Applied Probability II

Section 4: Confidence intervals and hypothesis tests

Professor Caroline Brophy

Semester 2, 2020-21

Applied Probability II. Section 4: Confidence intervals and hypothesis tests

Section 4.1: The sampling distribution of $\bar{Y}$

Using the sample mean to estimate the population mean

In Section 3.2 (Introduction to estimation), we introduced the idea of using the sample mean ($\bar{y}$) to estimate the population mean (μ).

That is, suppose we have collected $y_1, y_2, ..., y_n$, where these values are realisations of random variables $Y_1, Y_2, ..., Y_n$, and come from a random sample and so are independent and follow the same probability distribution.

Then

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \hat{\mu}$$

How good is our estimate $\bar{y}$? Or how close to μ is it?

To answer this, we need to make an assumption about the distribution of the Y_i's.

How precise is our estimate?

We will assume that $Y_1, Y_2, ..., Y_n \sim N(\mu, \sigma^2)$.

The expected value of any Y_i is $E[Y_i] = \mu$ and the variance is $Var(Y_i) = \sigma^2$.

Consider the random variable

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

If we can describe the distribution of $\bar{Y}$, then we can quantify how precise our estimates of μ are (and where we are using $\bar{y}$ to estimate μ).

Aside: properties of expectation and variance

Before we examine the distribution of $\bar{Y}$, here are two useful results, where a and b are constants:

1. The linearity of expectation

$$E[aY_1 + bY_2] = aE[Y_1] + bE[Y_2]$$

2. Properties of the variances:

$$Var(aY_1 + bY_2) = a^2 Var(Y_1) + b^2 Var(Y_2) + 2abCov(Y_1, Y_2)$$
$$Var(aY_1 - bY_2) = a^2 Var(Y_1) + b^2 Var(Y_2) - 2abCov(Y_1, Y_2)$$

The expected value of $\bar{Y}$

Remember, we have assumed that $Y_1, Y_2, ..., Y_n \sim N(\mu, \sigma^2)$.

Or we can say: $E[Y_i] = \mu$, $Var(Y_i) = \sigma^2$, and the Y_i are normally distributed.

Remember too: $E[aY_1 + bY_2] = aE[Y_1] + bE[Y_2]$.

Then:

$$\begin{aligned}
E[\bar{Y}] &= E[\frac{\Sigma_{i=1}^{n} Y_i}{n}] = \frac{1}{n} E[\Sigma_{i=1}^{n} Y_i] \\
&= \frac{1}{n} E[Y_1 + Y_2 + ... + Y_n] \\
&= \frac{1}{n} (E[Y_1] + E[Y_2] + ... + E[Y_n]) \\
&= \frac{1}{n} \Sigma_{i=1}^{n} E[Y_i] = \frac{1}{n} \Sigma_{i=1}^{n} \mu = \frac{1}{n} n\mu \\
&= \mu
\end{aligned}$$

So, our expected value of our estimator of μ is μ. This is good news!

What about the variance of $\bar{Y}$?

The variance of $\bar{Y}$

What is the $Var(\bar{Y})$?

Remember, we have assumed that $Y_1, Y_2, ..., Y_n \sim N(\mu, \sigma^2)$.

Remember too: $Var(aY_1 - bY_2) = a^2 Var(Y_1) + b^2 Var(Y_2) - 2abCov(Y_1, Y_2)$

$$
\begin{aligned}
Var(\bar{Y}) &= Var(\frac{\Sigma_{i=1}^n Y_i}{n}) = \frac{1}{n^2} Var(\Sigma_{i=1}^n Y_i) \\
&= \frac{1}{n^2} Var(Y_1 + Y_2 + ... + Y_n) \\
&= \frac{1}{n^2} [Var(Y_1) + Var(Y_2) + ... + Var(Y_n)] \\
&= \frac{1}{n^2} \Sigma_{i=1}^n Var(Y_i) = \frac{1}{n^2} \Sigma_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

What about the covariances in line 3 of the equation?

If two random variables are independent, then their covariance is equal to 0.

The sampling distribution of $\bar{Y}$

Recap, we want to know the distribution of $\bar{Y}$.

We have assumed

- $Y_i \sim N(\mu, \sigma^2)$, and
- the Y_i's are independent.

We have already proven

- $E[\bar{Y}] = \mu$, and
- $Var(\bar{Y}) = \frac{\sigma^2}{n}$

Final step

$\bar{Y}$ can be written as: $\bar{Y} = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + ... + \frac{1}{n}Y_n$

and so is a linear combination of normal random variables.

We can assume: the linear combination of independent normal random variables is also normal.

Therefore $\bar{Y}$ is itself a normal random variable. Hence, we've shown that

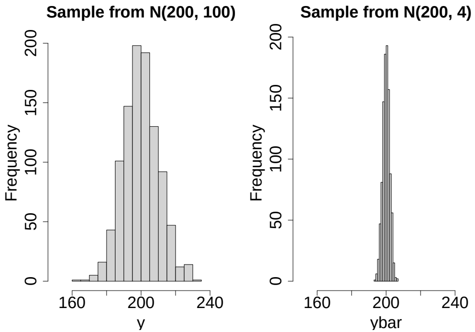
$$\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$$

An example

Assume we have $Y_i \sim N(200, 100)$.

Then what is the sampling distribution of $\bar{Y}$ for samples of size 25?

$\bar{Y} \sim N(200, \frac{100}{25}) = N(200, 4)$.



Final note

Since

$$\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$$

we can see that the variance of $\bar{Y}$ decreases as n increases. Or, we can say that our estimate of the mean is more precise the larger that n is (which makes sense intuitively!).

Recall from previous modules, that if $X \sim N(\mu_X, \sigma_X^2)$, then

$$\frac{X - \mu_X}{\sigma_X} \sim N(0, 1)$$

This is referred to as standardising X.

Applying this to $\bar{Y}$, we get:

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

We will use this in the next Section to construct a confidence interval for μ.

Section 4.2: Confidence interval for μ (variance known)

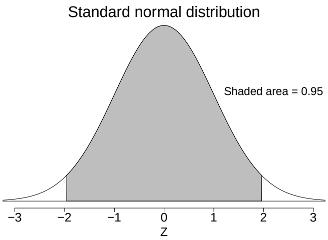
Assume we know σ

We will construct a CI for a population mean.

- To make things easier, assume we know the true value of σ.
- We are still assuming that $Y_1, Y_2, ..., Y_n \sim N(\mu, \sigma^2)$.
- Therefore we know from the previous section that $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$.

For the standard normal distribution, $Z \sim N(0, 1)$, we know that

$$Pr(-1.96 \leq Z \leq 1.96) = 0.95.$$



Constructing the CI

Since $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, this tells us that:

$$Pr(-1.96 \leq \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq 1.96) = 0.95$$
$$Pr(-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{Y}-\mu \leq 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$
$$Pr(-1.96\frac{\sigma}{\sqrt{n}} \leq \mu-\bar{Y} \leq 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$
$$Pr(\bar{Y}-1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y}+1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

This tells us that the probability the true value of μ lies in the interval

$$(\bar{Y}-1.96\frac{\sigma}{\sqrt{n}}, \bar{Y}+1.96\frac{\sigma}{\sqrt{n}})$$

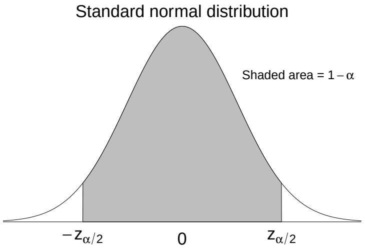
is 0.95, or we have 95% confidence that μ will be in any randomly selected interval.

95% of all realised intervals $\bar{y} \pm 1.96\frac{\sigma}{\sqrt{n}}$ will contain μ.

Confidence interval for $100(1-\alpha)$%

We don't need to restrict ourselves to 95% confidence. We can find the relevant $z_{\frac{\alpha}{2}}$ values (called critical values) from tables to construct a $100(1-\alpha)$% confidence interval for μ.

Where:



The 100(1-α)% confidence interval for μ is then:

$$\bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Section 4.3: Confidence interval for μ (variance unknown)

What about when we don't know σ?

In practice, we generally don't know σ, but in the previous section, we assumed that σ was known.

What to we do when σ is unknown?

We use the sample standard deviation, $S = \sqrt{\frac{\Sigma_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$, and replace σ by this estimate.

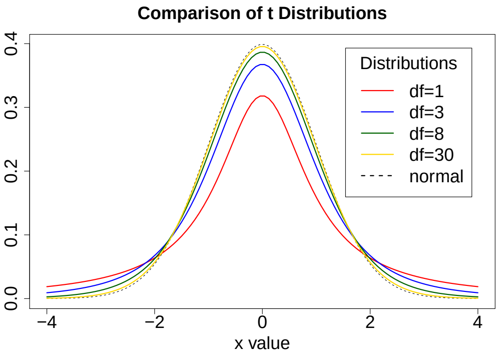
Comparing $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ with $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$, we know one less thing in the second one. Thus we have more uncertainty.

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

The t distribution

The t distribution has a heavier tail than the normal distribution.



As the degrees of freedom increase, the t-density curve approaches the N(0,1) density curve and the $t \approx N(0,1)$ curve for degrees of freedom ≥ 30.

Construct the CI

We are still assuming that $Y_1, Y_2, ..., Y_n \sim N(\mu, \sigma^2)$, but σ^2 is unknown.

We use the t-distribution to construct CIs.

We can write $t_{\nu,\frac{\alpha}{2}}$ to represent the critical values from the t distribution with $\nu = n-1$ the degrees of freedom.

$$P(-t_{\nu,\frac{\alpha}{2}} < \frac{\bar{Y}-\mu}{S/\sqrt{n}} < t_{\nu,\frac{\alpha}{2}}) = 1-\alpha$$

$$P(\bar{Y} - t_{\nu,\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{Y} + t_{\nu,\frac{\alpha}{2}}\frac{S}{\sqrt{n}}) = 1-\alpha$$

Therefore, a 100(1 - α)% confidence interval when σ isn't known is

$$\bar{y} \pm t_{\nu,\alpha/2}\frac{s}{\sqrt{n}}$$

where we read $t_{\nu,\alpha/2}$ from the t-tables.

Example

Let's go back to the example we did in Section 2 of the notes.

A pharmaceutical company has developed a drug they believe will help cure a particular disease and has acquired legal permission to test the drug on humans. However, there are safety concerns that the drug will have an adverse effect on blood pressure by increasing it. They run an experiment to estimate the average change in blood pressure for patients who take the new drug. They only have permission to test the drug on eleven patients.

Here are the data values:

```
x <- c(1.1, 1.8, 2, 2.4, 2.5, 2.8, 2.9, 3, 3.4, 3.4, 4)
x
```

```
##  [1] 1.1 1.8 2.0 2.4 2.5 2.8 2.9 3.0 3.4 3.4 4.0
```

Population parameter of interest: μ, the true mean change in blood pressure for the population of people who take the drug.

Assume that the data represent a random sample from the population of all people who take the drug.

Example contd.

95% confidence interval for μ: (2.11, 3.22)

This is computed as

$$\bar{x} \pm t_{\nu, \frac{0.05}{2}} \frac{s}{\sqrt{n}}$$

where $\bar{x} = 2.664$, $s = 0.8237$, $n = 11$, $t_{10,0.025} = 2.228$

Before we interpret the confidence, what assumptions have we made for the inference (interpretation) to be valid?

Recall: When we fit any statistical model to data, or perform any statistical analysis, we must:

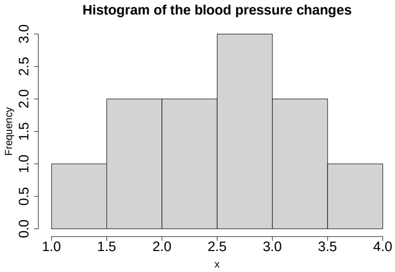
1. Know what (if any) assumptions are being made by the model or analysis.
2. Verify that those assumptions are reasonable.

Example contd.

We have assumed that the observations are independent of each other.

We have assumed that the population from which the observations were drawn is normally distributed.

- Is this reasonable?
- How might we check this?



We are 95% confident that μ, the true population mean change in blood pressure lies in the interval (2.11, 3.22).

Output from R

```
x
```

```
##  [1] 1.1 1.8 2.0 2.4 2.5 2.8 2.9 3.0 3.4 3.4 4.0
```

```
t.test(x, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  x
## t = 10.725, df = 10, p-value = 8.342e-07
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2.110241 3.217032
## sample estimates:
## mean of x
##  2.663636
```

Section 4.4: Hypothesis tests for μ (variance unknown)

Construction of a hypothesis test

We can carry out hypothesis tests for the population mean (μ) in the following way.

- Start by specifying the hypotheses. For example:

  H_0: $\mu = \mu_0$ versus H_A: $\mu \neq \mu_0$.

  This is a two-sided alternative hypothesis, one-sided would be:

  H_A: $\mu < \mu_0$ or H_A: $\mu > \mu_0$.

- Construct a test statistic.
- Evaluate the test statistic against the "null" distribution.
- Make a conclusion.

For now, we will assume $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ independent and that σ is unknown.

What is the logic behind a hypothesis test?

We assume that the null hypothesis, $\mu = \mu_0$, is true. We then examine the observed data to see if there is evidence to the contrary.

Computing the test statistic

If the H_0 is in fact true, then:

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

This is used as a reference to compare what we see in the data.

Calculate the observed test statistic (a realisation from the above distribution)

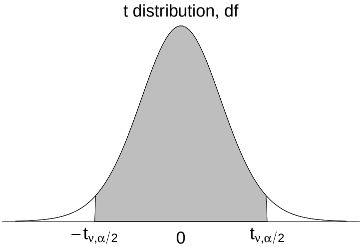
$$T_{obs} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

A test of level α will compare T_{obs} with the corresponding critical value(s) and based on the comparison we either:

- Reject the H_0 and accept the H_A, or
- Fail to reject the H_0 (which does not mean that we have proven it is true)

Evaluating the hypothesis test - two sided test

The sampling distribution of the test statistic T, with possible values along the x axis:

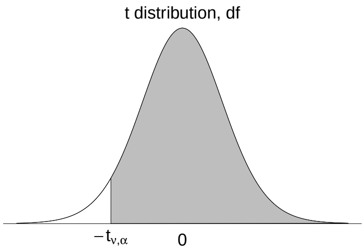


If an observed test statistic value T_{obs} lies along the region shaded in grey, this is 'typical' for this distribution and consistent with the null hypothesis. If the observed test statistic lies outside this, it is evidence for the alternative hypothesis, at the specified α level. Formally, for a two-sided test:

- Reject the H_0 and accept the H_A, when $T_{obs} \leq -t_{\nu,\alpha/2}$ or $T_{obs} \geq t_{\nu,\alpha/2}$.
- Fail to reject the H_0 when $-t_{\nu,\alpha/2} < T_{obs} < t_{\nu,\alpha/2}$.

Evaluating the hypothesis test for one-sided alternatives

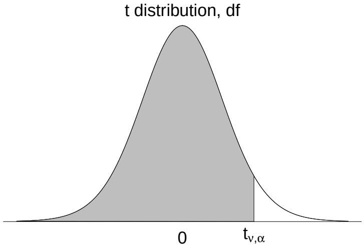
For a lower tailed test for a specified α level:



- Reject the H_0 and accept the H_A, when $T_{obs} \leq -t_{\nu,\alpha}$.
- Fail to reject the H_0 when $T_{obs} > -t_{\nu,\alpha}$.

Evaluating the hypothesis test for one-sided alternatives

For an upper tailed test for a specified α level:



- Reject the H_0 and accept the H_A, when $T_{obs} \geq t_{\nu,\alpha}$.
- Fail to reject the H_0 when $T_{obs} < t_{\nu,\alpha}$.

Additional notes

- It is possible that a null hypothesis is rejected, when in fact it is true. The probability of this happening here is α. We call this a Type I Error.

$$P(\text{Reject } H_0 \text{ when true}) = \alpha = P(\text{Type I Error})$$

- When specifying the hypotheses, the decision about what value μ_0 takes must be made **before** collecting and examining the data. We don't ever observe our sample mean, and then decide on a hypothesis to test.

- When it comes to specifying the alternative hypothesis, the default is to pick a two-sided test. If there is a reason before collecting the data for a belief that the true mean is in one direction, then a one-sided test can be done.

- NB: Don't ever decide on the value of μ_0, or decide to do a one-sided test based on the observed sample mean.

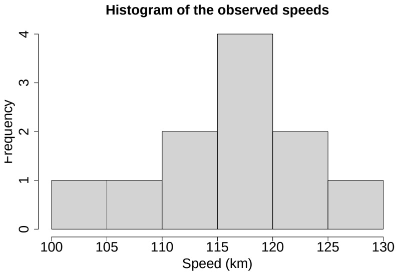
Example

There is a speed camera on a motorway where the speed limit is 120km. The camera observes the speed of 11 cars at random and records:

The 11 recordings are: 104, 109, 112, 114, 117, 118, 118, 120, 121, 124, 130.

Test the hypothesis that the true mean speed of vehicles on the motorway is under the speed limit at the $\alpha = 0.05$ level.

First, let's examine the data.



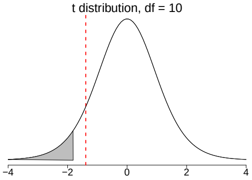
Is it reasonable to assume a normal population?

Example worked out

- Let μ be the mean population speed of all cars on the motorway.
- $H_0 : \mu = 120$ vs $H_A : \mu < 120$.
- Under the H_0, the test statistic $T \sim t(10)$. Find the observed test statistic:

$$T_{obs} = \frac{\bar{x} - \mu_0}{SE} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{117 - 120}{7.15542/\sqrt{11}} = -1.39$$

- The critical value is $t_{\nu=10,\alpha=0.05} = -1.812$.



- We fail to reject the H_0 ($\alpha = 0.05$, $t_{10,0.05} = -1.812 < -1.39$, lower-tailed test) and therefore have no evidence that the true mean speed of cars on the motorway is lower than 120, the speed limit.

Output from R

```
speeds <- c(104, 109, 112, 114, 117, 118, 118, 120, 121, 124, 130)
speeds
```

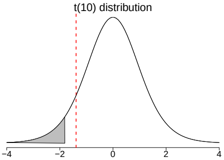
```
##  [1] 104 109 112 114 117 118 118 120 121 124 130
```

```
t.test(speeds, mu = 120,  alternative = "less")
```

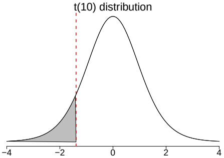
```
##
##  One Sample t-test
##
## data:  speeds
## t = -1.3905, df = 10, p-value = 0.09727
## alternative hypothesis: true mean is less than 120
## 95 percent confidence interval:
##      -Inf 120.9103
## sample estimates:
## mean of x
##       117
```

Using p-values or critical values to evaluate a hypothesis test

The critical value = -1.812; this means the P($t(10) <$ -1.812) = 0.05.



P-value = probability of observing a test statistic as extreme, or more extreme, under the null hypothesis. $T_{obs} = -1.39$. P-value = P($t(10) <$ -1.39) = 0.09727.



If the p-value $< \alpha$ we reject the H_0.

Section 4.5: Confidence intervals and hypothesis tests for σ^2

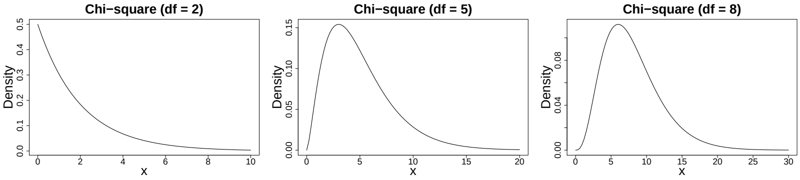
The chi-square distribution

If $Y_1, Y_2, ..., Y_n \sim N(\mu, \sigma^2)$, then we can use the following information to construct confidence intervals for σ^2:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(We will assume this without proof.)

The shape of the chi-square distribution depends on the degrees of freedom:



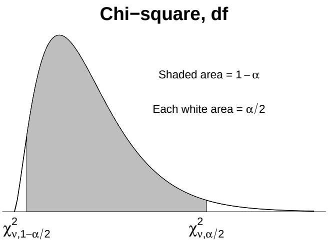
Construct a CI for σ^2

For degrees of freedom $\nu = n - 1$

$$P(\chi^2_{\nu,1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\nu,\alpha/2}) = 1 - \alpha$$

and a $100(1-\alpha)\%$ confidence interval for σ^2 can be shown to be

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



Hypothesis test for σ^2

We will follow the same general steps to do a hypothesis test as before. That is:

- Specify the hypotheses, for example

  H_0: $\sigma^2 = \sigma_0^2$ versus H_A: $\sigma^2 > \sigma_0^2$.
- Construct a test statistic.
- Evaluate the test statistic against the "null" distribution.
- Make a conclusion.

If the H_0 is true, then,

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

and for a given sample, the test statistic

$$\chi^2_{obs} = \frac{(n-1)s^2}{\sigma_0^2}$$

is a realisation of a $\chi^2(n-1)$ random variable.

Evaluating the test statistic

If H_0: $\sigma^2 = \sigma_0^2$ versus H_A: $\sigma^2 > \sigma_0^2$, and

the test statistic has been computed:

$$\chi^2_{obs} = \frac{(n-1)s^2}{\sigma_0^2}$$

We evaluate the test statistic against the "null" distribution. Under the H_0, the test statistic comes from a $\chi^2(n-1)$.

Here are the possible outcomes (with degrees of freedom $\nu = n - 1$):

- We reject the H_0 and accept the H_A if $\chi^2_{obs} \geq \chi^2_{\nu,\alpha}$.
- We fail to reject the H_0 if $\chi^2_{obs} < \chi^2_{\nu,\alpha}$.

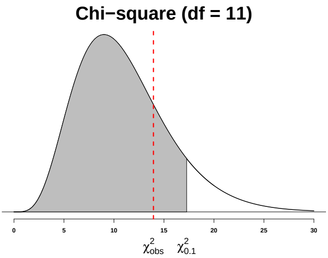
Example

A sample of data of size 12 was collected from a normally distributed population. The standard deviation was computed as 4.362.

We wish to test the hypothesis that H$_0$: $\sigma^2 = 15$ versus H$_A$: $\sigma^2 > 15$, using $\alpha = 0.1$.

$$\chi^2_{obs} = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(12-1)4.362^2}{15} = 13.95$$

The test statistic $\chi^2_{obs} = 13.95$ which is $< \chi^2_{0.1}(11) = 17.275$. At $\alpha = 0.1$, we fail to reject the H$_0$ and have no evidence that $\sigma^2 > 15$.



Final word

- Up to now, when you have done hypothesis testing or used confidence intervals, you have taken the methods at face value.
- Now, you know about the underlying theory.

- In this section, we have focused on data that is from populations that are normally distributed.
- In the next section, we will explore beyond that.