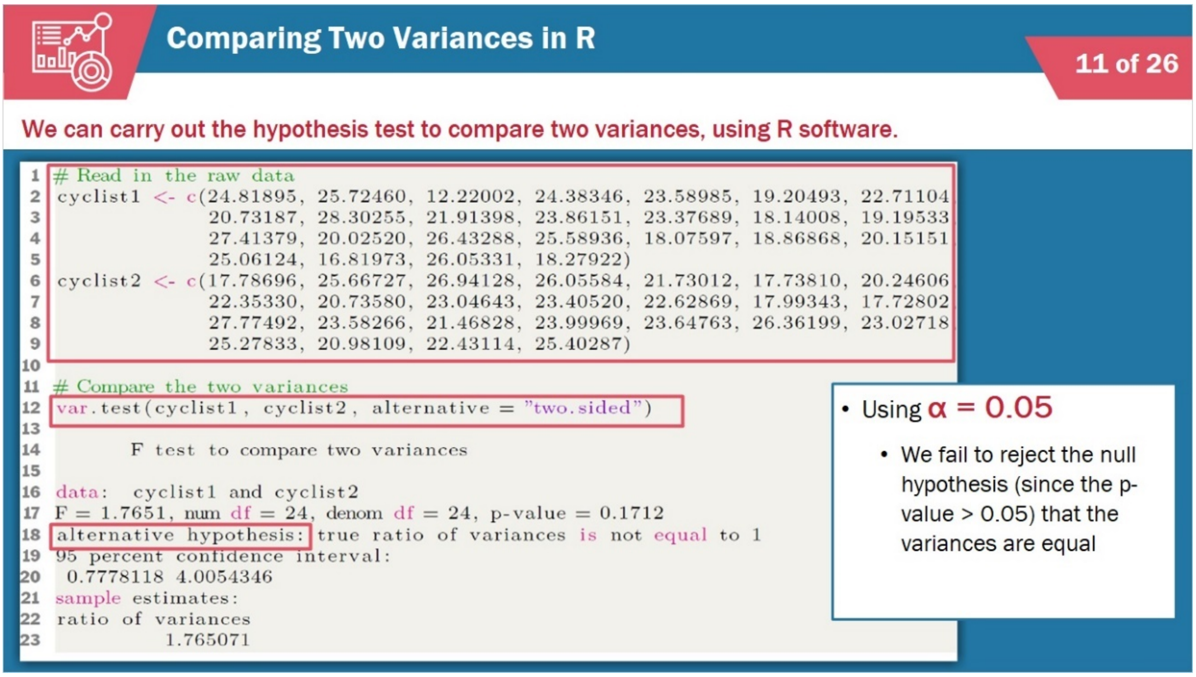
Tab 4: Conclusion



Since our observed test statistic, which is equal to 1.765, lies between our two critical values of 0.441 and 2.27, we fail to reject the null hypothesis and conclude that there is no evidence that sigma 1 squared differs from sigma 2 squared.

We have no evidence that the true population variances of the cycling times differ for the two cyclists.

Slide 11: Comparing Two Variances in R

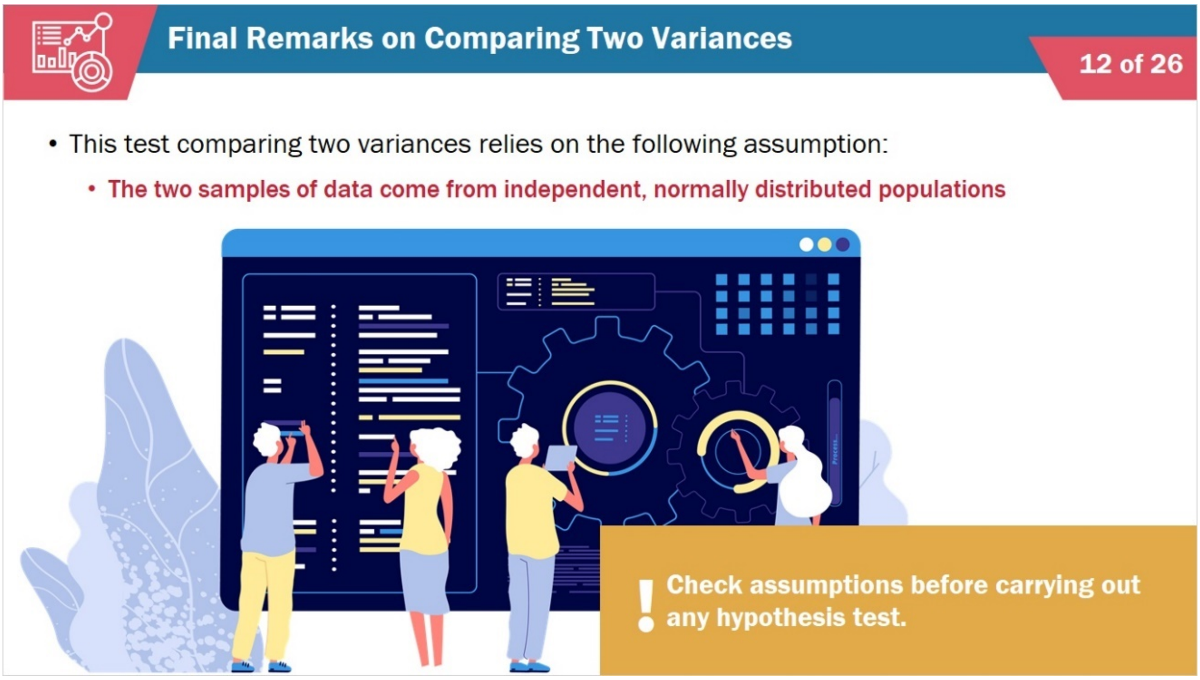


We can carry out the hypothesis test to compare two variances using R software. In the code, we start off by reading in the two vectors containing the cycling times for each cyclist.

We then use the var.test function to do the hypothesis test. We specify the two vectors containing the data and that we want to use a two-sided alternative hypothesis.

The output shows the observed test statistic of 1.7651, the degrees of freedom are 24 and 24, and the p-value is 0.1712. Using alpha = 0.05, we see that the p-value is greater than 0.05 and so we fail to reject the null hypothesis. We have no evidence that the true population variances for the two cyclists differ.

Slide 12: Final Remarks on Comparing Two Variances

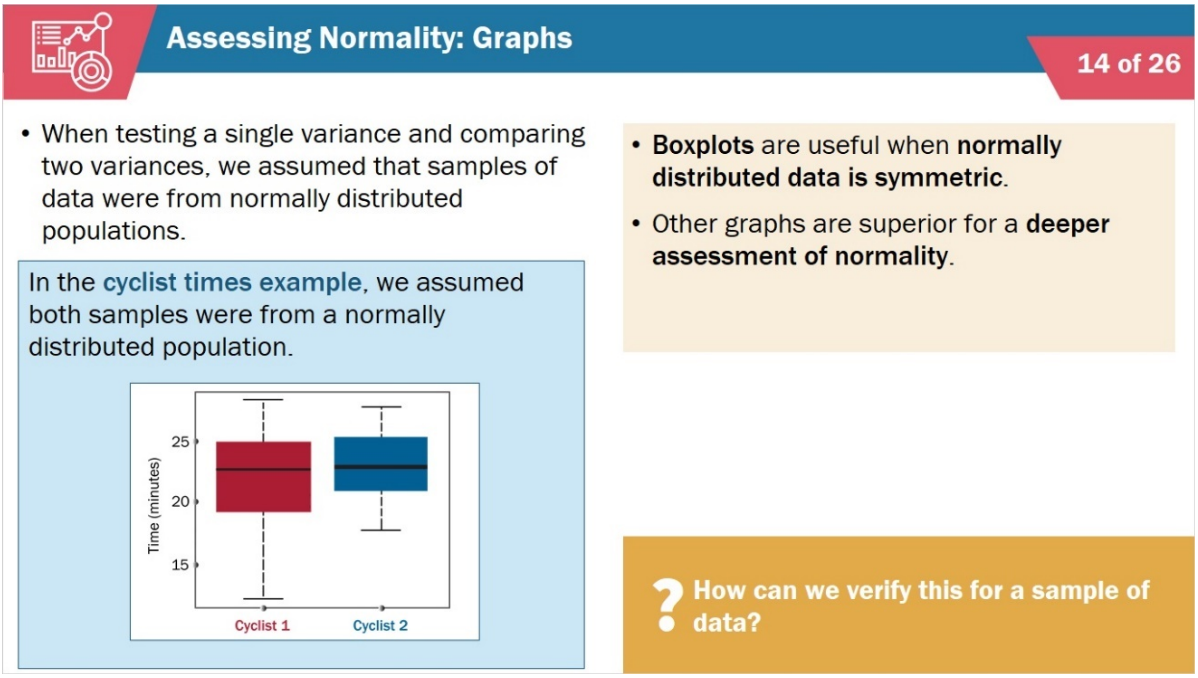


As with most statistical tests, the test we have just seen for comparing two variances has an assumption associated with it. This test relies on the assumption that the two samples of data come from independent normally distributed populations. This test can be quite sensitive to this assumption. It is advisable to check assumptions before doing this or indeed any hypothesis test.

Slide 13: Section 3: Graphical Assessments of Normality

We will now move on to graphical assessments of normality.

Slide 14: Assessing Normality: Graphs

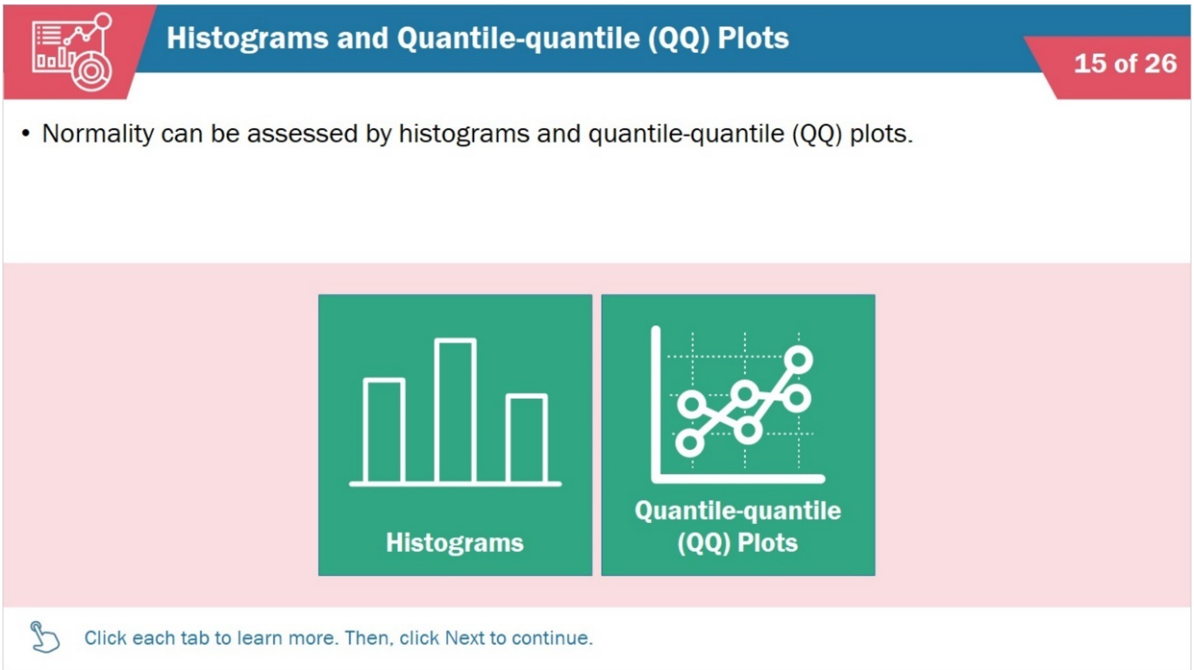


Many statistical tests rely on the assumption that the data comes from a normally distributed population, for example the two tests we have seen so far in this session for testing hypotheses related to variances both assumed that samples were from normally distributed populations.

In the example that recorded cyclist times, we saw boxplots of the two samples of data, here are those boxplots again. Boxplots are quite useful for an initial assessment of a sample of data; for instance, we can see that both samples are symmetric, and we can see that the medians are close. The medians are the black line in the centre of each box.

Normally distributed data is symmetric, and boxplots are useful for assessing symmetry, but there are other graphs that are superior for a deeper assessment of normality. We will explore some of them in this section.

Slide 15: Histograms and Quantile-quantile (QQ) Plots

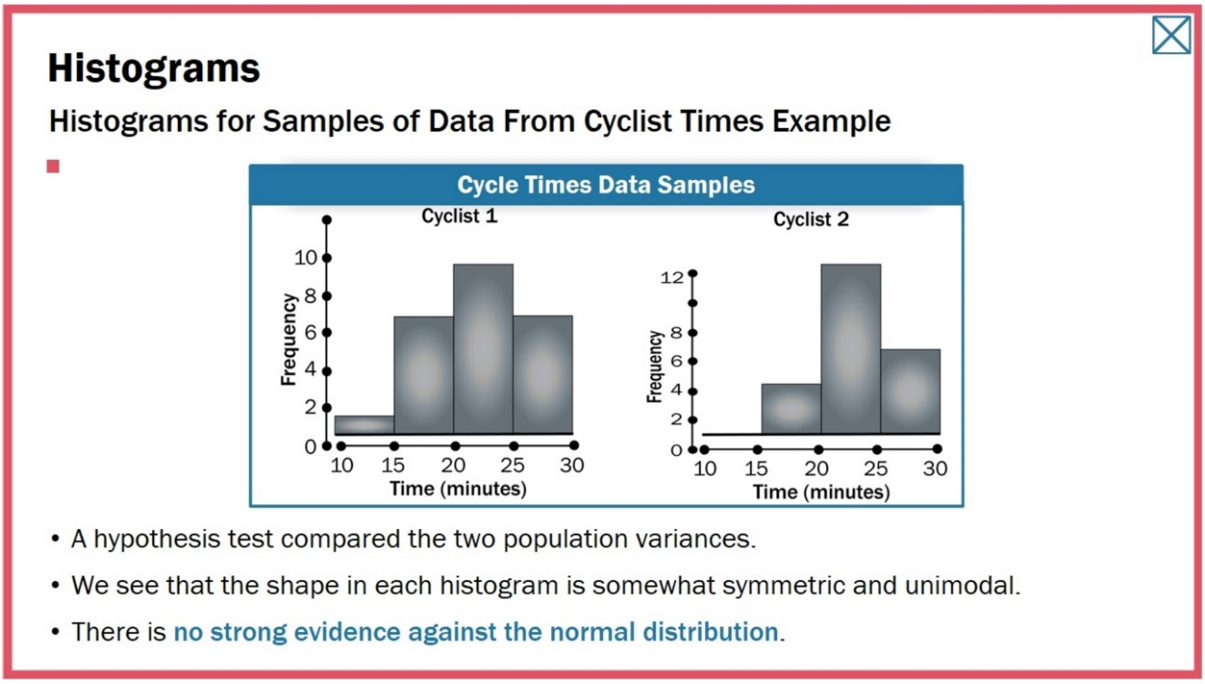


Two useful graphs for assessing normality are the histogram and the QQ plot.

A histogram gives an approximation of the underlying probability density function. While a QQ plot compares a sample to a theoretical distribution; it is a little bit more formal than a histogram but can also be subjective.

Click each tab to learn more. When you are ready, click next to continue.

Tab 1: Histograms



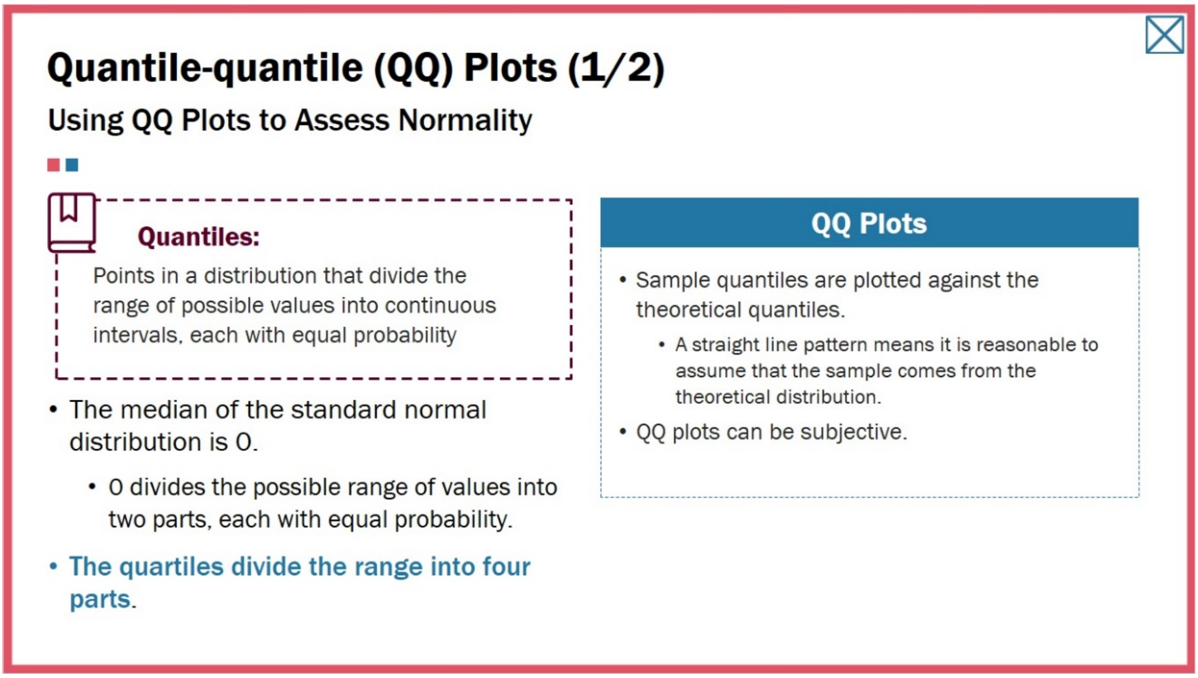
Here are histograms for the two samples of data from the cyclist times example.

We carried out a hypothesis test to compare the two population variances for these datasets that relied on the assumption that the populations were normally distributed.

To check this assumption, in each histogram we are looking for symmetry, unimodality and the distinctive bell-shaped curve that is typical of the normal distribution.

These histograms both seem unimodal and appear to be roughly symmetric. While these two graphs are not depicting perfect bell-shaped curves, there is no strong evidence against the normal distribution here.

Tab 2: Quantile-quantile (QQ) plots



We will now take a look at quantile-quantile (QQ) plots for assessing normality.

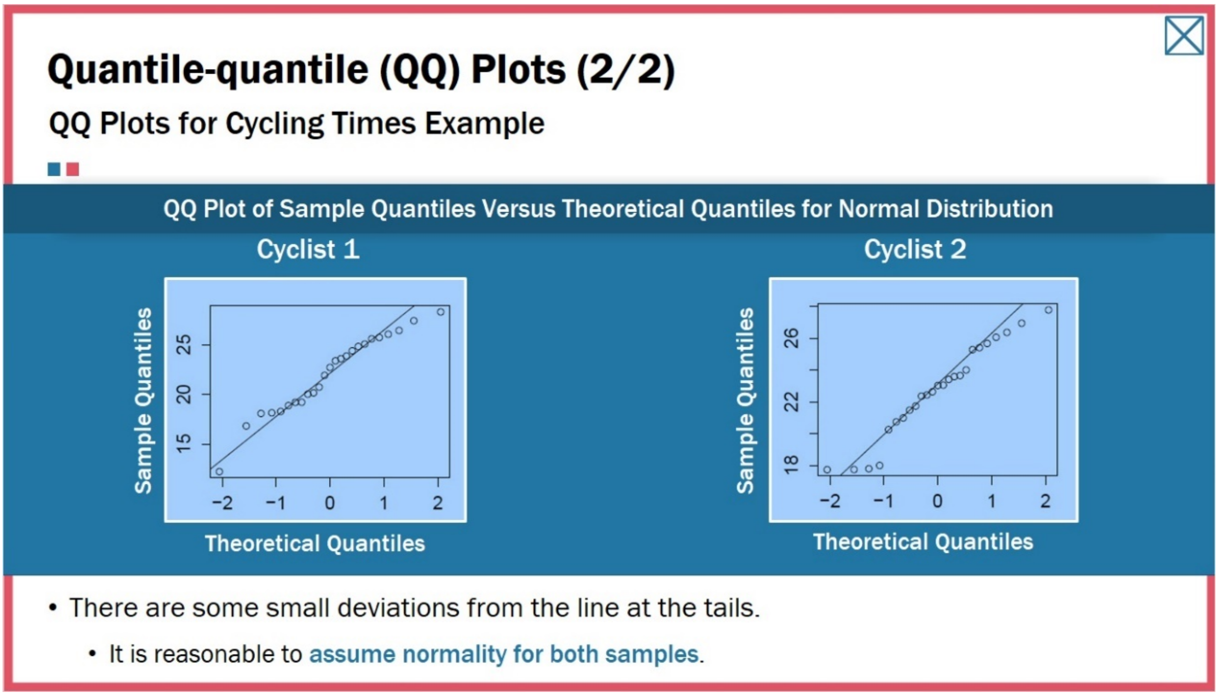
Quantiles are points in a distribution that divide the range of possible values into continuous intervals each with equal probability. For example, the median of the standard normal distribution is 0, and 0 divides the possible range of values into two parts, each with equal probability. If you select a value at random from a standard normal distribution, there is a probability of 0.5 that it will lie below 0, and also a probability of 0.5 that it will lie above 0.

Similarly, the quartiles divide the range into four parts, and so on.

In a QQ plot, the sample quantiles are plotted against the theoretical quantiles and a straight-line pattern means that it is reasonable to assume that the sample comes from the theoretical distribution.

However, QQ plots can be subjective, for example a small bit of deviation from the straight line might be fine, but how much is a small bit?

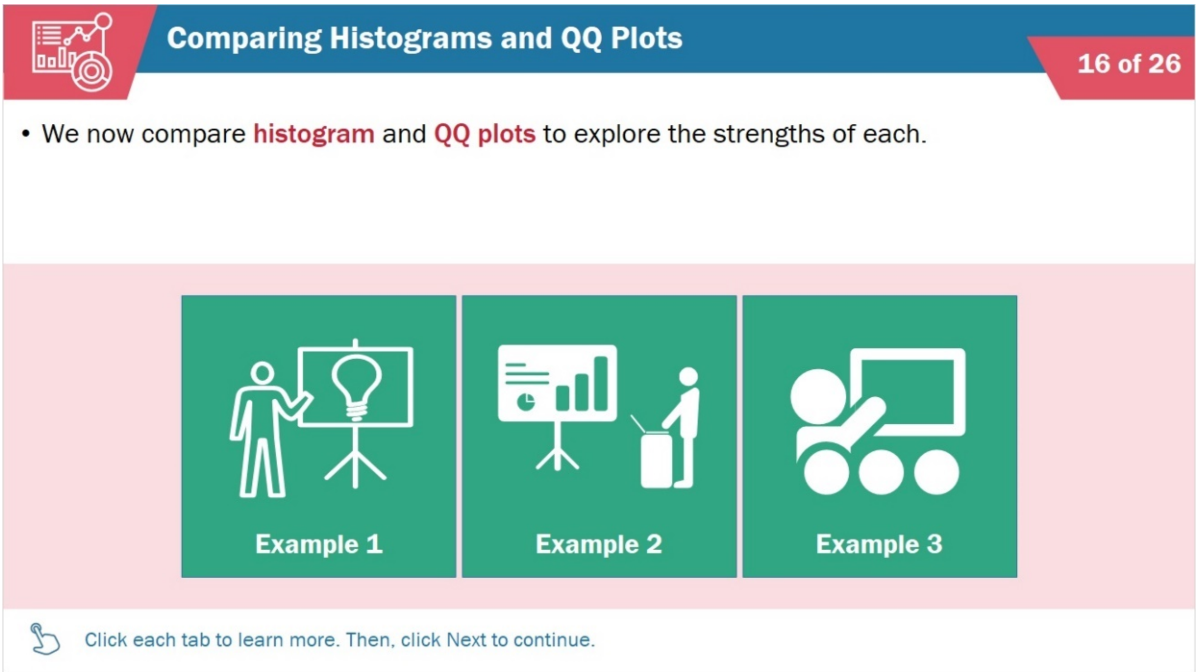
Tab 2.1: QQ Plots Cycling Times Example



Going back to the cyclist times data, here are the QQ plots showing the sample quantiles versus the theoretical quantiles for the normal distribution.

We can see now how QQ plots can be subjective: these plots don't show perfect straight lines, and there are some deviations from the lines, particularly at the tails for the cyclist 2 sample. However, the deviations are not very strong, and it seems somewhat reasonable to assume normality for both samples here.

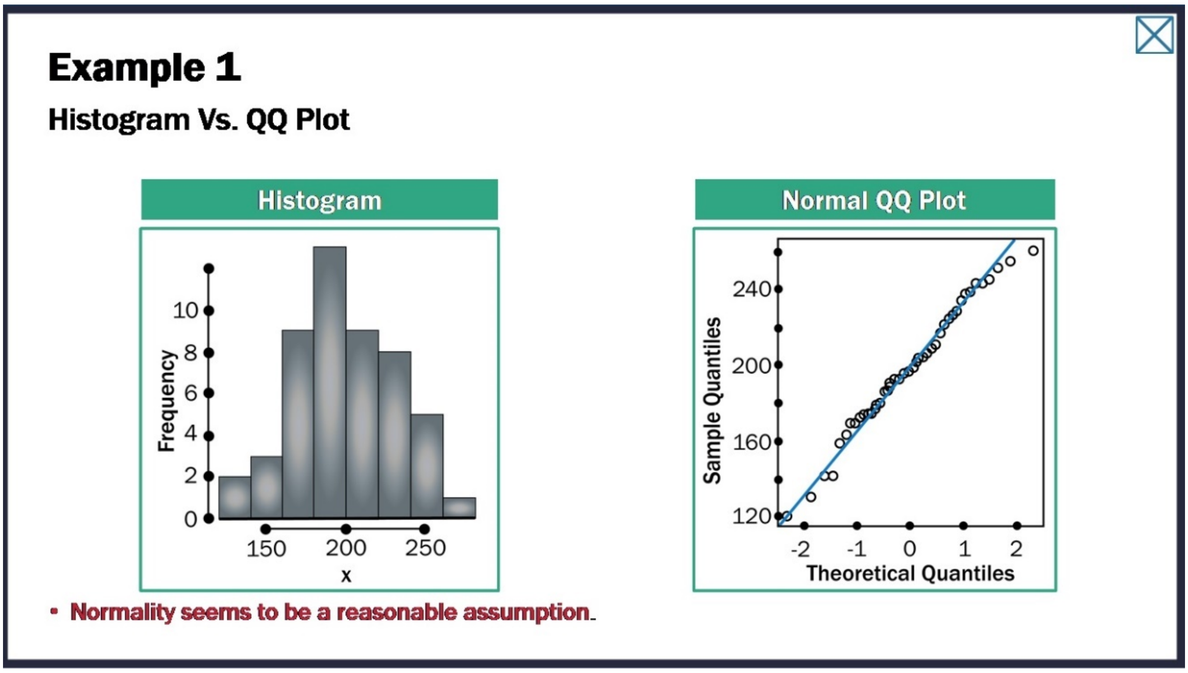
Slide 16: Comparing Histogram and QQ Plots



Let’s compare some examples of histograms and QQ plots to explore the strengths of each.

Click each tab to learn more. When you are ready, click next to continue.

Tab 1: Example 1

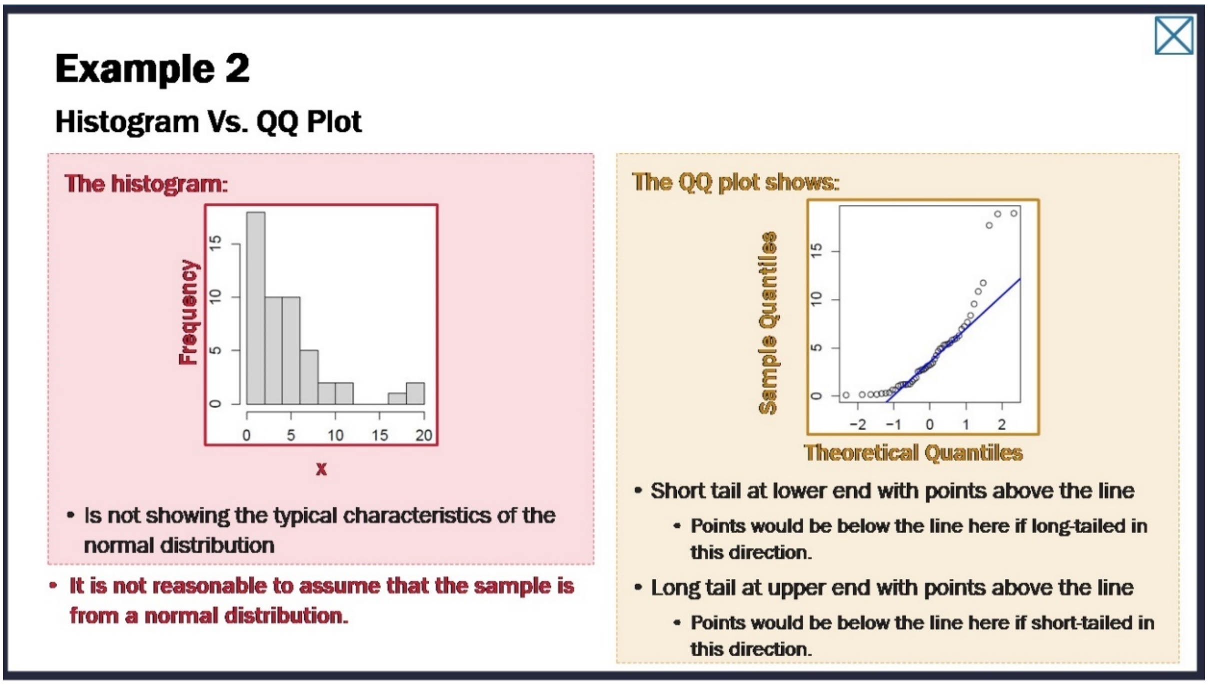


Here we can see a histogram on the left-hand side. The histogram shows a distribution that is unimodal, roughly symmetric, and approximates well the bell-shaped curve that is distinctive to the normal distribution.

In the QQ plot, the relationship between the sample quantiles and theoretical quantiles for the normal distribution appears to follow quite a straight line.

Interpreting both graphs, it certainly seems reasonable to assume that this sample comes from a population with a normal distribution.

Tab 2: Example 2



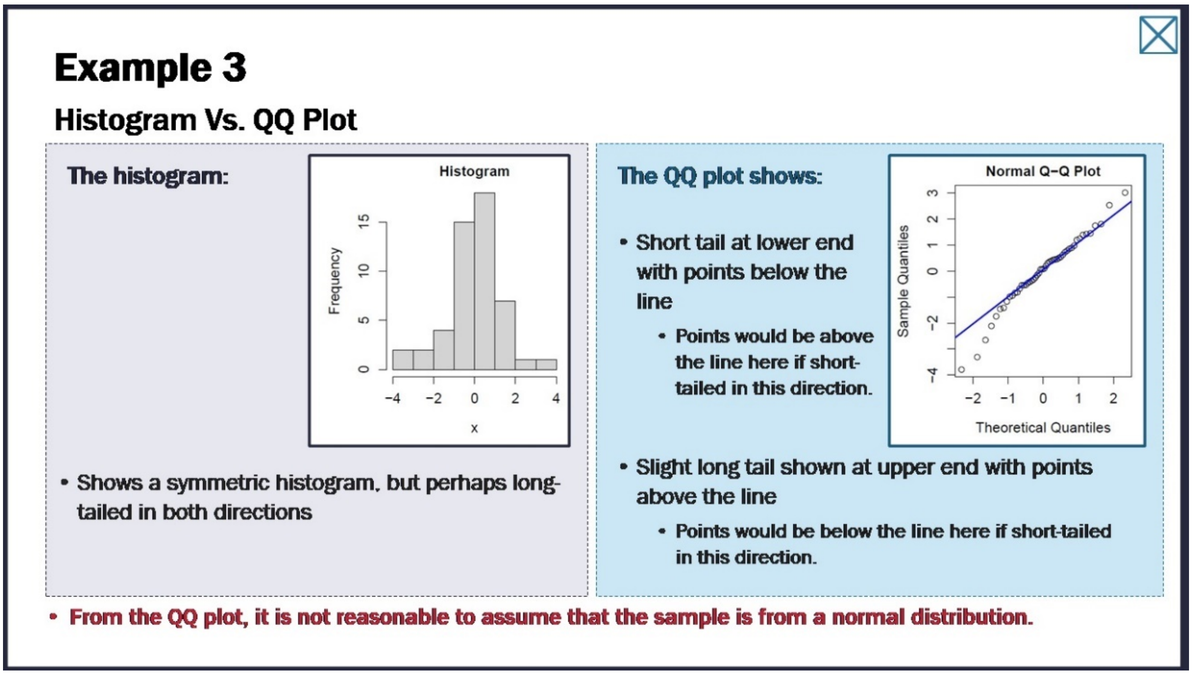
Here we can see a histogram on the left-hand side. The histogram shows a distribution that is skewed to the right, has a short tail on the left, and is asymmetric. It is not showing the typical characteristics of the normal distribution.

In the QQ plot, the relationship between the sample quantiles and theoretical quantiles for the normal distribution does not follow a straight line, it is quite curved. At the lower end, we can see the points are above the line and this matches with the short tail seen on the left-hand side of the histogram. If there were a long tail at the lower end of the sample, this would be illustrated by the points being below the line in the QQ plot.

At the higher end, we can see that the points are all above the line, and this matches with the long tail on the right-hand side of the histogram. If there were a short tail here, the points at this end would be below the line in the QQ plot.

Using the interpretations from both graphs, it is not reasonable to assume that this sample comes from a normal distribution.

Tab 3: Example 3



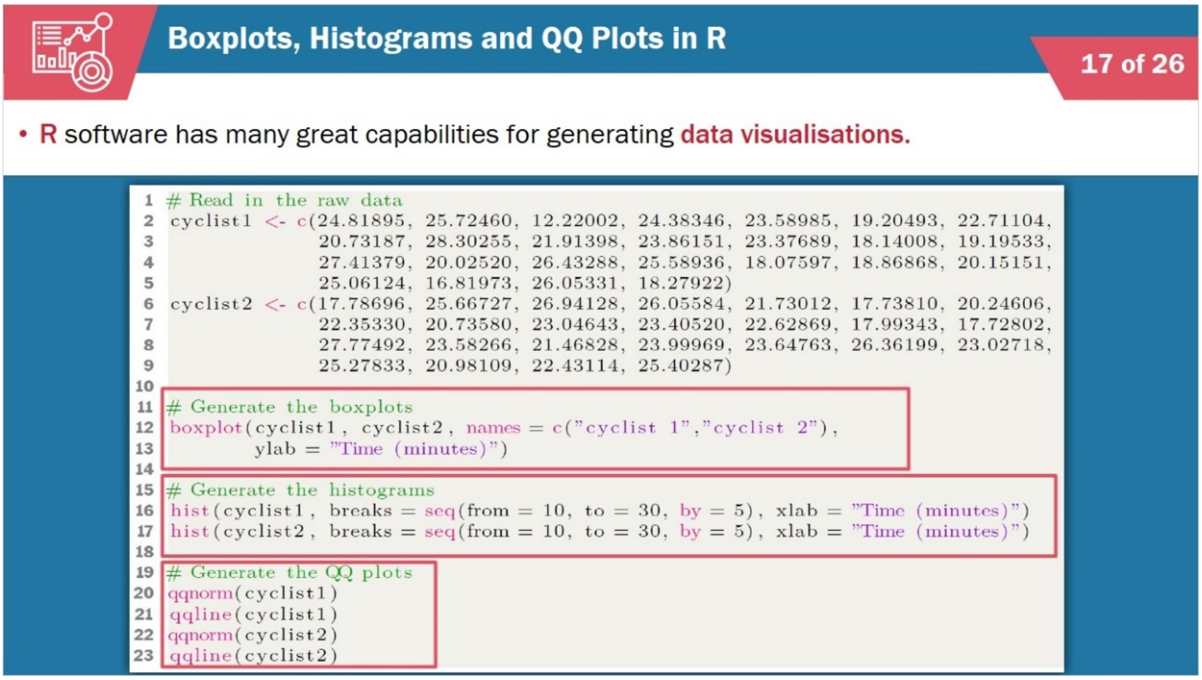
Here we can see a histogram on the left-hand side. The histogram shows a distribution that is symmetric but perhaps long-tailed in both directions, although this is not so easy to be sure about in a histogram.

In the QQ plot, the relationship between the sample quantiles and theoretical quantiles for the normal distribution deviates a bit from a straight line. At the lower end we can see the points are below the line and this matches with the long tail seen on the left-hand side in the histogram. A short tail at this end would be seen by the points falling above the line.

At the higher end, we can see a couple of points are above the line, but the deviation is not as strong as at the other end. It does however match with the slight long tail observed on the right-hand side of the histogram. If there were a short tail here, the points at this end would be below the line.

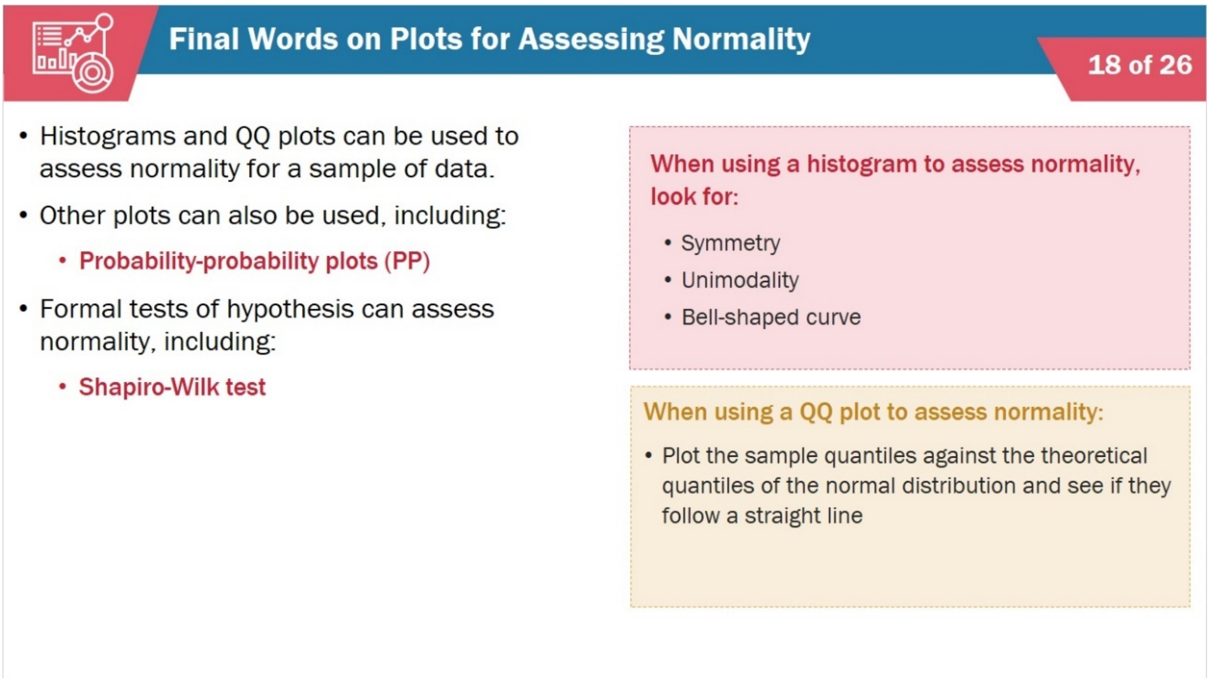
Using the QQ plot, it does not seem reasonable to assume that this sample comes from a normal distribution. This is reflected in the histogram too but might be a little bit harder to spot in the histogram alone.

Slide 17: Boxplots, Histograms and QQ plots in R



R software has many great capabilities for generating data visualizations. The code here shows how to generate the boxplots, histograms and QQ plots in R for the cyclist times data that we have seen already in this session.

Slide 18: Final Words on Plots for Assessing Normality



We have seen how to use histograms and QQ plots for assessing normality for a sample of data.

With a histogram, we are looking for symmetry, unimodality and the bell-shaped curve typical of the normal distribution.

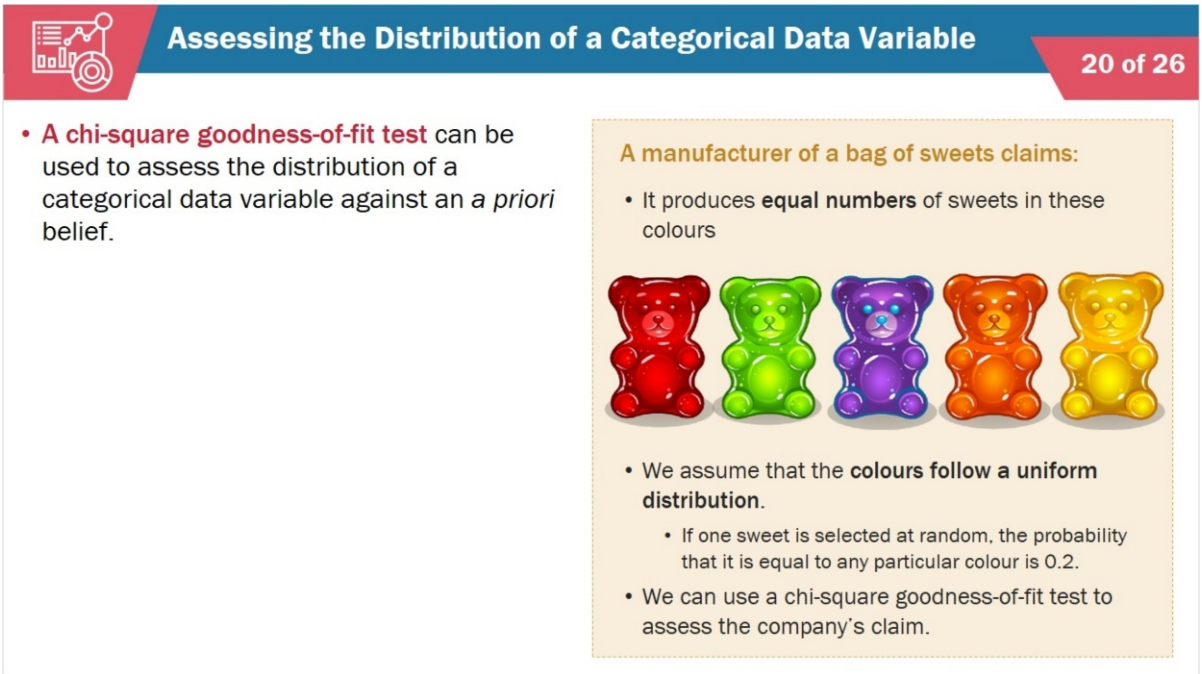
With QQ plots, we plot the sample quantiles against the theoretical quantiles of the normal distribution and see if they follow a straight line.

There are other plots that can also be used for assessing normality, for example PP or probability-probability plots, and there are also formal tests of hypothesis that can assess normality, for example the Shapiro-Wilk test.

Slide 19: Section 4: Chi-square Goodness-of-Fit Test

We will now move on to the final section of this session on Chi-square goodness-of-fit tests.

Slide 20: Assessing the Distribution of a Categorical Data Variable



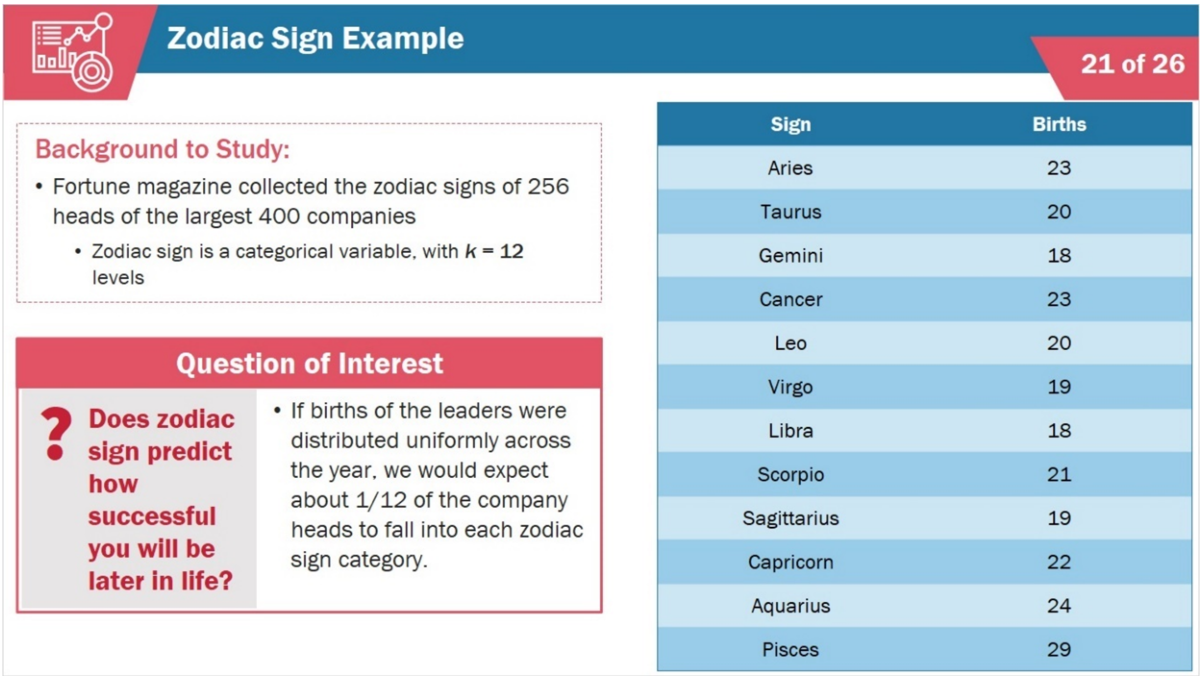
A Chi-square goodness-of-fit test can be used to assess the distribution of a categorical data variable against an *a priori* belief.

For example, suppose a company manufactures a bag of sweets and states that they produce equal numbers of each of the colours red, green, purple, orange and yellow. In this case, it is assumed that the colours follow a uniform distribution each with colour having probability 0.2. Suppose I select a sweet at random, the probably that it will be red is equal to 0.2.

We could use a Chi-square goodness-of-fit test to assess the company's claim. Suppose you like to eat these sweets and like the red ones the best, but really feel like they are less abundant than all the other colours! You could collect data and use it to test the uniform distribution assumption and see if your complaint has a basis.

Let's now explore how to conduct a Chi-square goodness-of-fit test.

Slide 21: Zodiac Sign Example



Sign	Births
Aries	23
Taurus	20
Gemini	18
Cancer	23
Leo	20
Virgo	19
Libra	18
Scorpio	21
Sagittarius	19
Capricorn	22
Aquarius	24
Pisces	29

We will start with a motivating example.

A study was carried out as follows.

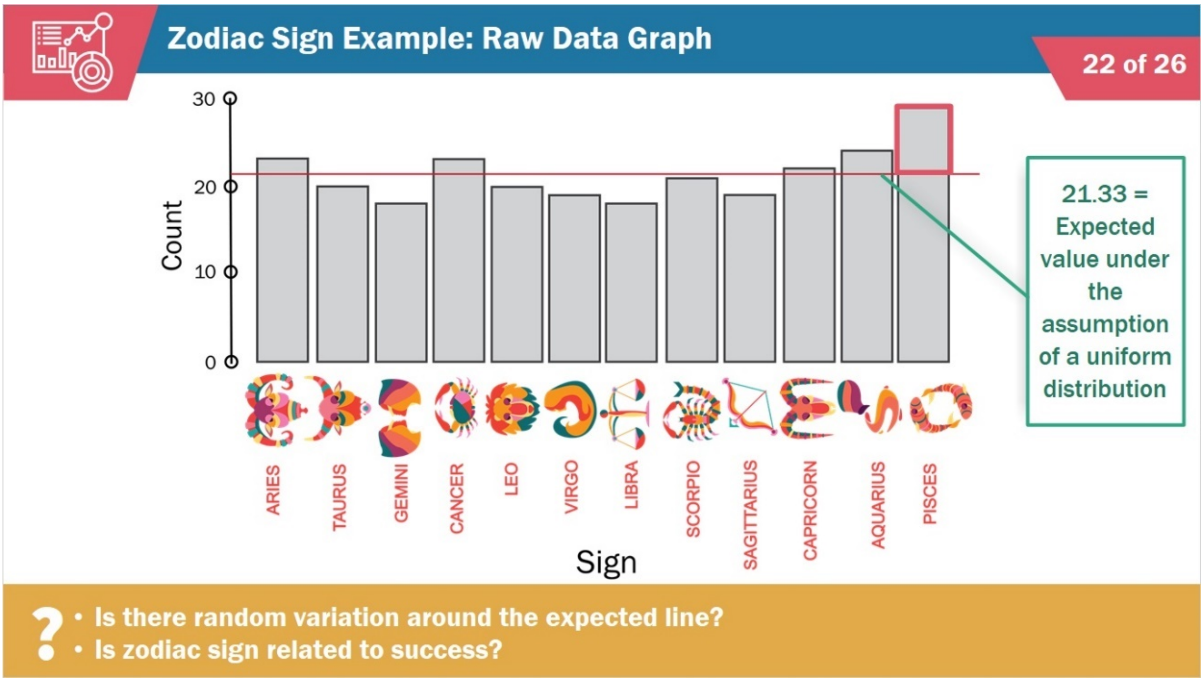
Fortune magazine collected the zodiac signs of 256 heads of the largest 400 companies. Zodiac sign is a categorical variable with k = 12 levels. The counts per zodiac sign are shown in this table.

A question of interest here is whether or not zodiac sign can predict how successful you will be later in life.

Putting this in statistical language, if success is not related to zodiac sign, we would expect the leaders to be uniformly distributed across the zodiac signs. That means we would expect approximately 256 divided by 12 people to lie in each category.

We can use a Chi-square goodness-of-fit test to test if the data deviates from a uniform distribution.

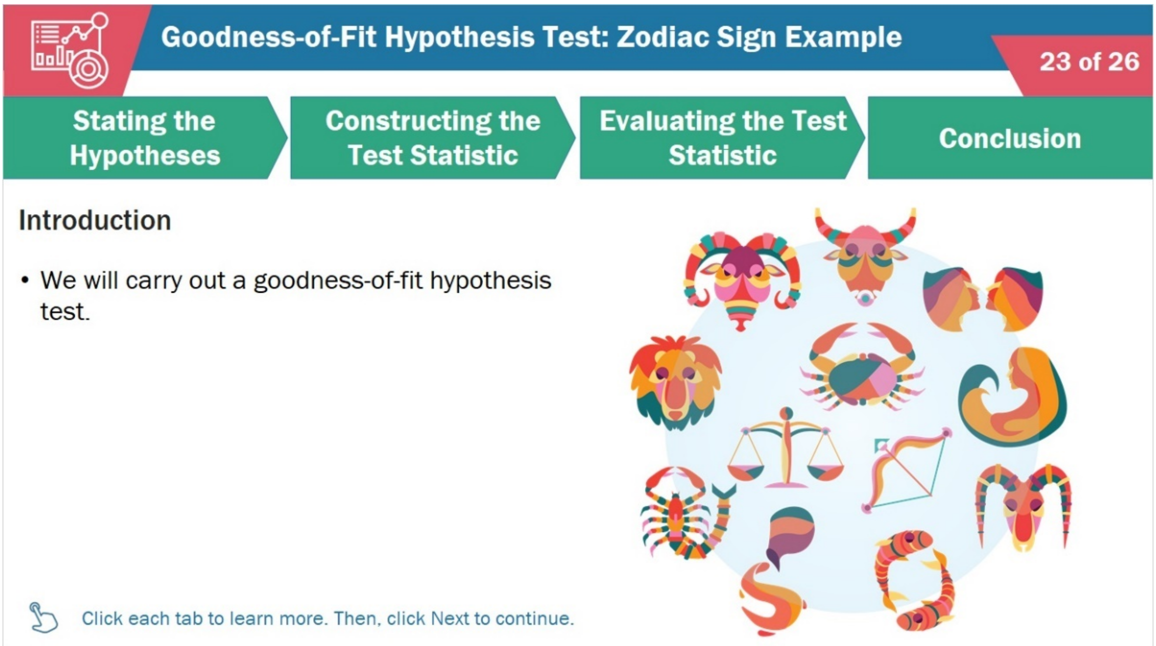
Slide 22: Zodiac Sign Example – Raw Data Graph



Here you can see a bar plot of the raw data counts across the zodiac signs. Also shown is a horizonal line at 21.33; this line is the expected value for each zodiac sign under the null hypothesis of a uniform distribution. The value 21.33 was computed by 256 divided by 12. We can see that some categories fall a little below and some a little above the 21.33 line. The Pisces category is the furthest away from the line.

Are we simply seeing random variation around the expected line? Or is zodiac sign related to success? Let’s test this formally.

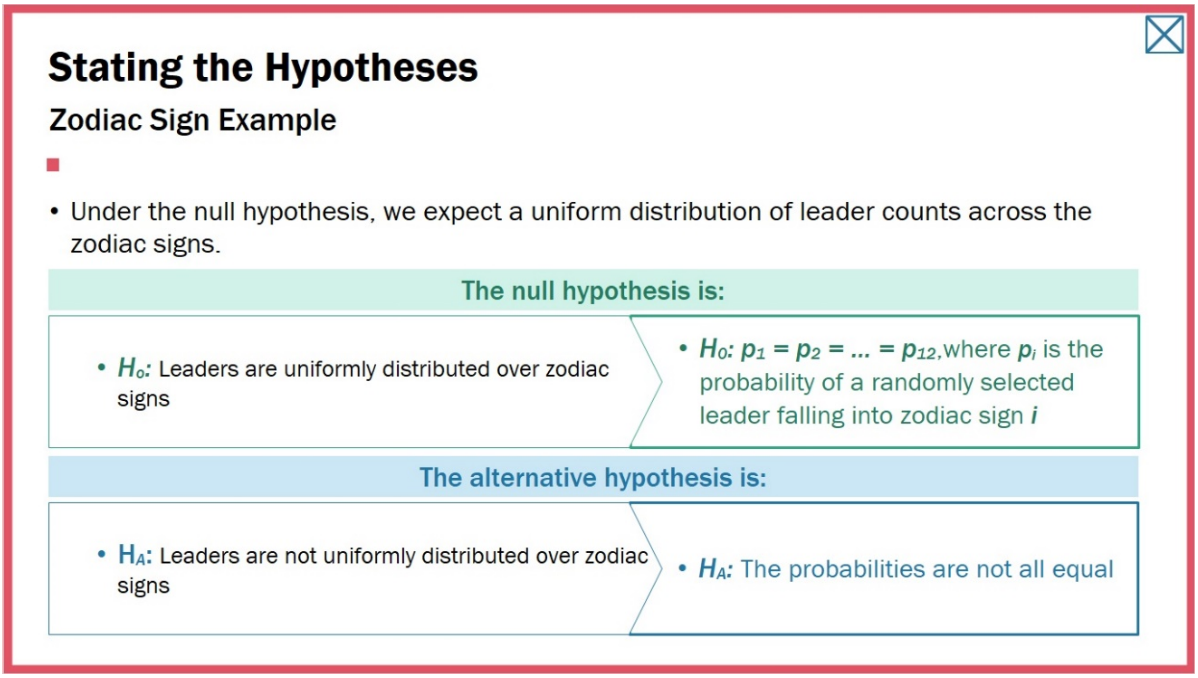
Slide 23: Goodness-of-Fit Hypothesis Test: Zodiac Sign Example



Once again, the now familiar steps are used to conduct this hypothesis test: specify the hypotheses, construct a test statistic, evaluate the test statistic against the reference distribution and make a conclusion.

Click each tab to learn more. When you are ready, click next to continue.

Tab 1: Stating the Hypotheses



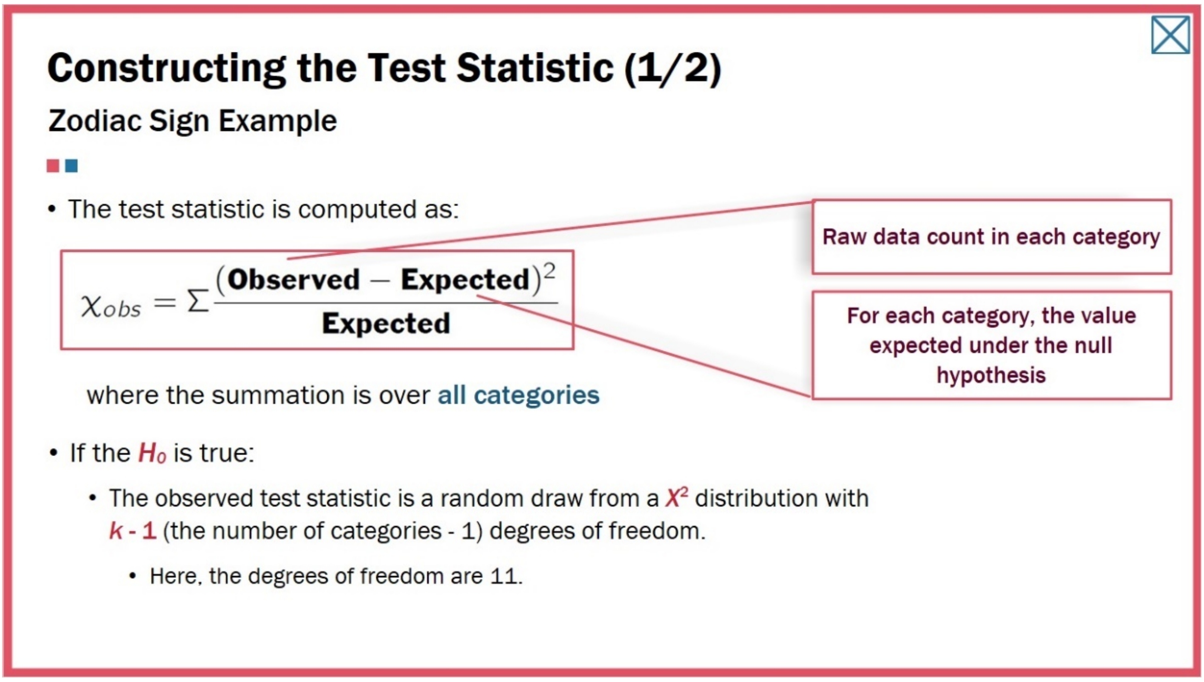
Let's start off by stating the hypotheses. Under the null hypothesis, we don't expect any relationship between zodiac sign and leaders, for example, we don't have *a priori* belief that leaders are more likely to come from some signs than others.

So, our null hypothesis is that leaders are uniformly distributed over the zodiac signs.

The alternative hypothesis is that leaders are not uniformly distributed over zodiac signs. With this alternative hypothesis, we don't speculate as to what signs might be more likely than others, it just says that the distribution isn't uniform across them.

To put these hypotheses in terms of statistical parameters, we could define p_i to be the probability of a randomly selected leader falling into zodiac sign i, where i goes from 1 to 12. Under the null hypothesis, all the p_i's are equal, and under the alternative hypothesis, these probabilities are not all equal. With this alternative hypothesis, we don't say that the p_i's all have to be different, so for example, the scenario of 11 p_i's being equal and one being different falls under the alternative hypothesis.

Tab 2: Constructing the Test Statistic (1/2)



The test statistic is computed by calculating the sum over all categories of observed minus expected to be squared, all divided by expected.

In our example, there are twelve categories to sum over. The observed values are the raw data counts given for each zodiac sign. The expected values are what we would expect in each category under the null hypothesis. In our example, we are assuming a uniform distribution across the zodiac categories, and so we have an equal number across all the expected values. There were 256 people sampled, and so 256 /12 is the expected value for each cell.

If the null hypothesis is true, the observed test statistic is a random draw from a chi square distribution with k - 1 degrees of freedom, where k is the number of categories. Here we had 12 zodiac sign categories and so the degrees of freedom are 12 – 1 = 11.

The form of this test statistic should be familiar to you from the Chi-square test of independence of two categorical variables that we saw in the last session.

Tab 2.1: Constructing the Test Statistic (2/2)

Constructing the Test Statistic (2/2)

Zodiac Sign Example

Observed and Expected Values for Calculation

	Aries	Taurus	Gemini	Cancer	Leo	Virgo	Libra	Scorpio	Sagittarius	Capricorn	Aquarius	Pisces
Observed	23	20	18	23	20	19	18	21	19	22	24	29
Expected	21.33	21.33	21.33	21.33	21.33	21.33	21.33	21.33	21.33	21.33	21.33	21.33

- This gives an observed test statistic of:

$$\begin{aligned} X_{obs} &= \sum \frac{(\textbf{Observed} - \textbf{Expected})^2}{\textbf{Expected}} \\ &= \frac{(23-21.33)^2}{21.33} + \frac{(20-21.33)^2}{21.33} + \ldots + \frac{(29-21.33)^2}{21.33} \\ &= \boxed{5.09} \end{aligned}$$

Let’s take a look at the values for the calculation.

In the table, you can see the observed values for each zodiac sign, for example the first one is 23 for Aries. We can also see the expected value for each cell underneath. This is the same for each cell and is computed as 256 / 12 = 21.33.

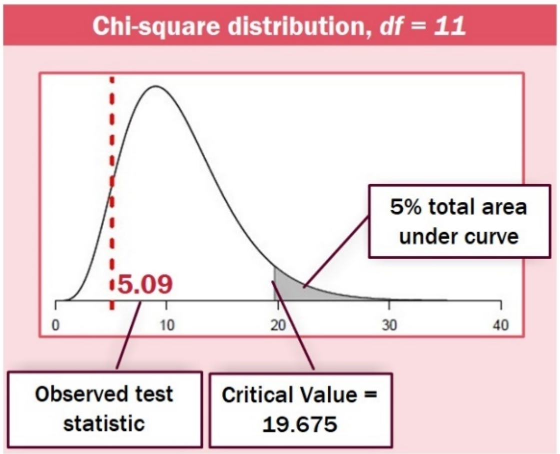
We compute observed minus expected squared, all divided by expected for each cell. For the first cell this is 23 – 21.33 to be squared, divided by 21.33. We do this for each cell and sum the values giving an observed test statistic value of 5.09.

Tab 3: **Evaluating the Test Statistic**

Evaluating the Test Statistic

Zodiac Sign Example

- The reference distribution for assessing X_{obs} is the Chi-square distribution with 11 degrees of freedom.
- Using α = 0.05, the critical value is **19.675**.
- The observed test statistic is not extreme for this distribution.
  - Therefore, we fail to reject the null hypothesis.



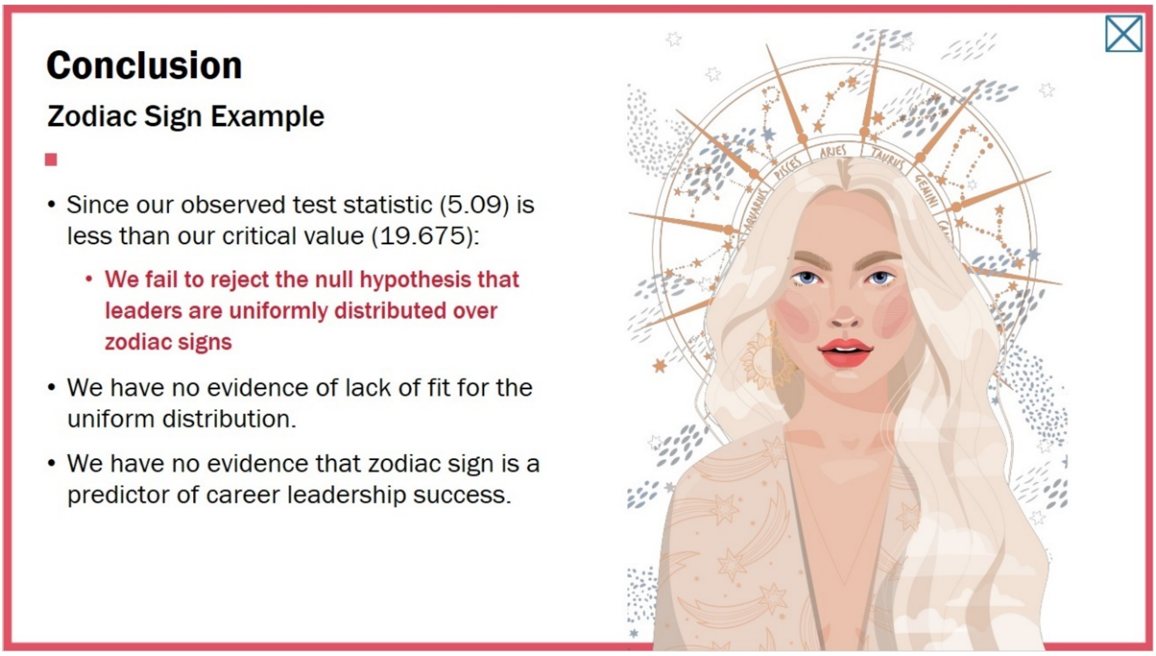
We will now evaluate the test statistic against the reference or null distribution.

In this example, if the null hypothesis is true, the observed test statistic will be a realisation from a Chi-square distribution with 11 degrees, where 11 is the number of categories 12 minus 1.

You can see the shape of the Chi-squared distribution with 11 degrees of freedom in the graph. Using alpha = 0.05, the critical value is 19.675. The shaded grey area in the graph highlights the area under the curve to the right of 19.675 and the shaded grey area equals 5% of the total area under the curve. Any value lying higher than our critical value is considered to be extreme for this distribution.

Our observed test statistic is 5.09. Since this is lower than 19.675, our observed test statistic is not extreme for this distribution and we fail to reject the null hypothesis.

Tab 4: Conclusion

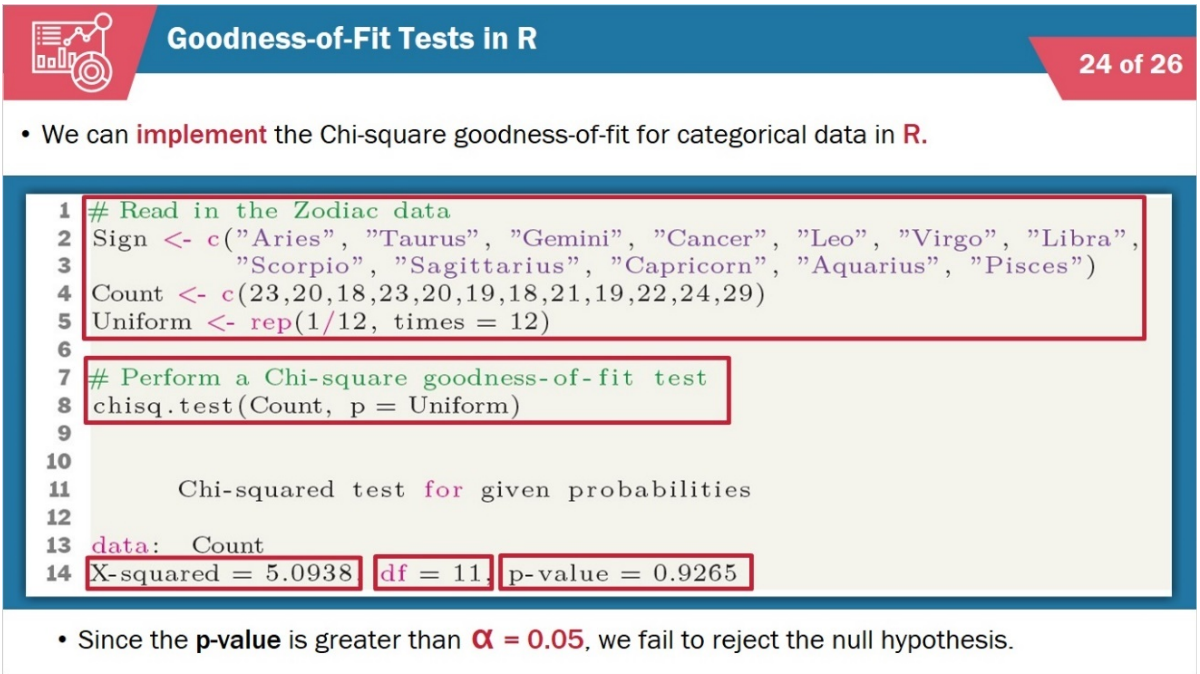


Since our observed test statistic of 5.09 is less than our critical value 19.675, for alpha = 0.05, we fail to reject the null hypothesis that leaders are uniformly distributed over zodiac signs.

This means that we have no evidence of lack of fit for the uniform distribution.

Or another way of putting this is that we have no evidence that zodiac sign is a predictor of career leadership success.

Slide 24: Goodness-of-Fit Tests in R



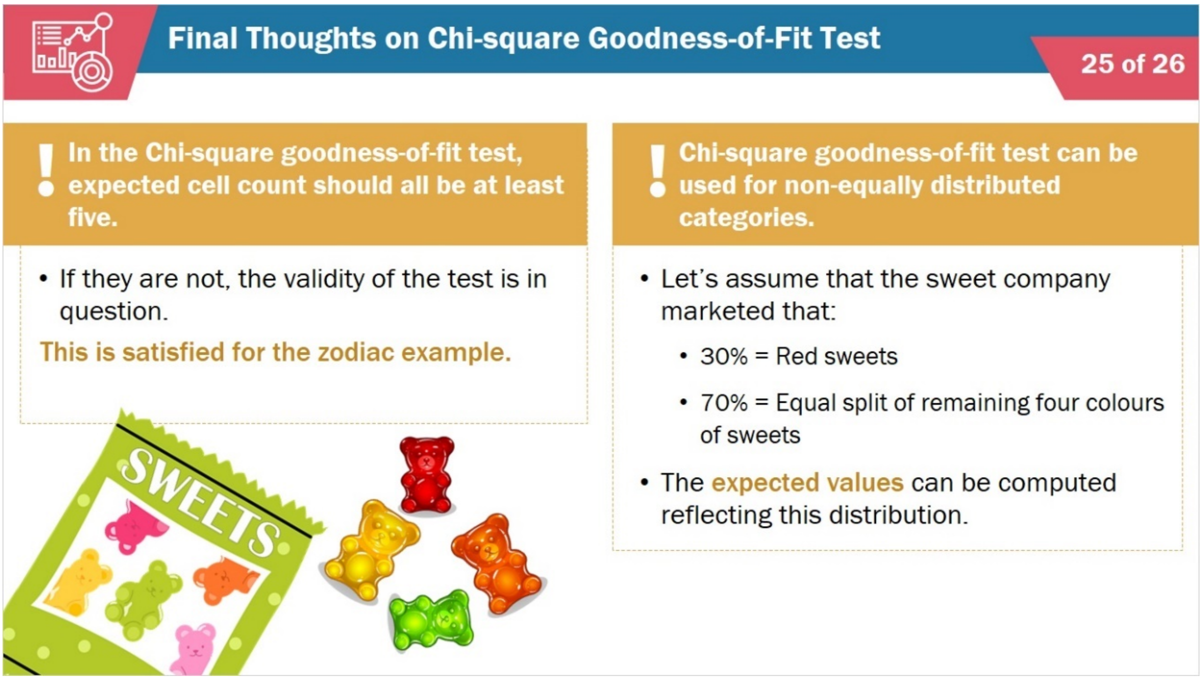
We can implement the Chi-square goodness-of-fit test for categorical data in R. Here is the code to perform the test in R for the zodiac data.

The first few lines of code are reading in the data; the Sign vector includes the zodiac signs, and the Count vector includes the raw data count values. The Uniform vector contains 12 probabilities all equal to 1/12, the expected probability of a leader falling into each category under the uniform distribution assumption.

We use the chisq.test function to do the test and specify the raw count values and the probabilities under the null hypothesis.

We see that the test statistic is equal to 5.0938, with degrees of freedom 11, and a p-value of 0.9265. Since the p-value is greater than alpha = 0.05, we fail to reject the null hypothesis and conclude as before.

Slide 25: Final Thoughts on Chi-square Goodness-of-Fit Test



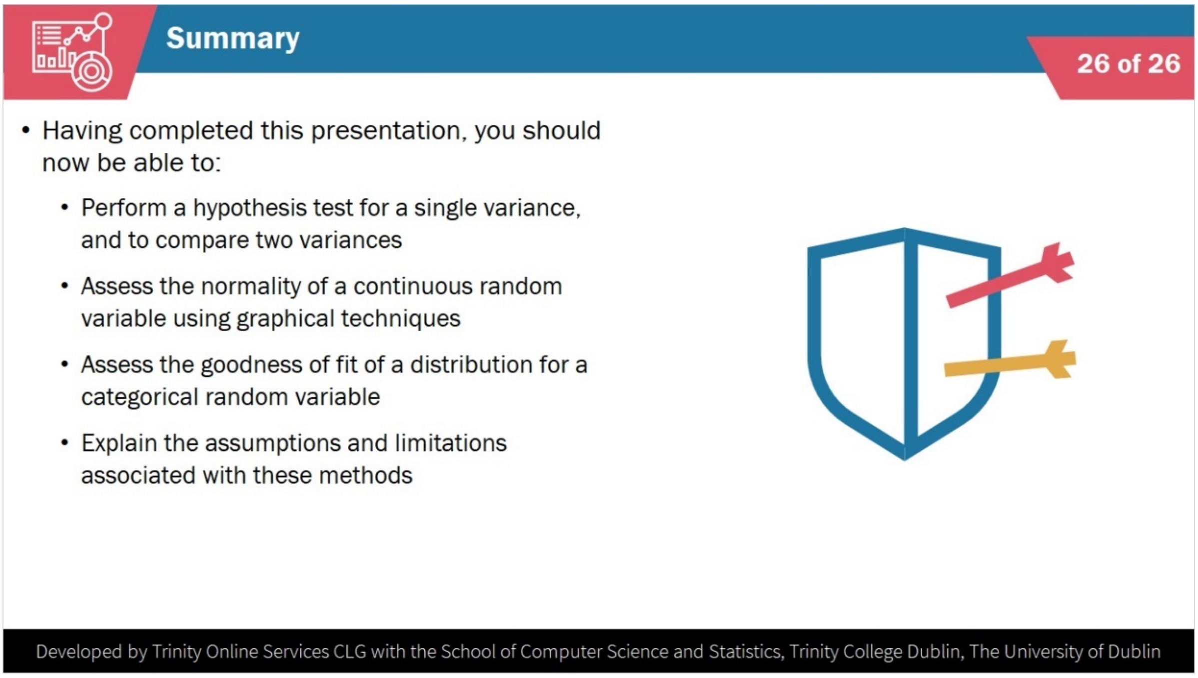
There are two final considerations related to this test.

The first is that in the Chi-square goodness-of-fit test, the expected cell counts should all be at least 5. If they are not, the validity of the test is in question. This is satisfied for the zodiac example.

The second point is that in our zodiac example, we assumed a uniform distribution, i.e., that each category had an equal probability under the null hypothesis. The Chi-square goodness-of-fit test can also be used for non-equally distributed categories.

For example, going back to the sweet example introduced earlier on, perhaps the company marketed that red sweets make up 30% of sweets, with the remaining 70% split equally among the other four colours. The expected values would be computed to reflect this, and in R the vector containing the null distribution probabilities can have varying probabilities across the categories.

Slide 26: Summary



Having completed this presentation, you should now be able to:

- Perform a hypothesis test for a single variance, and to compare two variances.
- Assess the normality of a continuous random variable using graphical techniques.
- Assess the goodness of fit of a distribution for a categorical random variable and
- Explain the assumptions and limitations associated with these methods.