Lecture 5: Hypothesis Testing

What we know now:

OLS is not only unbiased it is also the most precise (efficient) unbiased estimation technique
- ie the estimator has the smallest variance

(if the Gauss-Markov assumptions hold)
We also know that:

**OLS estimates given by**

\[ \hat{b}_0 = \bar{Y} - b_1 \bar{X} \]

\[ \hat{b}_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \]

**OLS estimates of precision given by**

\[ \text{Var}(\hat{\beta}_0) = \frac{s^2}{N} \left[ 1 + \frac{2}{\bar{X} \text{Var}(X)} \right] \]

\[ \text{Var}(\hat{\beta}_1) = \frac{s^2}{N \cdot \text{Var}(X)} \]

\[ s^2 = \frac{\text{RSS}}{N - k} \]

At same time usual to work with the square root to give **standard errors** of the estimates (**standard deviation** refers to the known variance, **standard error** refers to the **estimated** variance)

\[ \text{s.e.}(\hat{\beta}_0) = \sqrt{\frac{s^2}{N} \left( 1 + \frac{\bar{X}^2}{\text{Var}(X)} \right)} \]

\[ \text{s.e.}(\hat{\beta}_1) = \sqrt{\frac{s^2}{N \cdot \text{Var}(X)}} \]
Now need to learn how to test hypotheses about whether the values of the individual estimates we get are consistent with the way we think the world may work.

Eg if we estimate the model by OLS

\[ C = b_0 + b_1 Y + u \]

and the estimated coefficient on income is \( b_1 = 0.9 \) is this consistent with what we think the marginal propensity to consume should be?

Now common (economic) sense can help here, but in itself it is not rigorous (or impartial) enough.

which is where hypothesis testing comes in
Hypothesis Testing

If wish to make inferences about how close an estimated value is to a hypothesised value or even to say whether the influence of a variable is not simply the result of statistical chance (the estimates being based on a sample and so subject to random fluctuations)

then need to make one additional assumption about the behaviour of the (true, unobserved) residuals in the model
Hypothesis Testing

If wish to make inferences about how close an estimated value is to a hypothesised value or even to say whether the influence of a variable is not simply the result of statistical chance then need to make one additional assumption about the behaviour of the (true, unobserved) residuals in the model (the estimates being based on a sample and so subject to random fluctuations)

then need to make one additional assumption about the behaviour of the (true, unobserved) residuals in the model

We know already that \( u_i \sim (0, \sigma^2_u) \)

ie true residuals assumed to have a mean of zero and variance \( \sigma^2_u \)
Hypothesis Testing

If wish to make inferences about how close an estimated value is to a hypothesised value or even to say whether the influence of a variable is not simply the result of statistical chance then need to make one additional assumption about the behaviour of the (true, unobserved) residuals in the model (the estimates being based on a sample and so subject to random fluctuations)

then need to make one additional assumption about the behaviour of the (true, unobserved) residuals in the model

We know already that \( u_i \sim (0, \sigma^2_u) \)

ie true residuals assumed to have a mean of zero and variance \( \sigma^2_u \)

Now assume additionally that residuals follow a Normal distribution
Now assume additionally that residuals follow a Normal distribution

\[ u_i \sim N(0, \sigma^2_u) \]
Now assume additionally that residuals follow a **Normal distribution**

\[ u_i \sim \mathcal{N}(0, \sigma^2_u) \]

(Since residuals capture influence of many unobserved (random) variables, can use **Central Limit Theorem** which says that the sum of a large set of random variables will have a normal distribution)
If \( u \) is normal, then it is easy to show that the OLS coefficients (which are a linear function of \( u \)) are also normally distributed with the means and variances that we derived earlier. So

\[
\hat{\beta}_0 \sim N(\beta_0, Var(\beta_0)) \quad \text{and} \quad \hat{\beta}_1 \sim N(\beta_1, Var(\beta_1))
\]
If \( u \) is normal, then it is easy to show that the OLS coefficients (which are a linear function of \( u \)) are also normally distributed with the means and variances that we derived earlier. So

\[
\hat{\beta}_0 \sim N(\beta_0, Var(\beta_0)) \quad \text{and} \quad \hat{\beta}_1 \sim N(\beta_1, Var(\beta_1))
\]

Why is that any use to us?
If a variable is normally distributed we know that it is Symmetric.
If a variable is normally distributed we know that it is Symmetric.
If a variable is normally distributed we know that it is Symmetric centered on its mean (use the symbol $\mu$) and that the symmetry pattern is such that: 66% of values lie within mean $\pm$ 1*standard deviation
If a variable is normally distributed we know that it is

Symmetric

centred on its mean \( \mu \)  

and that the symmetry pattern is such that:

66\% of values lie within \( \text{mean} \pm 1 \times \text{standard deviation} \)  

(\text{use the symbol } \sigma )
If a variable is normally distributed we know that it is
Symmetric

centred on its mean \( \mu \) (use the symbol \( \mu \))

and that the symmetry pattern is such that:

66\% of values lie within mean \( \pm 1 \times \) standard deviation
(\( \sigma \))

95\% of values lie within mean \( \pm 1.96 \times \sigma \)
If a variable is normally distributed we know that it is

Symmetric

centred on its mean (use the symbol \( \mu \))

and that the symmetry pattern is such that:

66% of values lie within mean \( \pm 1 \times \text{standard deviation} \)

(use the symbol \( \sigma \))

95% of values lie within mean \( \pm 1.96 \times \sigma \)

99% of values lie within mean \( \pm 2.9 \times \sigma \)
The graph shows the probability density function $f(x)$ of a normal distribution with mean $\mu$ and standard deviation $\sigma$. The function is given by:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
Easier to work with the standard normal distribution which has a mean ($\mu$) of 0 and variance ($\sigma^2_u$) of 1

$$u_i \sim N(0, 1)$$
Easier to work with the **standard normal distribution**, $z$ which has a mean ($\mu$) of 0 and variance ($\sigma^2_u$) of 1

$$u_i \sim N(0, 1)$$

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation
Easier to work with the **standard normal distribution**, $z$ which has a mean ($\mu$) of 0 and variance ($\sigma^2_u$) of 1

$$u_i \sim \text{N}(0, 1)$$

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation

ie Given $\beta_1 \sim N(\beta_1, \text{Var}(\beta_1))$
Easier to work with the **standard normal distribution**, \( z \) which has a mean (\( \mu \)) of 0 and variance (\( \sigma^2 \)) of 1

\[
u_i \sim N(0, 1)
\]

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation

ie Given \( \beta_1 \sim N(\beta_1, Var(\beta_1)) \) then

\[
z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1)
\]
Easier to work with the standard normal distribution which has a mean of 0 and variance of 1

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation

\[
\hat{z} = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.(\beta_1)}}
\]

\[\hat{\beta}_1 \sim N(\beta_1, \text{Var}(\beta_1))\]

Since the mean of this variable is zero and because a normal distribution is symmetric and centred around its mean, and the standard deviation=1, we know that
Easier to work with the standard normal distribution which has a mean of 0 and variance of 1

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation

\[
\beta_1 \sim N(\hat{\beta}_1, Var(\hat{\beta}_1)) \text{ then } z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1)
\]

Since the mean of this variable is zero and because a normal distribution is symmetric and centred around its mean, and the standard deviation=1, we know that

66% of values lie within mean $\pm 1*\sigma$
Easier to work with the **standard normal distribution** which has a mean of 0 and variance of 1

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation

\[
\beta_1 \sim N(\hat{\beta}_1, Var(\hat{\beta}_1)) \implies z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1)
\]

Since the mean of this variable is zero and because a normal distribution is symmetric and centred around its mean, and the standard deviation=1, we know that

66% of values lie within mean $\pm 1*\sigma$

so now

66% of values lie within 0 $\pm 1$
Easier to work with the **standard normal distribution** which has a mean of 0 and variance of 1

- this can be obtained from any normal distribution by subtracting the mean and dividing by the standard deviation

\[ z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1) \]

Since the mean of this variable is zero and because a normal distribution is symmetric and centred around its mean, and the standard deviation=1, we know that

66% of values lie within mean ± 1*σ

so now 66% of values lie within 0 ± 1

Similarly

95% of values lie within mean ± 1.96*σ
99% of values lie within mean $\pm 2.9\sigma$
(these thresholds, ie 1, 1.96 and 2.9, are called the “critical values”)

A graph of a normal distribution with zero mean and unit variance

95% of values lie within 0 ± 1.96
so the *critical value* for the 95% significance level is 1.96
Can use all this to test hypotheses about the values of individual coefficients

**Testing A Hypothesis Relating To A Regression Coefficient**

Model: $Y = \beta_0 + \beta_1 X + u$
Testing A Hypothesis Relating To A Regression Coefficient

Model: \( Y = \beta_0 + \beta_1 X + u \)

Null hypothesis: \( H_0: \beta_1 = \beta_1^0 \)
Testing A Hypothesis Relating To A Regression Coefficient

Model: \( Y = \beta_0 + \beta_1 X + u \)

Null hypothesis: \( H_0: \beta_1 = \beta_1^0 \)

(which says that we think the true value is equal to a specific value)
Testing A Hypothesis Relating To A Regression Coefficient

Model: $Y = \beta_0 + \beta_1 X + u$

Null hypothesis: $H_0: \beta_1 = \beta_1^0$

(which says that we think the true value is equal to a specific value)

Alternative hypothesis: $H_1: \beta_1 \neq \beta_1^0$
Testing A Hypothesis Relating To A Regression Coefficient

Model: \[ Y = \beta_0 + \beta_1 X + u \]

Null hypothesis: \[ H_0: \beta_1 = \beta_1^0 \]

(which says that we think the true value is equal to a specific value)

Alternative hypothesis: \[ H_1: \beta_1 \neq \beta_1^0 \]

(which says that the true value is not equal to the specific value)
Testing A Hypothesis Relating To A Regression Coefficient

Model: \( Y = \beta_0 + \beta_1 X + u \)

Null hypothesis: \[ H_0: \beta_1 = \beta_{10} \]

(which says that we think the true value is equal to a specific value)

Alternative hypothesis: \[ H_1: \beta_1 \neq \beta_{10} \]

(which says that the true value is not equal to the specific value)

But because OLS is only an estimate we have to allow for the uncertainty in the estimate - as captured by its standard error
Testing A Hypothesis Relating To A Regression Coefficient

Model: \( Y = \beta_0 + \beta_1 X + u \)

Null hypothesis: \( H_0: \beta_1 = \beta_1^0 \)

(with which says that we think the true value is equal to a specific value)

Alternative hypothesis: \( H_1: \beta_1 \neq \beta_1^0 \)

(with which says that the true value is not equal to the specific value)

But because OLS is only an estimate we have to allow for the uncertainty in the estimate – as captured by its standard error

Implicitly the variance says the estimate may be centred on this value but because this is only an estimate and not the truth, there is a possible range of other plausible values associated with it
Testing A Hypothesis Relating To A Regression Coefficient

The most common hypothesis to be tested (and most regression packages default hypothesis) is that the true value of the coefficient is zero.
Testing A Hypothesis Relating To A Regression Coefficient

The most common hypothesis to be tested (and most regression packages default hypothesis) is that the true value of the coefficient is zero

- Which is the same thing as saying the effect of the variable is zero
  Since $\beta_1 = \frac{dY}{dX}$
Testing A Hypothesis Relating To A Regression Coefficient

The most common hypothesis to be tested (and most regression packages default hypothesis) is that the true value of the coefficient is zero

- Which is the same thing as saying the effect of the variable is zero
  Since $\beta_1 = dY/dX$

Example

Cons = $\beta_0 + \beta_1\text{Income} + u$
Testing A Hypothesis Relating To A Regression Coefficient

The most common hypothesis to be tested (and most regression packages default hypothesis) is that the true value of the coefficient is zero

- Which is the same thing as saying the effect of the variable is zero

Since $\beta_1 = dY/dX$

Example

Cons = $\beta_0 + \beta_1$Income + u

Null hypothesis: $H_0: \beta_1 = 0$
Testing A Hypothesis Relating To A Regression Coefficient

The most common hypothesis to be tested (and most regression packages default hypothesis) is that the true value of the coefficient is zero

- Which is the same thing as saying the effect of the variable is zero
  Since $\beta_1 = \frac{dY}{dX}$

Example

$$\text{Cons} = \beta_0 + \beta_1 \text{Income} + u$$

Null hypothesis: \quad H0: $\beta_1 = 0$

Alternative hypothesis: \quad H1: $\beta_1 \neq 0$
Testing A Hypothesis Relating To A Regression Coefficient

The most common hypothesis to be tested (and most regression packages default hypothesis) is that the true value of the coefficient is zero

- Which is the same thing as saying the effect of the variable is zero
  Since $\beta_1 = \frac{dY}{dX}$

Example

$\text{Cons} = \beta_0 + \beta_1 \text{Income} + u$

Null hypothesis: $H_0: \beta_1 = 0$

Alternative hypothesis: $H_1: \beta_1 \neq 0$

and this time $\beta_1 = \frac{d\text{Cons}}{d\text{Income}}$
In order to be able to say whether OLS estimate is close enough to hypothesized value so as to be acceptable, we take the range of estimates implied by the estimated OLS variance and look to see whether this range will contain the hypothesized value.
In order to be able to say whether OLS estimate is close enough to hypothesized value so as to be acceptable, we take the range of estimates implied by the estimated OLS variance and look to see whether this range will contain the hypothesized value.

To do this we can use the range of estimates implied by the standard normal distribution
In order to be able to say whether OLS estimate is close enough to hypothesized value so as to be acceptable, we take the range of estimates implied by the estimated OLS variance and look to see whether this range will contain the hypothesized value.

To do this we can use the range of estimates implied by the standard normal distribution

So given we now know $\hat{\beta}_1 \sim N(\beta_1, Var(\beta_1))$
In order to be able to say whether OLS estimate is close enough to hypothesized value so as to be acceptable, we take the range of estimates implied by the estimated OLS variance and look to see whether this range will contain the hypothesized value.

To do this we can use the range of estimates implied by the standard normal distribution

\[
\beta_1 \sim N(\beta_1, \text{Var}(\beta_1))
\]

So given we now know \( \beta_1 \sim N(\beta_1, \text{Var}(\beta_1)) \),

then we can transform the estimates into a standard normal

\[
z = \frac{\beta_1 - \beta_1^\wedge}{\text{s.d.}(\beta_1)} \sim N(0,1)
\]
In order to be able to say whether OLS estimate is close enough to hypothesized value so as to be acceptable, we take the range of estimates implied by the estimated OLS variance and look to see whether this range will contain the hypothesized value.

To do this we can use the range of estimates implied by the standard normal distribution

So given we now know \( \beta_1 \sim N(\beta_1, \text{Var}(\beta_1)) \)

then we can transform the estimates into a standard normal

\[
z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1)
\]

and we know that 95% of all values of a variable that has mean 0 and variance 1 will lie within \( 0 \pm 1.96 \)
We know 95% of all values of a variable that has mean 0 and variance 1 will lie within

\[0 \pm 1.96\]

(95 times out of a 100, estimates will lie in the range \(0\pm 1.96\times\text{standard deviation}\))

or written another way
We know 95% of all values of a variable that has mean 0 and variance 1 will lie within

$$0 \pm 1.96$$

(95 times out of a 100, estimates will lie in the range 0± 1.96*standard deviation)

or written another way

$$\Pr[-1.96 \leq z \leq 1.96] = 0.95$$
We know 95% of all values of a variable that has mean 0 and variance 1 will lie within

\[ 0 \pm 1.96 \]

(95 times out of a 100, estimates will lie in the range 0±1.96*standard deviation)

or written another way

\[ \Pr[-1.96 \leq z \leq 1.96] = 0.95 \]

Since

\[ z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1) \]

sub. this into \( \Pr[-1.96 \leq z \leq 1.96] = 0.95 \)
Since \( z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1) \)

Substitute this into \( \Pr[-1.96 \leq z \leq 1.96] = 0.95 \)

\[
\Pr\left[ -1.96 \leq \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \leq 1.96 \right] = 0.95
\]
Since \( z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1) \)

sub. this into \( \Pr[-1.96 \leq z \leq 1.96] = 0.95 \)

\[
\Pr\left[ -1.96 \leq \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \leq 1.96 \right] = 0.95
\]

or equivalently multiplying the terms in square brackets in by \( s.d.(\beta_1) \)
Since \( z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1) \)

sub. this into \( \Pr[-1.96 \leq z \leq 1.96] = 0.95 \)

\[
\Pr\left[-1.96 \leq \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \leq 1.96\right] = 0.95
\]

or equivalently multiplying the terms in square brackets in by \( s.d.(\beta_1) \)

\( -1.96 \times s.d.(\beta_1) \leq \beta_1 - \beta_1 \)
Since \( z = \frac{\^\beta_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1) \)

sub. this into \( \Pr[-1.96 \leq z \leq 1.96] = 0.95 \)

\[
\begin{bmatrix}
-1.96 \leq \frac{\^\beta_1 - \beta_1}{s.d.(\beta_1)} \leq 1.96
\end{bmatrix} = 0.95
\]

or equivalently multiplying the terms in square brackets in by \( s.d.(\beta_1) \)

\( -1.96 \times s.d.(\beta_1) \leq \^\beta_1 - \beta_1 \) and \( 1.96 \times s.d.(\beta_1) \leq \^\beta_1 - \beta_1 \)

\( \beta_1 - \beta_1 \leq 1.96 \times s.d.(\beta_1) \)
Since \( z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1) \)

sub. this into \( \Pr[-1.96 \leq z \leq 1.96] = 0.95 \)

\[
\Pr\left[-1.96 \leq \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \leq 1.96 \right] = 0.95
\]

or equivalently multiplying the terms in square brackets in by \( \text{s.d.}(\beta_1) \)

\[
-1.96 \times \text{s.d.}(\beta_1) \leq \hat{\beta}_1 - \beta_1 \quad \text{and} \quad 1.96 \times \text{s.d.}(\beta_1) \leq \hat{\beta}_1 - \beta_1 \leq 1.96 \times \text{s.d.}(\beta_1)
\]
and taking $\beta_1$ to the other sides of the equality gives

$$z = \frac{\beta_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1)$$

sub. this into $\Pr[-1.96 \leq z \leq 1.96] = 0.95$

$$\Pr\left[-1.96 \leq \frac{\beta_1 - \beta_1}{s.d.(\beta_1)} \leq 1.96\right] = 0.95$$

or equivalently multiplying the terms in square brackets in by $s.d.(\beta_1)$
\[-1.96 \times \text{s.d.}(\beta_1) \leq \beta_1 - \beta_1 \quad \text{and} \quad 1.96 \times \text{s.d.}(\beta_1) \leq \beta_1 - \beta_1 \leq 1.96 \times \text{s.d.}(\beta_1)\]

and taking $\beta_1$ to the other sides of the equality gives

$$\Pr\left[\beta_1 - 1.96 \times \text{s.d.}(\beta_1) \leq \beta_1 \leq \beta_1 + 1.96 \times \text{s.d.}(\beta_1)\right] = 0.95$$
Pr\left[\hat{\beta}_1 - 1.96 \cdot s.d.(\hat{\beta}_1) \leq \hat{\beta}_1 \leq \hat{\beta}_1 + 1.96 \cdot s.d.(\hat{\beta}_1)\right] = 0.95

is called the 95\% confidence interval
\[
\Pr\left[\hat{\beta}_1 - 1.96 \times \text{s.d.}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + 1.96 \times \text{s.d.}(\hat{\beta}_1)\right] = 0.95
\]

is called the 95% confidence interval and says that given an OLS estimate and its standard deviation we can be 95% confident that the true (unknown) value for \( \beta_1 \) will lie in this region.
Pr\left[ \hat{\beta}_1 - 1.96 \times s.d.(\hat{\beta}_1) \leq \hat{\beta}_1 \leq \hat{\beta}_1 + 1.96 \times s.d.(\hat{\beta}_1) \right] = 0.95

is called the 95\% confidence interval

and says that given an OLS estimate and its standard deviation we can be 95\% confident that the true (unknown) value for \( \beta_1 \) will lie in this region

(If an estimate falls within this range it is said to lie in the acceptance region)
\[
\Pr \left[ \beta_1 - 1.96 \times s.d.(\beta_1) \leq \beta_1 \leq \beta_1 + 1.96 \times s.d.(\beta_1) \right] = 0.95
\]

is called the 95\% confidence interval

and says that given an OLS estimate and its standard deviation we can be 95\% confident that the true (unknown) value for $\beta_1$ will lie in this region

(If an estimate falls within this range it is said to lie in the acceptance region)

**Unfortunately** we never know the true standard deviation of $\beta_1$, only ever have an estimate, the standard error
\[
\Pr \left[ \beta_1 - 1.96 \times \text{s.d.}(\beta_1) \leq \beta_1 \leq \beta_1 + 1.96 \times \text{s.d.}(\beta_1) \right] = 0.95
\]

Unfortunately we never know the true standard deviation of \( \beta_1 \), only ever have an estimate, the standard error

\[
\text{s.d.}(\beta_1) = \sqrt{\frac{\sigma^2}{N \times \text{Var}(X)}}
\]

ie don’t have

since \( \sigma^2 \) (the true residual variance) is unobserved
\[ \Pr\left[ \hat{\beta}_1 - 1.96 \times s.d.(\beta_1) \leq \beta_1 \leq \hat{\beta}_1 + 1.96 \times s.d.(\beta_1) \right] = 0.95 \]

Unfortunately we never know the true standard deviation of \( \beta_1 \), only ever have an estimate, the standard error

ie don’t have \( s.e.(\beta_1) = \sqrt{\frac{\sigma^2}{N \times \text{Var}(X)}} \)

since \( \sigma^2 \) (the true residual variance) unobserved

but we now know that

\[ s.e.(\beta_1) = \sqrt{\frac{s^2}{N \times \text{Var}(X)}} \]
\[ \Pr \left[ \beta_1 - 1.96 \times s.d.(\beta_1) \leq \beta_1 \leq \beta_1 + 1.96 \times s.d.(\beta_1) \right] = 0.95 \]

Unfortunately we never know the true standard deviation of \( \beta_1 \), only ever have an estimate, the standard error

\[ \text{ie don't have } s.d.(\beta_1) = \sqrt{\frac{\sigma^2}{N \times \text{Var}(X)}} \]

since \( \sigma^2 \) (the true residual variance) unobserved

but we now know that

\[ s.e.(\beta_1) = \sqrt{\frac{s^2}{N \times \text{Var}(X)}} \quad \text{(replacing } \sigma^2 \text{ with } s^2) \]

so can substitute this into the equation above
\[
\Pr \left[ \hat{\beta}_1 - 1.96 \times \text{s.d.}(\beta_1) \leq \beta_1 \leq \hat{\beta}_1 + 1.96 \times \text{s.d.}(\beta_1) \right] = 0.95
\]

Unfortunately we never know the true standard deviation of \( \beta_1 \), only ever have an estimate, the standard error

\[
\text{s.d.}(\beta_1) = \sqrt{\frac{\sigma^2}{N \times \text{Var}(X)}}
\]

since \( \sigma^2 \) (the true residual variance) unobserved

but we now know that

\[
\text{s.e.}(\beta_1) = \sqrt{\frac{s^2}{N \times \text{Var}(X)}}
\]

(replacing \( \sigma^2 \) with \( s^2 \))

so can substitute this into the equation above
\[
\Pr\left[\hat{\beta}_1 - 1.96 \times s.e.(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + 1.96 \times s.e.(\hat{\beta}_1)\right] = 0.95
\]
However when we do this we no longer have a standard normal distribution.
However when we do this we no longer have a standard normal distribution and this has implications for the 95% critical value thresholds in the equation

$$\Pr\left[\hat{\beta}_1 - 1.96 \times \text{s.e.}(\hat{\beta}_1) \leq \hat{\beta}_1 \leq \hat{\beta}_1 + 1.96 \times \text{s.e.}(\hat{\beta}_1)\right] = 0.95$$
ie no longer \( z = \frac{\beta_1 - \beta_1}{\hat{\sigma}(\beta_1)} \sim N(0,1) \)
\[ z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1) \]

but instead the statistic
However when we do this we no longer have a standard normal distribution,

\[ z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0, 1) \]

but instead the statistic

\[ t = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\beta_1)} \]
ie no longer \( z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1) \)

but instead the statistic

\[
\hat{t} = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\beta_1)} = \frac{\hat{\beta}_1 - \beta_1 \sqrt{N \cdot Var(X)}}{s} \sim t(N - k)
\]
ie no longer \( z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1) \)

but instead the statistic

\[
t = \frac{\hat{\beta}_1 - \beta_1}{\text{s.e.}(\beta_1)} = \frac{\hat{\beta}_1 - \beta_1 \sqrt{N \cdot \text{Var}(X)}}{s} \sim t(N - k)
\]

The "\( \sim \)" means the statistic is said to follow a \( t \) distribution with \( N-k \) "degrees of freedom"
\[
\text{ie no longer } z = \frac{\beta_1 - \beta_1}{s.d.(\beta_1)} \sim N(0,1)
\]

but instead the statistic

\[
t = \frac{\beta_1 - \beta_1}{s.e.(\beta_1)} = \frac{\beta_1 - \beta_1 \sqrt{N*\text{Var}(X)}}{s} \sim t(N - k)
\]

The “\( \sim \)” means the statistic is said to follow a t distribution with \( N-k \) “degrees of freedom”

\( N = \text{sample size} \)
ie no longer \( z = \frac{\hat{\beta}_1 - \beta_1}{\text{s.d.}(\beta_1)} \sim N(0,1) \)

but instead the statistic

\[
t = \frac{\hat{\beta}_1 - \beta_1}{\text{s.e.}(\beta_1)} = \frac{\hat{\beta}_1 - \beta_1}{s} \sqrt{\frac{N \cdot \text{Var}(X)}{\text{s}}}
\]

\( \sim t(N-k) \)

The " ~ " means the statistic is said to follow a t distribution with N-k "degrees of freedom"

N = sample size
k = no. of right hand side coefficients in the model
(so includes the constant)
Student's t distribution introduced by
William Gosset 1976-1937 and developed by Ronald Fisher 1890-1962
Now this t distribution is symmetrical about its mean but has its own set of critical values at given significance levels.

When the number of degrees of freedom is large, the t distribution looks very much like a normal distribution (and as the number increases, it converges on one).
Now this $t$ distribution is symmetrical about its mean but has its own set of critical values at given significance levels which vary, unlike the standard normal distribution, with the degrees of freedom in the model,
Now this t distribution is symmetrical about its mean but has its own set of critical values at given significance levels which vary, unlike the standard normal distribution, with the degrees of freedom in the model,

N-K
Now this t distribution is symmetrical about its mean but has its own set of **critical values** at given significance levels which vary, unlike the standard normal distribution, with the degrees of freedom in the model, N-K which means the “degrees of freedom” depend on the sample size and the number of variables in the model)
Now this \( t \) distribution is symmetrical about its mean but has its own set of \textbf{critical values} at given significance levels which vary, unlike the standard normal distribution, with the degrees of freedom in the model, \( N-K \) which means the “degrees of freedom” depend on the sample size and the number of variables in the model.

Also since the true mean is unknown we can replace it with a hypothesized value and still have a \( t \) distribution.
Now this t distribution is symmetrical about its mean but has its own set of **critical values** at given significance levels which vary, unlike the standard normal distribution, with the degrees of freedom in the model, \( N-K \) which means the “degrees of freedom” depend on the sample size and the number of variables in the model.

Also since the true mean is unknown we can replace it with a hypothesized value and still have a t distribution

\[
\text{So } t = \frac{\hat{\beta}_1 - \beta_1}{\text{s.e.} (\beta_1)}
\]
Now this t distribution is symmetrical about its mean but has its own set of **critical values** at given significance levels which vary, unlike the standard normal distribution, with the degrees of freedom in the model, \( N-K \) which means the “degrees of freedom” depend on the sample size and the number of variables in the model.

Also since the true mean is unknown we can replace it with any hypothesized value and still have a t distribution

\[
\hat{t} = \frac{\hat{\beta}_1 - \beta_1}{\text{s.e.}(\beta_1)}
\]

and also

\[
\hat{t} = \frac{\hat{\beta}_1 - \beta_0}{\text{s.e.}(\beta_1)} \sim t_{(N-k)}
\]