Hypothesis Testing

Now know how to estimate and interpret OLS regression coefficients
We know also that OLS gives us unbiased and efficient (smallest variance) estimates

Hypothesis Testing

Now know how to estimate and interpret OLS regression coefficients
We know also that OLS gives us unbiased and efficient (smallest variance) estimates

But 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

Hypothesis Testing

Now know how to estimate and interpret OLS regression coefficients
We know also that OLS gives us unbiased and efficient (smallest variance) estimates

But 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

We know already that $u_i \sim (0, \sigma^2_u)$
ie true residuals assumed to have a mean of zero and variance σ^2_u

Hypothesis Testing

Now know how to estimate and interpret OLS regression coefficients
We know also that OLS gives us unbiased and efficient (smallest variance) estimates

But 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

We know already that $u_i \sim (0, \sigma^2_u)$
ie true residuals assumed to have a mean of zero and variance σ^2_u

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

Hypothesis Testing

Now know how to estimate and interpret OLS regression coefficients
We know also that OLS gives us unbiased and efficient (smallest variance) estimates

But 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

We know already that $u_i \sim (0, \sigma^2_u)$
ie true residuals assumed to have a mean of zero and variance σ^2_u

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

(Since residuals capture influence of many unobserved (random) variables, can use **Central Limit Theorem** which says that the sum of a set of random variables will have a normal distribution)

Hypothesis Testing

Now know how to estimate and interpret OLS regression coefficients
We know also that OLS gives us unbiased and efficient (smallest variance) estimates

But 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

We know already that $u_i \sim (0, \sigma^2_u)$
ie true residuals assumed to have a mean of zero and variance σ^2_u

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

(Since residuals capture influence of many unobserved (random) variables, can use **Central Limit Theorem** which says that the sum of a set of random variables will have a normal distribution)

and 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

$\hat{\beta}_0 \sim N(\beta_0, Var(\hat{\beta}_0))$ and $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$

If a variable is normally distributed we know that it is

If a variable is normally distributed we know that it is

Symmetric

If a variable is normally distributed we know that it is

Symmetric

centred on its mean

If a variable is normally distributed we know that it is

Symmetric

centred on its mean

and 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

and that:

66% of values lie within mean $\pm$1*standard deviation
95% of values lie within mean $\pm$1.96*standard dev.

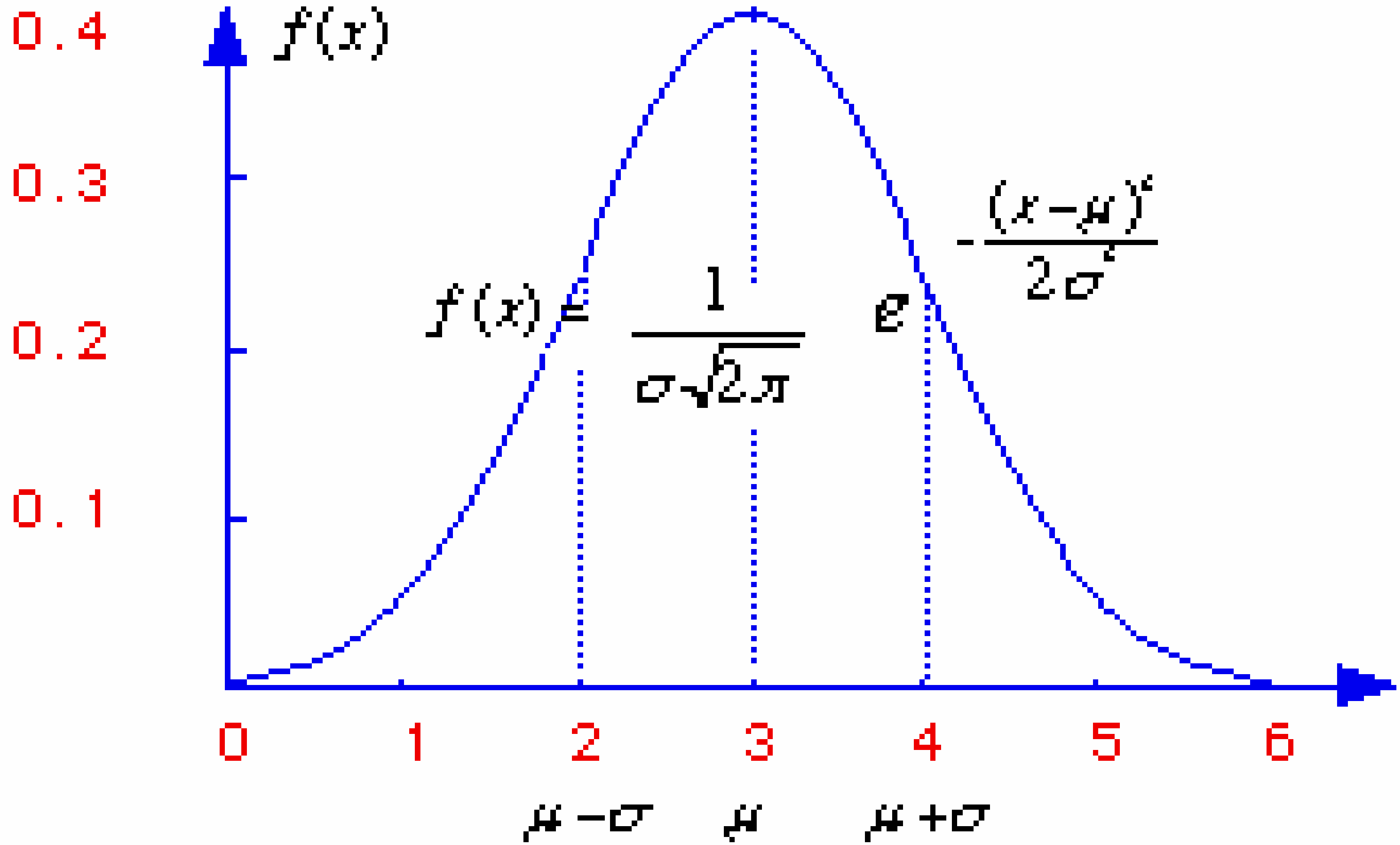
If a variable is normally distributed we know that it is

Symmetric

centred on its mean

and that:

66% of values lie within	mean $\pm$1*standard deviation
95% of values lie within	mean $\pm$1.96*standard dev.
99% of values lie within	mean $\pm$2.9*standard dev.



0.4
0.3
0.2
0.1
f(x)
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
0
1
2
3
4
5
6
μ − σ
μ
μ + σ

Easier to work with the *standard normal distribution* which has a mean of 0 and variance of 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

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $$z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$$

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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 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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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 0 ± 1
95% of values lie within 0 ± 1.96

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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 0 ± 1
95% of values lie within 0 ± 1.96
99% of values lie within 0 ± 2.9

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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 0 $\pm$1
95% of values lie within 0 $\pm$1.96
99% of values lie within 0 $\pm$2.9

(these thresholds, ie 1, 1.96 and 2.9, are called the “critical values”)

Lecture 5

Interested in testing hypotheses about likely value of a variable
(eg the marginal propensity to consume =0.9)

To do this need to understand the meaning of t values and confidence intervals

To do this need to understand a little about the standard normal distribution

If a variable is normally distributed we know that it is

Symmetric

centred on its mean

and that:

66% of values lie within	mean $\pm$1*standard deviation
95% of values lie within	mean $\pm$1.96*standard dev.
99% of values lie within	mean $\pm$2.9*standard dev.

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

ie Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

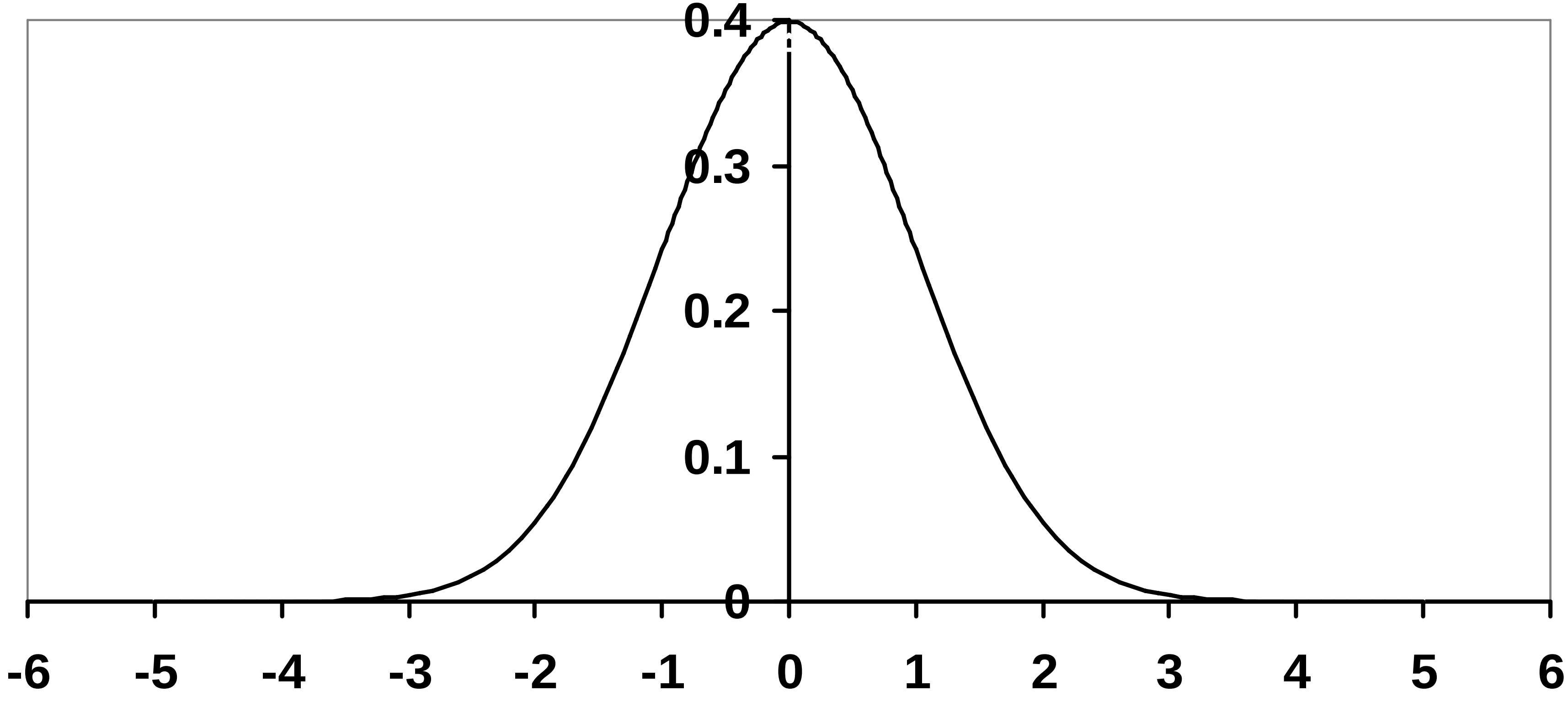
Since the mean of this variable is zero and because a normal distribution is symmetric and centred around its mean, and the standard deviation=1, we know that

66% of values lie within 0 ± 1
95% of values lie within 0 ± 1.96
99% of values lie within 0 ± 2.9

(these thresholds, ie 1, 1.96 and 2.9, are called the “critical values”)

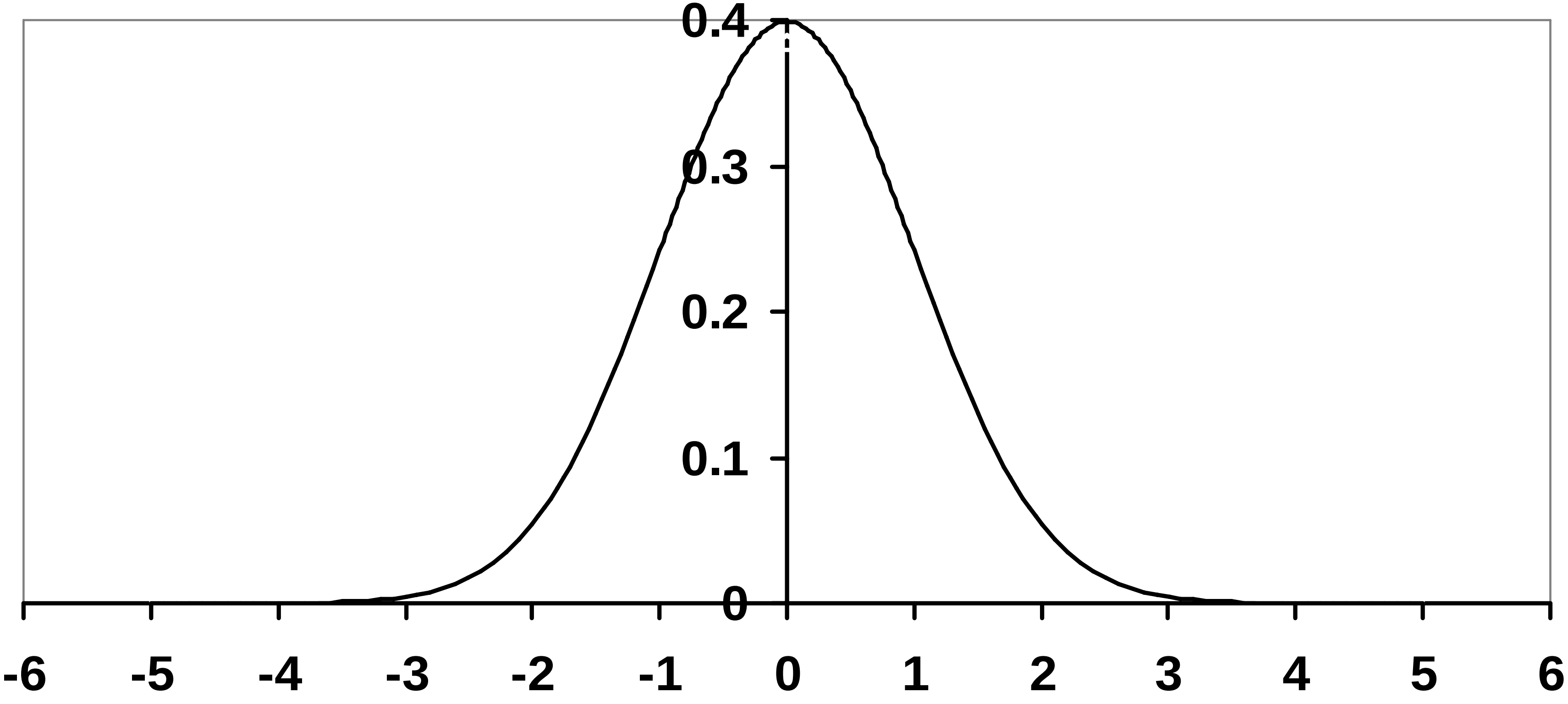
(and the percentage values appropriate to these thresholds are called the “significance levels”

Here is a graph of a normal distribution with zero mean and unit variance



95% of values lie within 0 ± 1.96

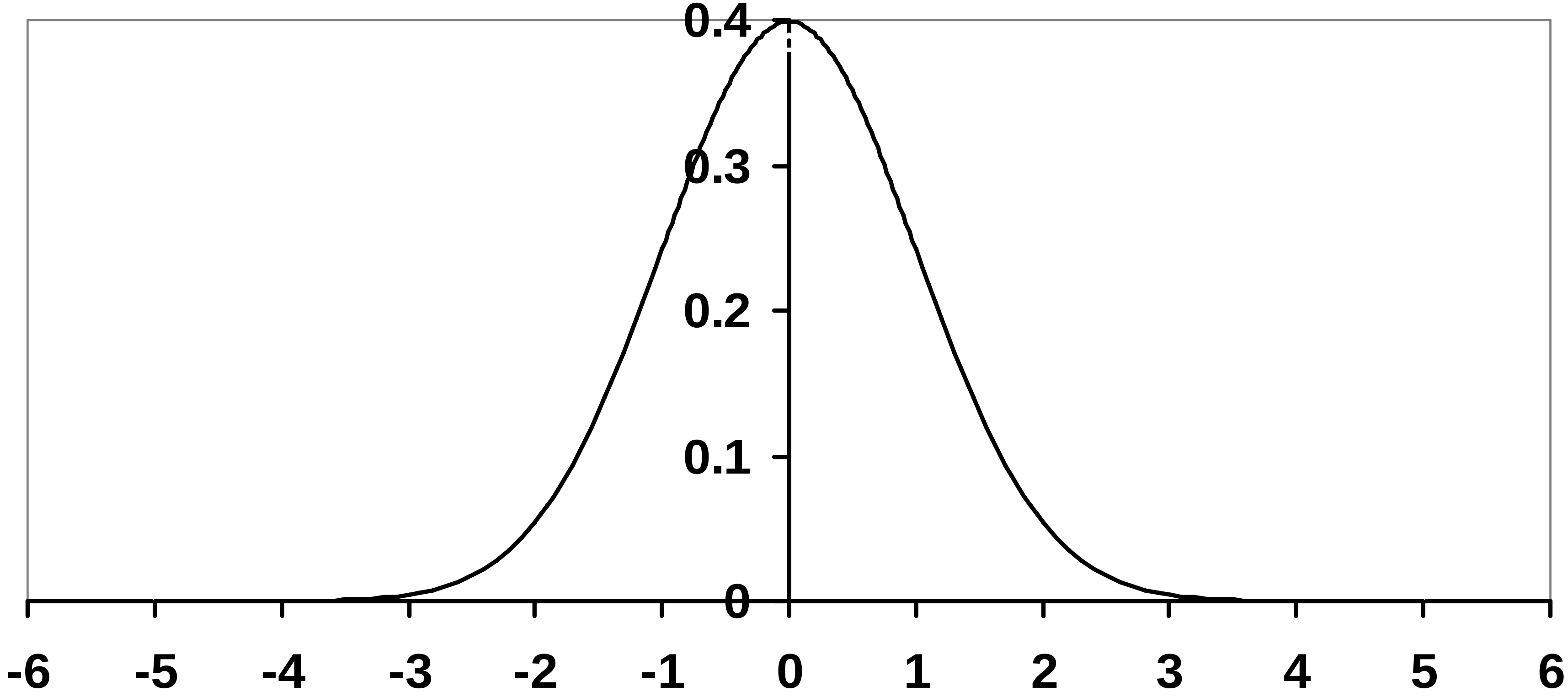
Here is a graph of a normal distribution with zero mean and unit variance



95% of values lie within 0 ± 1.96

so the critical value for the 95% significance level is 1.96

Here is 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 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: $H0: \beta_1 = \beta_1^0$

Testing A Hypothesis Relating To A Regression Coefficient

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

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

(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: H0: $\beta_1 = \beta_1^0$

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

Alternative hypothesis: H1: $\beta_1 \neq \beta_1^0$

Testing A Hypothesis Relating To A Regression Coefficient

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

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

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

Alternative hypothesis: H1: $\beta_1 \neq \beta_1^0$

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

Testing A Hypothesis Relating To A Regression Coefficient

Example

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

Testing A Hypothesis Relating To A Regression Coefficient

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Testing A Hypothesis Relating To A Regression Coefficient

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

Testing A Hypothesis Relating To A Regression Coefficient

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

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.

So given we now know $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \frac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

and also that 95% of all values of a variable that has mean 0 and variance 1 will lie within

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

and also that 95% of all values of a variable that has mean 0 and variance 1 will lie within

$$0 \pm 1.96$$

Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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)

Given $\hat{\beta}_1 \sim N(\beta_1, Var(\hat{\beta}_1))$ then $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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 <= z <= 1.96] = 0.95

Since $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

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

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

Since $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

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

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

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

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

or equivalently multiplying the terms in square brackets in by $s.d.(\hat{\beta}_1)$

Since $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

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

or equivalently multiplying the terms in square brackets in by $s.d.(\hat{\beta}_1)$

-1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1$

Since $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

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

or equivalently multiplying the terms in square brackets in by $s.d.(\hat{\beta}_1)$

-1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1$ and 1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1 <= 1.96 * s.d.(\hat{\beta}_1)$

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

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

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

or equivalently multiplying the terms in square brackets in by $s.d.(\hat{\beta}_1)$

-1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1$ and 1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1 <= 1.96 * s.d.(\hat{\beta}_1)$

and taking $\hat{\beta}_1$ to the other sides of the equality gives

Since $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

sub. this into Pr[-1.96 <= z <= 1.96] = 0.95

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

or equivalently multiplying the terms in square brackets in by $s.d.(\hat{\beta}_1)$

-1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1$ and 1.96* $s.d.(\hat{\beta}_1) <= \hat{\beta}_1 - \beta_1 <= 1.96 * s.d.(\hat{\beta}_1)$

and taking $\hat{\beta}_1$ to the other sides of the equality gives

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

is called the 95% *confidence interval*

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

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 β_1 will lie in this region

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

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 β_1 will lie in this region

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

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

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 β_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 β_1, only ever have an estimate, the standard error

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

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 β_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 β_1, only ever have an estimate, the standard error

ie don't have $s.e.(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{N * Var(X)}}$ since σ^2 (the true residual variance) unobserved

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

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 β_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 β_1, only ever have an estimate, the standard error

ie don't have $s.e.(\hat{\beta}_1) = \sqrt{\dfrac{\sigma^2}{N * Var(X)}}$ since σ^2 (the true residual variance) unobserved

but we now know that

$$s.e.(\hat{\beta}_1) = \sqrt{\frac{s^2}{N * Var(X)}}$$

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

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 β_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 β_1, only ever have an estimate, the standard error

ie don't have $s.d.(\hat{\beta}_1) = \sqrt{\dfrac{\sigma^2}{N * Var(X)}}$ since σ^2 (the true residual variance) unobserved

but we now know that

$$s.e.(\hat{\beta}_1) = \sqrt{\frac{s^2}{N * Var(X)}}$$ (replacing σ^2 with s^2)

so can substitute this into the equation above

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,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

However when we do this we no longer have a standard normal distribution,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

but instead the statistic

However when we do this we no longer have a standard normal distribution,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

but instead the statistic

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

However when we do this we no longer have a standard normal distribution,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

but instead the statistic

$$t = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1 \sqrt{N * Var(X)}}{s} \sim t(N-k)$$

However when we do this we no longer have a standard normal distribution,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

but instead the statistic

$$t = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)} \qquad = \frac{\hat{\beta}_1 - \beta_1 \sqrt{N * Var(X)}}{s} \sim t(N-k)$$

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

However when we do this we no longer have a standard normal distribution,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

but instead the statistic

$$t = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1 \sqrt{N * Var(X)}}{s} \sim t(N-k)$$

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

N = sample size

However when we do this we no longer have a standard normal distribution,

ie no longer $z = \dfrac{\hat{\beta}_1 - \beta_1}{s.d.(\hat{\beta}_1)} \sim N(0,1)$

but instead the statistic

$$t = \frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1 \sqrt{N * Var(X)}}{s} \sim t(N-k)$$

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)

Now this t distribution is symmetrical about its mean but has its own set of critical values at given significance levels

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 ***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

So $t = \dfrac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$

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 a hypothesized value and still have a t distribution

So $t = \dfrac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$ and also $t = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ $\sim t_{(N-k)}$

Nevertheless the general rules established above apply

In particular the t distribution will be centred on zero

Nevertheless the general rules established above apply

In particular the t distribution will be centred on zero

Why?

Nevertheless the general rules established above apply

In particular the t distribution will be centred on zero

Why?

if the hypothesised value β_1^0 and estimated value $\hat{\beta}_1$ are the same then

Nevertheless the general rules established above apply

In particular the t distribution will be centred on zero

Why?

if the hypothesised value β_1^0 and estimated value $\hat{\beta}_1$ are the same then

$$\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)} = 0$$

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

So the idea then is to ask whether the estimated t value

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

So the idea then is to ask whether the estimated t value

$$\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$$

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

So the idea then is to ask whether the estimated t value

$$\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$$

is close enough to zero

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

So the idea then is to ask whether the estimated t value

$$\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$$

is close enough to zero

allowing for sampling variation

Now the reality of sampling means that would not expect an estimate to be exactly equal to the hypothesized value

(randomness means the stochastic world is not a deterministic world)

But…

If the hypothesis is true we would expect the estimated value to be close enough to the hypothesized value *allowing for the inevitable variability thrown up by the fact that we are using a sample*

So the idea then is to ask whether the estimated t value

$$\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$$

is close enough to zero

allowing for sampling variation

to be consistent with the null hypothesis

Close enough in this context means that the estimated t value will lie in the acceptance region

Close enough in this context means that the estimated t value will lie in the acceptance region

- whose critical thresholds depend on

Close enough in this context means that the estimated t value will lie in the acceptance region

- whose critical thresholds depend on

the significance level (α)

Close enough in this context means that the estimated t value will lie in the acceptance region

- whose critical thresholds depend on

the significance level (α)
the degrees of freedom (N-k)

Close enough in this context means that the estimated t value will lie in the acceptance region

- whose critical thresholds depend on

the significance level (α)
the degrees of freedom (N-k)

Since the t distribution is symmetric (about zero) can just look at the absolute value

Close enough in this context means that the estimated t value will lie in the acceptance region

- whose critical thresholds depend on

the significance level (α)
the degrees of freedom (N-k)

Since the t distribution is symmetric (about zero) can just look at the absolute value

In general if the **absolute** value of $\hat{t}$ > critical value, this means the estimated coefficient lies so far away from the hypothesised value as to be inconsistent with the null hypothesis even allowing for sampling variation

So if

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| > t^{\alpha}_{N-k}$$

reject the null at the α% significance level.

So if

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| > t^{\alpha}_{N-k}$$

reject the null at the α% significance level.

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| \leq t^{\alpha}_{N-k}$$

accept the null

So if

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^{\,0}}{s.e.(\hat{\beta}_1)}\right| > t_{N-k}^{\alpha}$$

reject the null at the α% significance level.

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^{\,0}}{s.e.(\hat{\beta}_1)}\right| \le t_{N-k}^{\alpha}$$

accept the null

The most common levels of significance of a test are
$\alpha = 0.05$ (5% significance)
$\alpha = 0.01$ (1% significance)

So if

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^{\,0}}{s.e.(\hat{\beta}_1)}\right| > t_{N-k}^{\alpha}$$

reject the null at the α% significance level.

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^{\,0}}{s.e.(\hat{\beta}_1)}\right| \leq t_{N-k}^{\alpha}$$

accept the null

The most common levels of significance of a test are
α =0.05 (5% significance)
α =0.01 (1% significance)

Usual convention is to base a t-test using the critical values at the 5% level of significance (though this may vary a little depending on the size of the sample).

So if

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| > t^{\alpha}_{N-k}$$

reject the null at the α% significance level.

$$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| \le t^{\alpha}_{N-k}$$

accept the null

The most common levels of significance of a test are
α =0.05 (5% significance)
α =0.01 (1% significance)

Usual convention is to base a t-test using the critical values at the 5% level of significance (though this may vary a little depending on the size of the sample).

Most econometric computer packages routinely report the t value for a null hypothesis that that particular coefficient is zero (the variable has no effect)

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$

Example

Cons = β_0 + β_1*Income* + *u*

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$ $t = 0.917 - 0$

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$ $t = 0.917 - 0$ $t = \dfrac{0.917 - 0}{0.012}$

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$ $t = 0.917 - 0$ $t = \frac{0.917 - 0}{0.012}$

$t = 77.26$

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$ $t = 0.917 - 0$ $t = \dfrac{0.917 - 0}{0.012}$

$t = 77.26$

Does this t value lie in acceptance region of the t distribution?

Example

Cons = β_0 + β_1*Income* + *u*

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$ $t = 0.917 - 0$ $t = \dfrac{0.917 - 0}{0.012}$

$t = 77.26$

Does this t value lie in acceptance region of the t distribution?
If so estimate of β_1 is consistent with the null hypothesis

Example

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

Null hypothesis: H0: $\beta_1 = 0$

Alternative hypothesis: H1: $\beta_1 \neq 0$

In this case N=45 k = 2 (constant income)

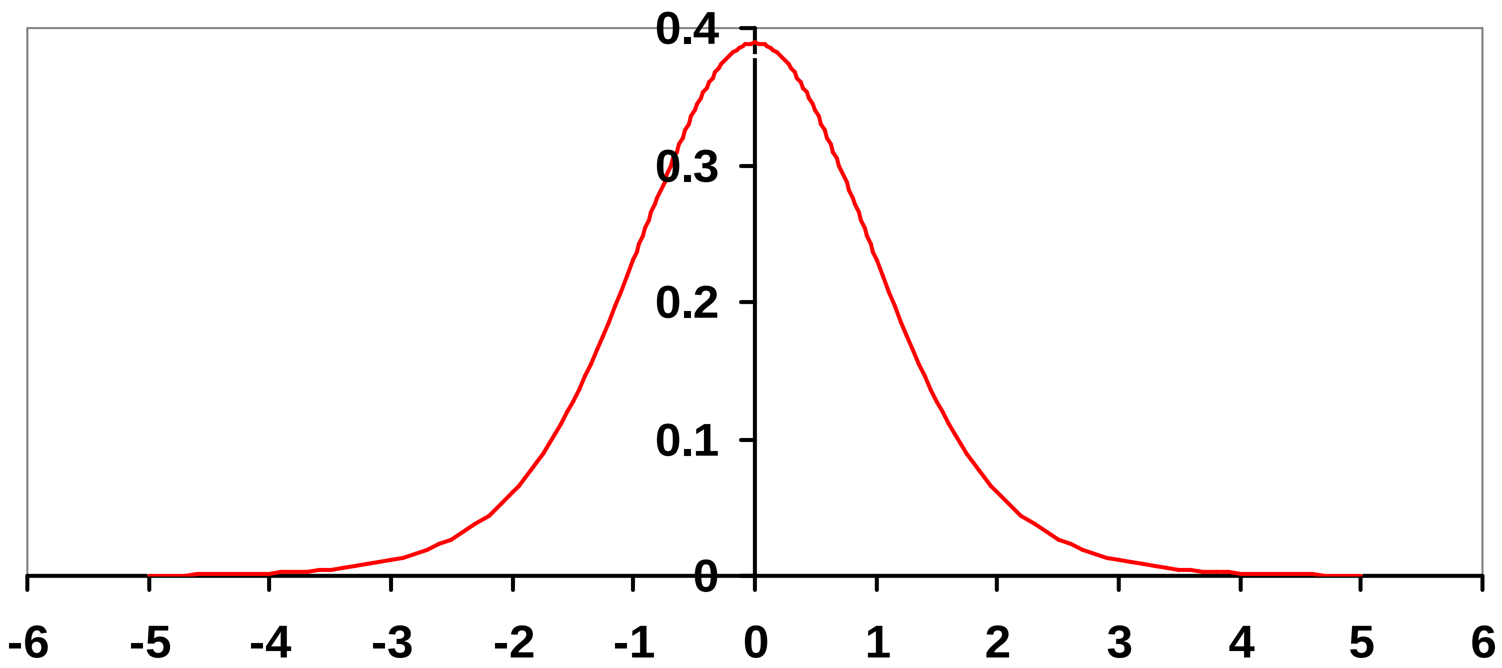
So given $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then here $t = 0.917$ $t = 0.917 - 0$ $t = \dfrac{0.917 - 0}{0.012}$

$t = 77.26$

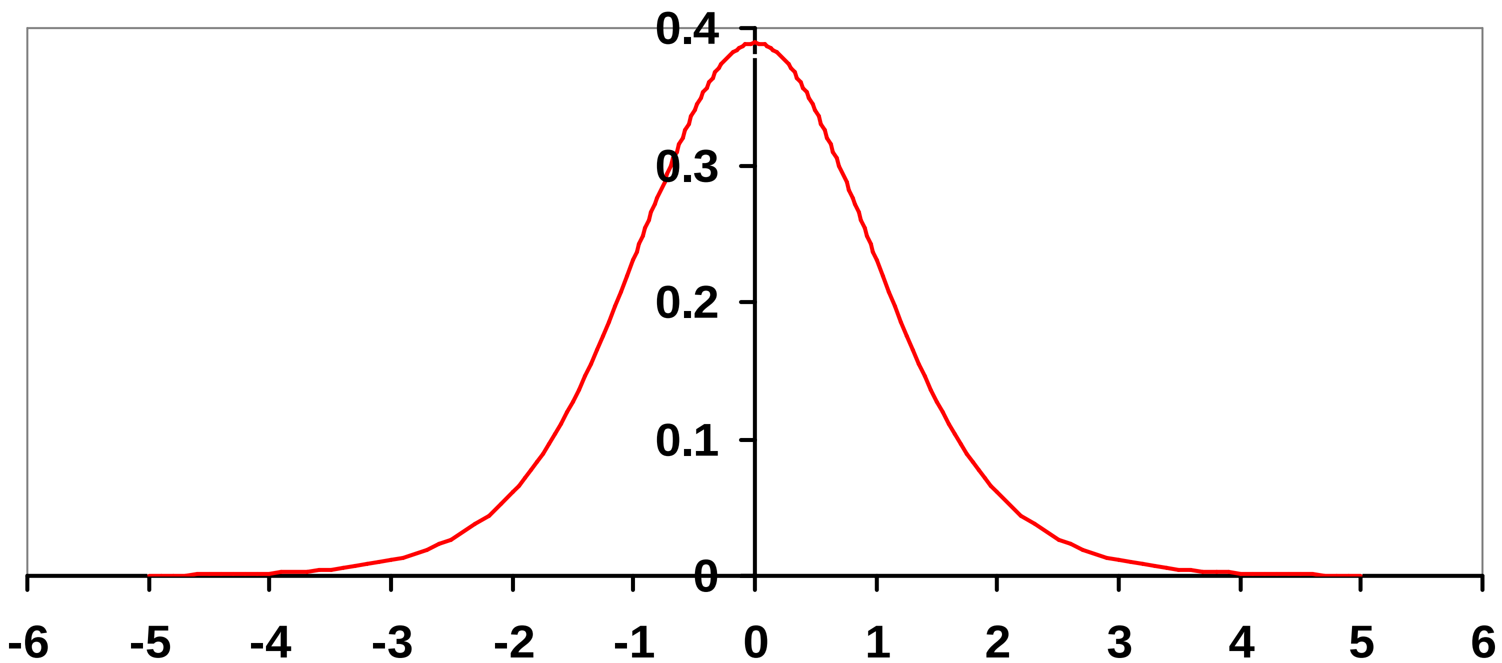
Does this t value lie in acceptance region of the t distribution?
If so estimate of β_1 is consistent with the null hypothesis

If not it is inconsistent with the null (estimate too far away from hypothesised value)

Given a t distribution need to find critical values for acceptance region – and these depend on sample size N and number of coefficients in model k



Given a t distribution need to find critical values for acceptance region – and these depend on sample size N and number of coefficients in model k



There are tables that will tell us the critical values

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
11.	1.363	1.796	2.201	2.718	3.106	4.024
12.	1.356	1.782	2.179	2.681	3.055	3.929
13.	1.350	1.771	2.160	2.650	3.012	3.852
14.	1.345	1.761	2.145	2.624	2.977	3.787
15.	1.341	1.753	2.131	2.602	2.947	3.733
16.	1.337	1.746	2.120	2.583	2.921	3.686
17.	1.333	1.740	2.110	2.567	2.898	3.646
18.	1.330	1.734	2.101	2.552	2.878	3.610
19.	1.328	1.729	2.093	2.539	2.861	3.579
20.	1.325	1.725	2.086	2.528	2.845	3.552
21.	1.323	1.721	2.080	2.518	2.831	3.527
22.	1.321	1.717	2.074	2.508	2.819	3.505
23.	1.319	1.714	2.069	2.500	2.807	3.485
24.	1.318	1.711	2.064	2.492	2.797	3.467
25.	1.316	1.708	2.060	2.485	2.787	3.450

26.	1.315	1.706	2.056	2.479	2.779	3.435
27.	1.314	1.703	2.052	2.473	2.771	3.421
28.	1.313	1.701	2.048	2.467	2.763	3.408
29.	1.311	1.699	2.045	2.462	2.756	3.396
30.	1.310	1.697	2.042	2.457	2.750	3.385
31.	1.309	1.696	2.040	2.453	2.744	3.375
32.	1.309	1.694	2.037	2.449	2.738	3.365
33.	1.308	1.692	2.035	2.445	2.733	3.356
34.	1.307	1.691	2.032	2.441	2.728	3.348
35.	1.306	1.690	2.030	2.438	2.724	3.340
36.	1.306	1.688	2.028	2.434	2.719	3.333
37.	1.305	1.687	2.026	2.431	2.715	3.326
38.	1.304	1.686	2.024	2.429	2.712	3.319
39.	1.304	1.685	2.023	2.426	2.708	3.313
40.	1.303	1.684	2.021	2.423	2.704	3.307
41.	1.303	1.683	2.020	2.421	2.701	3.301
42.	1.302	1.682	2.018	2.418	2.698	3.296
43.	1.302	1.681	2.017	2.416	2.695	3.291
44.	1.301	1.680	2.015	2.414	2.692	3.286
45.	1.301	1.679	2.014	2.412	2.690	3.281
46.	1.300	1.679	2.013	2.410	2.687	3.277
47.	1.300	1.678	2.012	2.408	2.685	3.273
48.	1.299	1.677	2.011	2.407	2.682	3.269
49.	1.299	1.677	2.010	2.405	2.680	3.265
50.	1.299	1.676	2.009	2.403	2.678	3.261
51.	1.298	1.675	2.008	2.402	2.676	3.258
52.	1.298	1.675	2.007	2.400	2.674	3.255
53.	1.298	1.674	2.006	2.399	2.672	3.251
54.	1.297	1.674	2.005	2.397	2.670	3.248
55.	1.297	1.673	2.004	2.396	2.668	3.245
56.	1.297	1.673	2.003	2.395	2.667	3.242

88.	1.291	1.662	1.987	2.369	2.633	3.185
89.	1.291	1.662	1.987	2.369	2.632	3.184
90.	1.291	1.662	1.987	2.368	2.632	3.183
91.	1.291	1.662	1.986	2.368	2.631	3.182
92.	1.291	1.662	1.986	2.368	2.630	3.181
93.	1.291	1.661	1.986	2.367	2.630	3.180
94.	1.291	1.661	1.986	2.367	2.629	3.179
95.	1.291	1.661	1.985	2.366	2.629	3.178
96.	1.290	1.661	1.985	2.366	2.628	3.177
97.	1.290	1.661	1.985	2.365	2.627	3.176
98.	1.290	1.661	1.984	2.365	2.627	3.175
99.	1.290	1.660	1.984	2.365	2.626	3.175
100.	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090

In the above example the degrees of freedom = 45

In the above example the degrees of freedom = 45 (N)

In the above example the degrees of freedom = 45 (N) – 2

In the above example the degrees of freedom = 45 (N) – 2 (K)

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail (and 2.5% in the bottom since the distribution is symmetric)

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail (and 2.5% in the bottom since the distribution is symmetric) is 2.017

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

	Probability of exceeding the critical value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail
(and 2.5% in the bottom since the distribution is symmetric)
is 2.017

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

	Probability of exceeding the critical value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail
(and 2.5% in the bottom since the distribution is symmetric)
is 2.017

then ask whether the estimated t value lies above or below this value (in absolute terms)

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail
(and 2.5% in the bottom since the distribution is symmetric)
is 2.017

then ask whether the estimated t value lies above or below this value (in absolute terms)

Since estimate t=77.26

Lecture 6

More about t values

Confidence intervals

Type I and Type II error & P values

Some different hypothesis tests

In the above example the degrees of freedom = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail
(and 2.5% in the bottom since the distribution is symmetric)
is 2.017

then ask whether the estimated t value lies above or below this value (in absolute terms)
Since estimate t=77.26

Then $$\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| > t^{\alpha}_{N-k}$$ reject the null – that the variable has no effect - at the α% significance level.

77.26 > 2.017

While most econometric software routinely reports the t value based on a null hypothesis that a single coefficient is zero

(and remember a null of zero effectively implies that the effect of that variable is zero since

While most econometric software routinely reports the t value based on a null hypothesis that a single coefficient is zero

(and remember a null of zero effectively implies that the effect of that variable is zero since

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

Implies that the slope estimate $\dfrac{\partial Cons}{\partial Income} = \beta_1 = 0$

While most econometric software routinely reports the t value based on a null hypothesis that a single coefficient is zero

(and remember a null of zero effectively implies that the effect of that variable is zero since

Cons = β_0 + β_1*Income* + *u*

that the slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0$

(a change in income has NO effect on consumption)

While most econometric software routinely reports the t value based on a null hypothesis that a single coefficient is zero

(and remember a null of zero effectively implies that the effect of that variable is zero since

Cons = β_0 + β_1*Income* + *u*

that the slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0$

(a change in income has NO effect on consumption)

However the principle of using a t value to compare the difference between the estimate and the hypothesized value applies to ANY hypothesized value

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

(a unit change in income leads to a 0.9unit change in consumption)

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

(a unit change in income leads to a 0.9unit change in consumption)

Alternative hypothesis: H1: $\beta_1 \neq 0.9$

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

(a unit change in income leads to a 0.9unit change in consumption)

Alternative hypothesis: H1: $\beta_1 \neq 0.9$

Still the case that N=45 k = 2 (constant income)

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

(a unit change in income leads to a unit change in consumption)

Alternative hypothesis: H1: $\beta_1 \neq 0.9$

Still the case that N=45 k = 2 (constant income)

So given $\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

(a unit change in income leads to a unit change in consumption)

Alternative hypothesis: H1: $\beta_1 \neq 0.9$

Still the case that N=45 k = 2 (constant income)

So given $\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then now $t = \frac{0.917 - 0.9}{0.012}$

So suppose we have a new null

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

Null hypothesis: H0: $\beta_1 = 0.9$

slope estimate $\frac{\partial Cons}{\partial Income} = \beta_1 = 0.9$

(a unit change in income leads to a unit change in consumption)

Alternative hypothesis: H1: $\beta_1 \neq 0.9$

Still the case that N=45 k = 2 (constant income)

So given $\hat{t} = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ then now $t = \frac{0.917 - 0.9}{0.012}$

$t = 1.43$

Does this new t value lie in acceptance region of the t distribution? If so estimate of β_1 is consistent with the null hypothesis

Does this new t value lie in acceptance region of the t distribution? If so estimate of β_1 is consistent with the null hypothesis

In the above example the degrees of freedom again = 45 (N) – 2 (K) = 43

Does this new t value lie in acceptance region of the t distribution? If so estimate of β_1 is consistent with the null hypothesis

In the above example the degrees of freedom again = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail (and 2.5% in the bottom since the distribution is symmetric) is again 2.017

Does this new t value lie in acceptance region of the t distribution? If so estimate of β_1 is consistent with the null hypothesis

In the above example the degrees of freedom again = 45 (N) – 2 (K) = 43

So look across the row with 43 degrees of freedom

Upper critical values of Student's t distribution with ν degrees of freedom

Probability of exceeding the critical value

ν	0.10	0.05	0.025	0.01	0.005	0.001
43.	1.302	1.681	2.017	2.416	2.695	3.291

to find that the critical value that puts 2.5% of the distribution in the top tail (and 2.5% in the bottom since the distribution is symmetric) is again 2.017

This time $\left|\hat{t}\right| = \left|\frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}\right| > t^{\alpha}_{N-k}$

accept the null – that the variables true effect is 0.9 - at the 5% significance level.

$1.43 < 2.017$

Confidence Intervals Revisited

Confidence Intervals Revisited

Already seen for the standard normal distribution that the 95% confidence interval (the interval in which 95% of observations lie) can be written as

Confidence Intervals Revisited

Already seen for the standard normal distribution that the 95% confidence interval (the interval in which 95% of observations lie) can be written as

$$Pr[-1.96 \le z \le 1.96] = 0.95$$

Confidence Intervals Revisited

Already seen for the standard normal distribution that the 95% confidence interval (the interval in which 95% of observations lie) can be written as

Pr[-1.96 <= z <= 1.96] = 0.95

or

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

Confidence Intervals Revisited

Already seen for the standard normal distribution that the 95% confidence interval (the interval in which 95% of observations lie) can be written as

Pr[-1.96 <= z <= 1.96] = 0.95

or

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

or

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

Confidence Intervals Revisited

Already seen for the standard normal distribution that the 95% confidence interval (the interval in which 95% of observations lie) can be written as

Pr[-1.96 <= z <= 1.96] = 0.95

or

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

or

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

when working with the t rather than the standard normal distribution have to modify this a little

Given

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

Given

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

- 1. Replace the standard deviation of the standard normal distribution with the standard error of the t distribution

Given

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

1. Replace the standard deviation of the standard normal distribution with the standard error of the t distribution $s.e.(\hat{\beta}_1) = \sqrt{\dfrac{s^2}{N * Var(X)}}$

Given

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

1. Replace the standard deviation of the standard normal distribution with the standard error of the t distribution $s.e.(\hat{\beta}_1) = \sqrt{\frac{s^2}{N * Var(X)}}$

2. – 1.96 & +1.96 are the critical values for the 95% acceptance region of the standard normal so have to replace those with the critical values for the 95% acceptance region of the t distribution

Given

$$\Pr\left[\hat{\beta}_1 - 1.96 * s.d.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 1.96 * s.d.(\hat{\beta}_1)\right] = 0.95$$

1. Replace the standard deviation of the standard normal distribution with the standard error of the t distribution $s.e.(\hat{\beta}_1) = \sqrt{\dfrac{s^2}{N * Var(X)}}$

2. – 1.96 & +1.96 are the critical values for the 95% acceptance region of the standard normal so have to replace those with the critical values for the 95% acceptance region of the t distribution

and have already seen that this depends on the sample size N and the number of right hand side parameters k

So

$$\Pr\left[\hat{\beta}_1 - t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

So

$$\Pr\left[\hat{\beta}_1 - t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

or equivalently

$$\Pr\left[\hat{\beta}_1 \pm t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

is the 95% confidence interval for in which the true value of the coefficient will lie for the t distribution

So

$$\Pr\left[\hat{\beta}_1 - t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

or equivalently

$$\Pr\left[\hat{\beta}_1 \pm t_{N-k}^{.05/2} * s.e.(\hat{\beta}_1)\right] = 0.95$$

is the 95% confidence interval for in which the true value of the coefficient will lie for the t distribution

(remember it is a 2-tailed test so the significance level α =0.05 is divided by 2 to allow 5% of the distribution to appear outside the acceptance region with 2.5% in either tail)

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region - so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region - so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

By **increasing** the size of the test (α goes from 1% to 5% for example) we **reduce** the acceptance region for a given t estimate (and increase the range of estimate that fall in the rejection region)

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

By **increasing** the size of the test (α goes from 1% to 5% for example) we **reduce** the acceptance region for a given t estimate (and increase the range of estimate that fall in the rejection region)

The danger of this is that increase the chance that reject the null hypothesis even though it is true.

Whether we accept or reject a hypothesis is partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

By **increasing** the size of the test (α goes from 1% to 5% for example) we **reduce** the acceptance region for a given t estimate (and increase the range of estimate that fall in the rejection region)

The danger of this is that increase the chance that reject the null hypothesis even though it is true.

So whether we accept or reject a hypothesis is therefore partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

By **increasing** the size of the test (α goes from 1% to 5% for example) we **reduce** the acceptance region for a given t estimate (and increase the range of estimate that fall in the rejection region)

The danger of this is that increase the chance that reject the null hypothesis even though it is true.

This called Type I error

So whether we accept or reject a hypothesis is therefore partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

By **increasing** the size of the test (α goes from 1% to 5% for example) we **reduce** the acceptance region for a given t estimate (and increase the range of estimate that fall in the rejection region)

The danger of this is that increase the chance that reject the null hypothesis even though it is true.

This called Type I error

Clearly to reduce Type I error we increase the acceptance region by choosing a smaller size of the test, but the cost of this is that it increase the risk of Type II error

So whether we accept or reject a hypothesis is therefore partly dependent on the choice of significance levels of a test

By **reducing** the size (significance level) of the test (α goes from 5% to 1% for example)) we **increase** the acceptance region for a given t estimate (and reduce the range of estimates that fall in the rejection region- so making it harder to reject a null)

The danger of this is that increase the chance that the null hypothesis could be false, but that we accept it.

This called Type II error

By **increasing** the size of the test (α goes from 1% to 5% for example) we **reduce** the acceptance region for a given t estimate (and increase the range of estimate that fall in the rejection region)

The danger of this is that increase the chance that reject the null hypothesis even though it is true.

This called Type I error

Clearly to reduce Type I error we increase the acceptance region by choosing a smaller size of the test, but the cost of this is that it increase the risk of Type II error

So what to do?

So what to do?

- depends on which is the most serious

So what to do?

- depends on which is the most serious

- the more you believe that avoiding Type 1 error (rejecting a true null) is more important than avoiding Type II error (accepting a false null), then would want a larger acceptance region (confidence interval) and a smaller the size of the test (α= 1% rather than 5% for example).

So what to do?

- depends on which is the most serious

- the more you believe that avoiding Type 1 error (rejecting a true null) is more important than avoiding Type II error (accepting a false null), then would want a larger acceptance region (confidence interval) and a smaller the size of the test (α= 1% rather than 5% for example).

Since rarely know the costs of these errors, have to settle on critical values which balance the two errors

So what to do?

- depends on which is the most serious

- the more you believe that avoiding Type 1 error (rejecting a true null) is more important than avoiding Type II error (accepting a false null), then would want a larger acceptance region (confidence interval) and a smaller the size of the test (α= 1% rather than 5% for example).

Since rarely know the costs of these errors, have to settle on critical values which balance the two errors

- which is why the 5% level of significance is the one usually used.

So what to do?

- depends on which is the most serious

- the more you believe that avoiding Type 1 error (rejecting a true null) is more important than avoiding Type II error (accepting a false null), then would want a larger acceptance region (confidence interval) and a smaller the size of the test (α= 1% rather than 5% for example).

Since rarely know the costs of these errors, have to settle on critical values which balance the two errors

- which is why the 5% level of significance is the one usually used.

It is true however that there is some **trade-off** between the sample size and the significance level of the test. Might want to use the 1% significance level in large samples (a tougher test, since we know large N will produce small standard errors and hence larger t values other things equal)

So what to do?

- depends on which is the most serious

- the more you believe that avoiding Type 1 error (rejecting a true null) is more important than avoiding Type II error (accepting a false null), then would want a larger acceptance region (confidence interval) and a smaller the size of the test (α= 1% rather than 5% for example).

Since rarely know the costs of these errors, have to settle on critical values which balance the two errors

- which is why the 5% level of significance is the one usually used.

It is true however that there is some **trade-off** between the sample size and the significance level of the test. Might want to use the 1% significance level in large samples (a tougher test, since we know large N will produce small standard errors and hence larger t values other things equal)

Since we know $s.e.(\hat{\beta}_1) = \sqrt{\dfrac{s^2}{N * Var(X)}}$

Then ↑N → ↓ $s.e.(\hat{\beta}_1)$

and hence because $\hat{t} = \dfrac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)}$ this will ↑ estimated t value

There are no rules on this but might want to think about the following guidelines:

N in the 10's	use α =10%
N in the 100's	use α =5%
N in the 1000's	use α =1%

It is more common to use the term “significance level” of the test than “size”. Unfortunately it is sometimes confusing, since to increase the significance level of the test usually means reduce the size of the test ie go from 5% to 1% level.

Also some texts refer to 95% levels which is the same as a 5% level (95% acceptance probability, 5% rejection probability)

A useful corollary is to be aware of the **p value** of a test

A useful corollary is to be aware of the **p value** of a test

The p value is the lowest significance level at which the null hypothesis can be rejected (the exact probability of committing Type I error)

A useful corollary is to be aware of the **p value** of a test

The p value is the lowest significance level at which the null hypothesis can be rejected (the exact probability of committing Type I error)

In practice this amounts to finding the significance level α which, given sample size N and no. right hand side coefficients k, equates the critical and estimated t values

$$\hat{t} = t_{N-k}^{\alpha/2}$$

A useful corollary is to be aware of the **p value** of a test

The p value is the lowest significance level at which the null hypothesis can be rejected (the exact probability of committing Type I error)

In practice this amounts to finding the significance level α which, given sample size N and no. right hand side coefficients k, equates the critical and estimated t values

$$\hat{t} = t_{N-k}^{\alpha/2}$$

(ie most estimated t values will not be exactly equal to the 5% or 1% critical values by they will be equal to the p% critical value)

A useful corollary is to be aware of the **p value** of a test

The p value is the lowest significance level at which the null hypothesis can be rejected (the exact probability of committing Type I error)

In practice this amounts to finding the significance level α which, given sample size N and no. right hand side coefficients k, equates the critical and estimated t values

$$\hat{t} = t_{N-k}^{\alpha/2}$$

(ie most estimated t values will not be exactly equal to the 5% or 1% critical values by they will be equal to the p% critical value)

Intuitively if can ↑α a lot (and in so doing reduce the acceptance region) and still accept the null hypothesis, this suggests the null hypothesis is likely to be true.

A useful corollary is to be aware of the **p value** of a test

The p value is the lowest significance level at which the null hypothesis can be rejected (the exact probability of committing Type I error)

In practice this amounts to finding the significance level α which, given sample size N and no. right hand side coefficients k, equates the critical and estimated t values

$$\hat{t} = t_{N-k}^{\alpha/2}$$

(ie most estimated t values will not be exactly equal to the 5% or 1% critical values by they will be equal to the p% critical value)

Intuitively if can ↑α a lot (and in so doing reduce the acceptance region) and still accept the null hypothesis, this suggests the null hypothesis is likely to be true.

So a **high** p value (high α) is evidence **in favour** of the null and a **low p** value is evidence in **against** the null

A useful corollary is to be aware of the **p value** of a test

The p value is the lowest significance level at which the null hypothesis can be rejected (the exact probability of committing Type I error)

In practice this amounts to finding the significance level α which, given sample size N and no. right hand side coefficients k, equates the critical and estimated t values

$$\hat{t} = t_{N-k}^{\alpha/2}$$

(ie most estimated t values will not be exactly equal to the 5% or 1% critical values by they will be equal to the p% critical value)

Intuitively if can ↑α a lot (and in so doing reduce the acceptance region) and still accept the null hypothesis, this suggests the null hypothesis is likely to be true.

So a **high** p value (high α) is evidence **in favour** of the null and a **low p** value is evidence in **against** the null

More formally if in the regression output

p < chosen level of α (say 0.05 = 5%)

reject the null hypothesis

Statistical v. Practical Significance

Just because a variable is statistically significant in a regression does not mean that it has a large economic impact on the dependent variable

May have a large t values (and ∴ be statistically significant from zero) but if the estimated coefficient

If $\hat{\beta} = \frac{\partial y}{\partial X}$ is small then

the impact of a change in X on y $\hat{\beta} \partial X = \partial y$
would also be very small

Moral: Significance and size of effect are both important. When reporting the effect of coefficients

1. Check variable is statistically significant from zero
2. If it is then (and only then) discuss the size of the effect as implied by the regression coefficient

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

So how do we know that the value for the R^2 reflects a true relationship rather than the result of chance?

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

So how do we know that the value for the R^2 reflects a true relationship rather than the result of chance?

We know that

$$\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(\hat{u})$$

F test of Goodness of Fit of the Model

We know that the R^2 gives an idea of the strength of association between dependent and explanatory variable

Because of statistical variation unlikely that R^2 would ever fall to zero in absence of any (true) correlation between y and X variables

So how do we know that the value for the R^2 reflects a true relationship rather than the result of chance?

We know that

$$\mathrm{Var}(Y) = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(\hat{u})$$

and that this can be also written as

$$\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum \hat{u}^2$$

$$TSS = ESS + RSS$$

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \qquad \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \sim F[k\text{-}1, N\text{-}k]$$

to test whether the R^2 is statistically significantly different from zero

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Just how high depends (like t values) on the sample size N and the number of right hand side coefficients k (hence the corrections above)

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \qquad \sim F[k\text{-}1, N\text{-}k]$$

to test whether the R^2 is statistically significantly different from zero

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Just how high depends (like t values) on the sample size N and the number of right hand side coefficients k (hence the corrections above)

The F statistic is said to follow an F distribution
(it is the ratio of 2 chi-squared distributions divided by their respective degrees of freedom)

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k - 1}{RSS / N - k} \qquad \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

Just how high depends (like t values) on the sample size N and the number of right hand side coefficients k (hence the corrections above)

The F statistic is said to follow an F distribution
(it is the ratio of 2 chi-squared distributions divided by their respective degrees of freedom)

which has its own set of degrees of freedom

F[k-1, N-k]

Can show (don't need to learn proof) that can use the value

$$F = \frac{ESS / k-1}{RSS / N-k} \quad \sim F[k-1, N-k]$$

to test whether the R^2 is statistically significantly different from zero

Intuitively if the model is a good fit then the ESS will be high relative to the RSS.

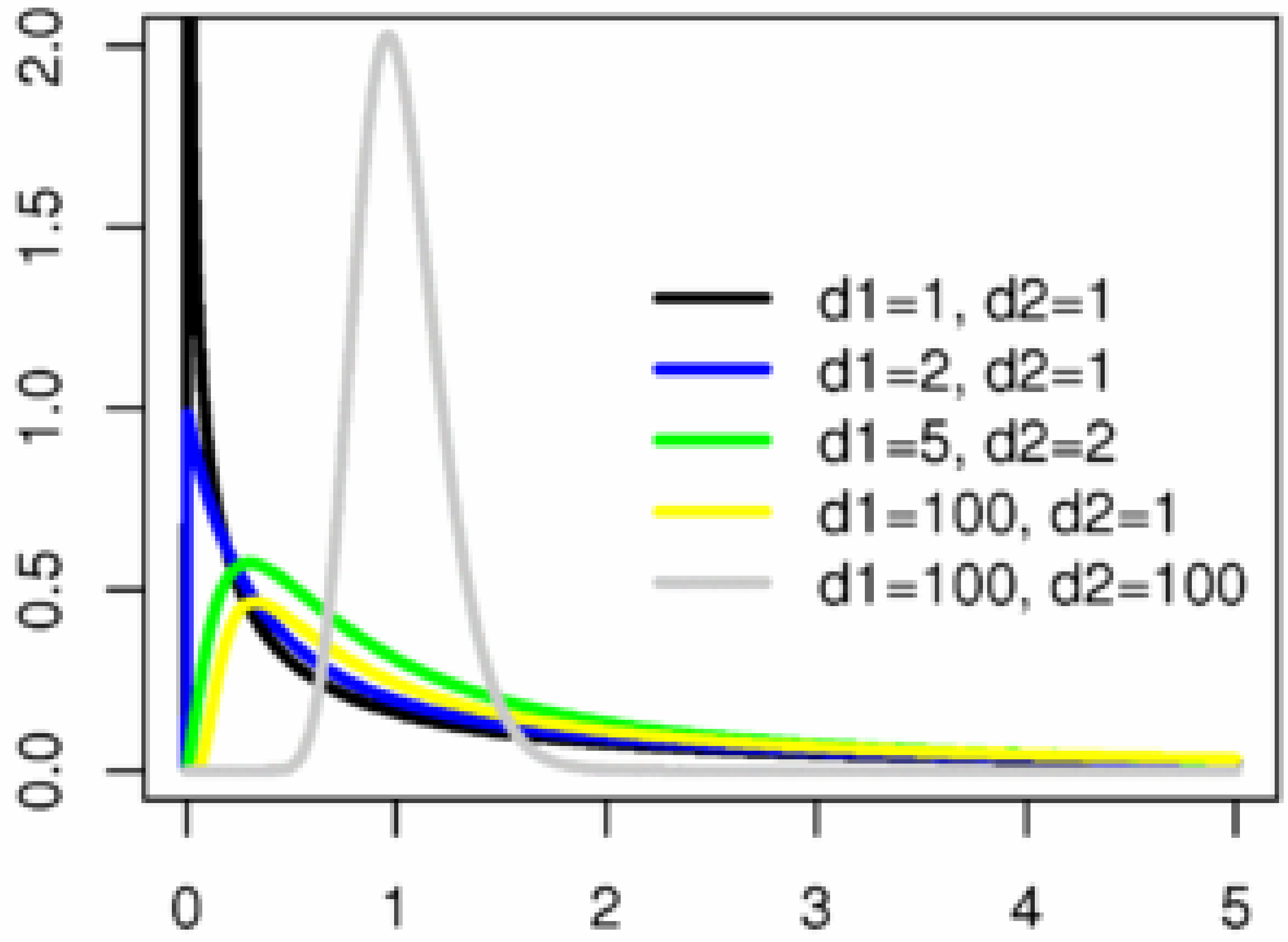
Just how high depends (like t values) on the sample size N and the number of right hand side coefficients k (hence the corrections above)

The F statistic is said to follow an F distribution
(it is the ratio of 2 chi-squared distributions divided by their respective degrees of freedom)

which has its own set of degrees of freedom

F[k-1, N-k]

k-1 said to be the numerator
N-k said to be the denominator (for obvious reasons)



2.0
1.5
1.0
0.5
0.0
d1=1, d2=1
d1=2, d2=1
d1=5, d2=2
d1=100, d2=1
d1=100, d2=100
0
1
2
3
4
5

Now, unlike the t distribution, the F distribution (because it is “squared”) is bounded from below at zero.

Now, unlike the t distribution, the F distribution (because it is “squared”) is bounded from below at zero.

This means that any hypothesis testing is based on a “1-tailed test” since the rejection region now only lies at the far right-end tail of the distribution

Now, unlike the t distribution, the F distribution (because it is “squared”) is bounded from below at zero.

This means that any hypothesis testing is based on a “1-tailed test” since the rejection region now only lies at the far right-end tail of the distribution

The basic rule is that

$$\text{if } \hat{F} > F^{(k-1,N-k)}_{\alpha\_critical}$$

Now, unlike the t distribution, the F distribution (because it is "squared") is bounded from below at zero.

This means that any hypothesis testing is based on a "1-tailed test" since the rejection region now only lies at the far right-end tail of the distribution

The basic rule is that

$$\text{if } \hat{F} > F_{\alpha\_critical}^{(k-1, N-k)}$$

then can be 1-α% confident that result is unlikely to have arisen by chance (the ESS is high wrt the RSS)

Since the F-statistic can also be manipulated to give

$$F = \frac{ESS/k-1}{RSS/N-k} \qquad = \frac{(ESS/TSS)/k-1}{(RSS/TSS)/N-k}$$

Since the F-statistic can also be manipulated to give

$$F = \frac{ESS/k-1}{RSS/N-k} \qquad = \frac{(ESS/TSS)/k-1}{(RSS/TSS)/N-k}$$

then $F = \dfrac{R^2/k-1}{(1-R^2)/N-k}$

which is an equivalent way of doing the test

Since the F-statistic can also be manipulated to give

$$F = \frac{ESS / k-1}{RSS / N-k} \qquad = \frac{(ESS / TSS) / k-1}{(RSS / TSS) / N-k}$$

then $F = \dfrac{R^2 / k-1}{(1-R^2) / N-k}$

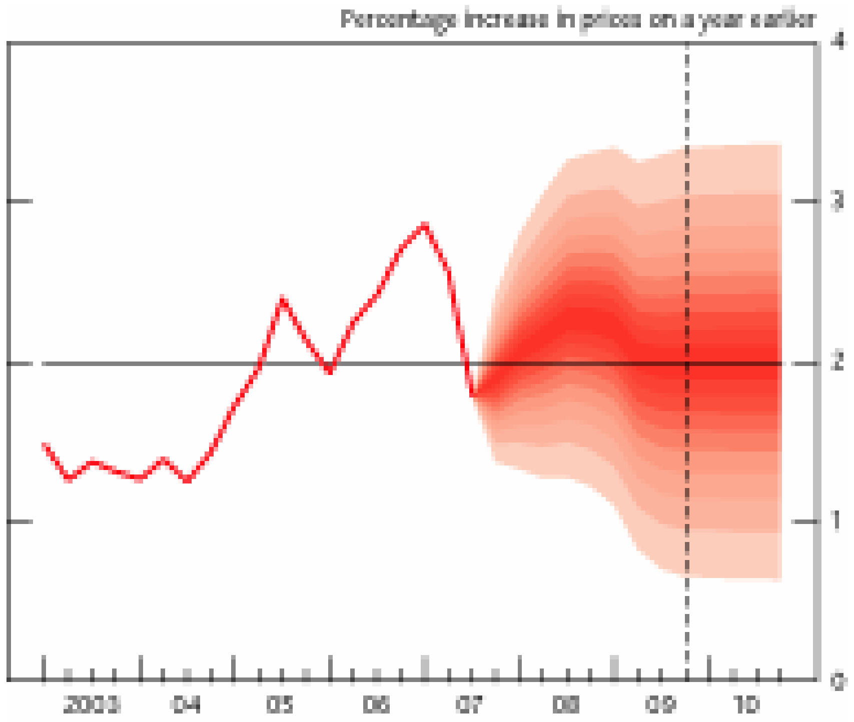
which is an equivalent way of doing the test

and still the rule is reject the null hypothesis if $\hat{F} > F^{(k-1, N-k)}_{\alpha\_critical}$

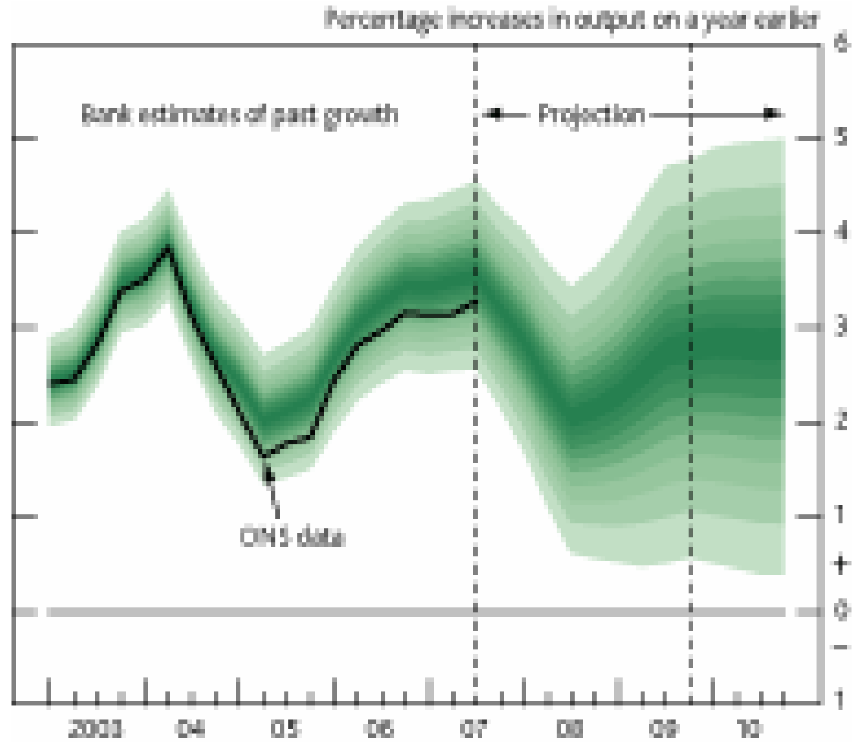
Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

November 2007 CPI Fan Chart



November 2007 GDP Fan Chart



Source Bank of England Inflation report

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$
and a set of observations on the variable X can use this to predict a value for y

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$
and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$
and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

and can do this for any X value even if the X value did not belong to the original data set

(Good practice when estimating equations to leave some observations (eg end few time periods or a random sub-sample of a cross-section) out

If the model is a good one it should predict well out-of-sample.

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Already seen that given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$
and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

and can do this for any X value even if the X value did not belong to the original data set

Forecasting Using OLS

One of the many uses of regression analysis is that can use the coefficient estimates to make predictions about the behaviour of economic agents (individuals, time periods) that may not be included in the original data

Given a set of estimates $\hat{\beta}_0$ and $\hat{\beta}_1$

and a set of observations on the variable X can use this to predict a value for y

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

even if the x value did not belong to the original data set

So it is good practice when estimating equations to leave some observations (eg end few time periods or a random sub-sample of a cross-section) out

If the model is a good one it should predict well out-of-sample.

How do we know how good the prediction is ?

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error which is just the difference between the prediction (the forecast) and the actual (out-of-sample) value

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error which is just the difference between the prediction (the forecast) and the actual (out-of-sample) value

$\hat{u}_0 = \hat{y}_o - y_o$ (o is "out of sample")

How do we know how good the prediction is ?

- use the idea of a **forecast error**

the closer is the estimate $\hat{y}$ to its actual value, the smaller will be the forecast error which is just the difference between the prediction (the forecast) and the actual (out-of-sample) value

$\hat{u}_0 = \hat{y}_o - y_o$ (o is “out of sample”)

and if we use OLS to generate the estimates to make the forecasts then can show that the forecast error will on average be zero and have the smallest variance of any other (linear unbiased) technique.

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(y_o - \hat{y}_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

$$E(\hat{u}_0) = \beta_0 + \beta_1 X_o - \beta_0 - \beta_1 X_o - E(u_0)$$

$$= 0 - E(u_0)$$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(\hat{y}_o - y_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

$$E(\hat{u}_0) = \beta_0 + \beta_1 X_o - \beta_0 - \beta_1 X_o - E(u_0)$$
$$= 0 - E(u_0)$$

and since Gauss-Markov assumption is that for any observation the expected value of the *true* residual is zero, $E(u_0)=0$

Proof: The expected value of the forecast error is

$$E(\hat{u}_0) = E(y_o - \hat{y}_o)$$

sub. in for $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_o$ and $y_O = \beta_0 + \beta_1 X_o + u_o$

$$E(\hat{u}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 X_o - (\beta_0 + \beta_1 X_o + u_o))$$

using what we know about the expected value of OLS estimates,

$E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$

$$E(\hat{u}_0) = \beta_0 + \beta_1 X_o - \beta_0 - \beta_1 X_o - E(u_0)$$

$$= 0 - E(u_0)$$

and since Gauss-Markov assumption is that for any observation the expected value of the *true* residual is zero, $E(u_0)=0$

$E(\hat{u}_0) = 0$ the mean value of any OLS forecast error is zero

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) the larger the sample size

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) the larger the sample size
(not expect a good forecast if only have little information)

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) Larger sample size
(not expect a good forecast if only have little information)
c) Larger variance of in-sample X's

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2\left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
c) Larger sample size
(not expect a good forecast if only have little information)
c) Larger variance of in-sample X's
(wider variance means more likely already encountered a big range of possible X values to make a forecast with)

The variance of each individual forecast error is given by

$$Var(\hat{Y}_0 - Y_o) = Var(\hat{u}_o) = s^2 \left[1 + \frac{1}{N} + \frac{(X_o - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \right]$$

where N is the number of in-sample observations,

s^2 is the residual variance = RSS/N-k

So precision of forecast is increased by

a) the closer is the out of sample observation X_o to mean value of in-sample X's
b) Larger sample size
(not expect a good forecast if only have little information)

c) Larger variance of in-sample X's
(wider variance means more likely already encountered a big range of possible X values to make a forecast with)

d) Better fitting model (implies smaller RSS and smaller s^2. The better the model is at predicting in-sample, the more accurate will be the prediction out-of-sample)

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$

follows a t distribution

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

$$\frac{\hat{u}_0 - \bar{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

$$\frac{\hat{u}_0 - \bar{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

and since $\bar{u}_0 = E(u_0) = 0$

Follows that can construct α% confidence interval for each individual forecast observation (just as with the confidence interval for an unknown true coefficient value) using the square root of the above variance and substituting this into the formula for the confidence interval

Just as

$$\frac{\hat{\beta}_1 - \beta_1}{s.e.(\hat{\beta}_1)}$$ follows a t distribution

using $\hat{u}_0$ rather than $\hat{\beta}_1$ then

$$\frac{\hat{u}_0 - \bar{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

and since $\bar{u}_0 = E(u_0) = 0$

$$\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$$ also follows a t distribution

Given $\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

Given $\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

becomes $\frac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)}$

Given $\dfrac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = y_o - \hat{y}_o$

becomes $\dfrac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)}$

Since this follows a t distribution we know that 95% of observations will lie in the region

$$\Pr\left[-t_{N-k}^{.05/2} \le \frac{y_o - \hat{y}_o}{s.e.(\hat{u}_0)} \le +t_{N-k}^{.05/2}\right] = 0.95$$

Given $\frac{\hat{u}_0}{s.e.(\hat{u}_0)}$

Sub. in $\hat{u}_0 = \hat{y}_o - y_o$

becomes $\frac{\hat{y}_o - y_o}{s.e.(\hat{u}_0)}$

Since this follows a t distribution we know that 95% of observations will lie in the region

$$\Pr\left[-t_{N-k}^{.05/2} \leq \frac{\hat{y}_o - y_o}{s.e.(\hat{u}_0)} \leq +t_{N-k}^{.05/2}\right] = 0.95$$

and rearranging terms gives

$$\Pr[\hat{Y}_o - t_{N-k}^{\alpha/2} s.e.(\hat{u}_o) \leq Y_o \leq \hat{Y}_o + t_{N-k}^{\alpha/2} s.e.(\hat{u}_o)] = 0.95$$

Given

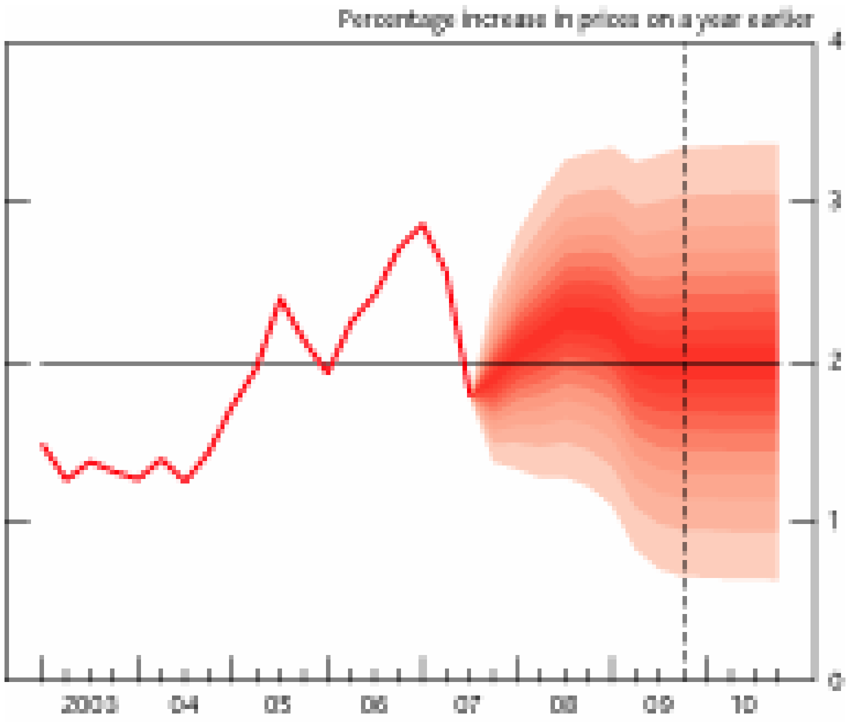
$$\Pr[\hat{Y}_o - t_{N-k}^{\alpha/2} s.e.(\hat{u}_o) \le Y_o \le \hat{Y}_o + t_{N-k}^{\alpha/2} s.e.(\hat{u}_o)] = 0.95$$

$$= \hat{Y}_o \pm t_{\alpha/2} SE(\hat{u}_o)$$

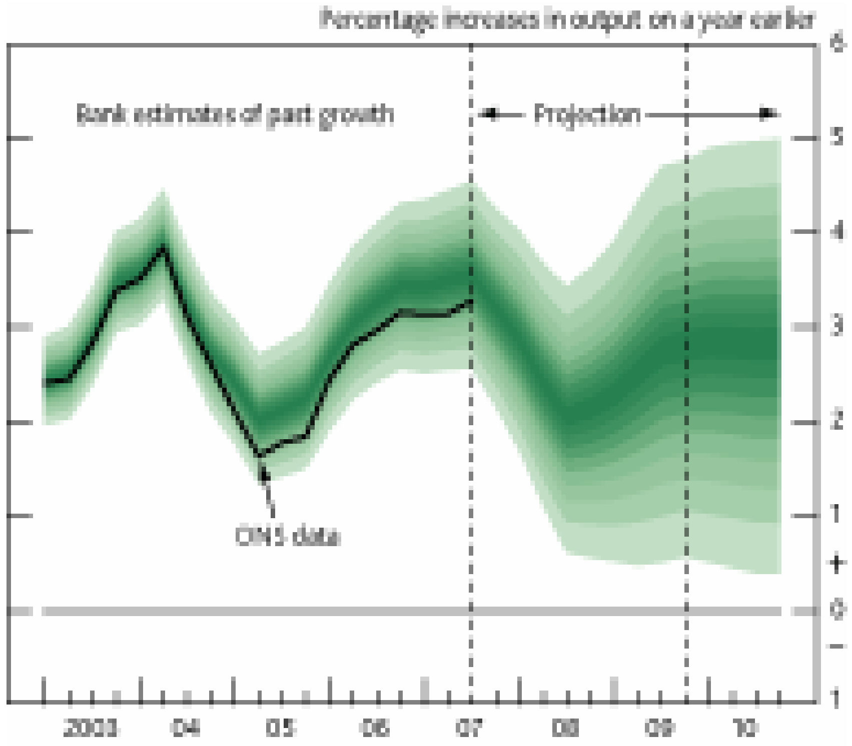
then we can be 95% confident that the true value will lie within this range

If it does not then the model does not forecast very well

November 2007 CPI Fan Chart



November 2007 GDP Fan Chart

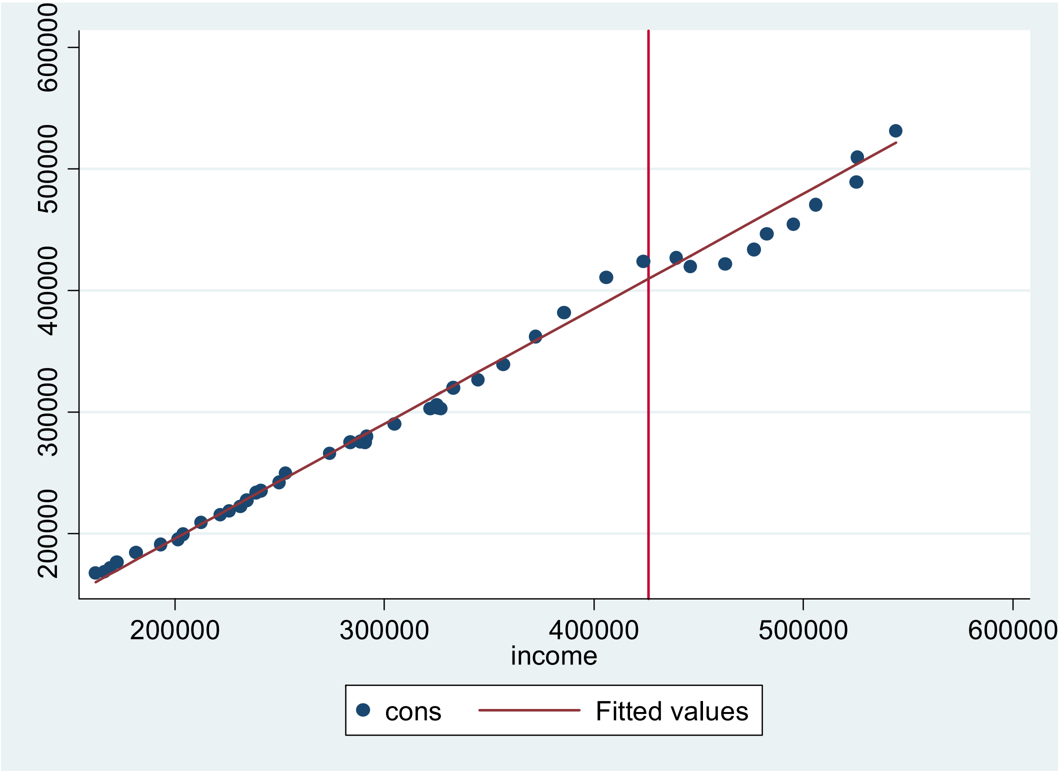


Source Bank of England Inflation report

Example

```
 reg cons income if year<90   /* use 1st 35 obs and save last 10 for forecasts*/

  Source |       SS       df       MS                  Number of obs =      35
---------+------------------------------               F(  1,    33) = 3190.74
   Model |  1.5750e+11     1  1.5750e+11               Prob > F      =  0.0000
Residual |  1.6289e+09    33  49361749.6               R-squared     =  0.9898
---------+------------------------------               Adj R-squared =  0.9895
   Total |  1.5913e+11    34  4.6803e+09               Root MSE      =  7025.8

------------------------------------------------------------------------------
    cons |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  income |   .9467359   .0167604     56.487   0.000       .9126367    .9808351
   _cons |   6366.214   4704.141      1.353   0.185      -3204.433    15936.86

predict chat            /* gets fitted (predicted) values from regression */

predict forcse, stdf    /* gets standard error around forecast value */

/* graph actual and fitted values, draw line through OLS predictions */
```

```
twoway (scatter cons income) (line chat income, xline(426000))
```

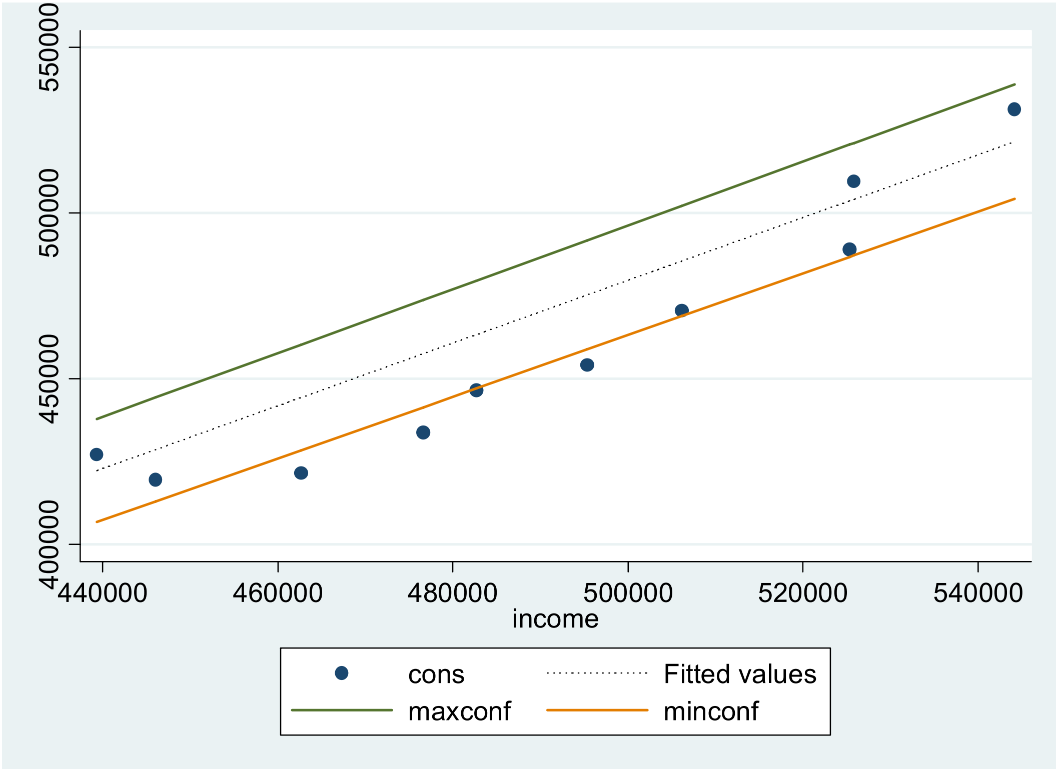


Vertical line denotes boundary between in and out of sample observations

```
/* now calculate confidence intervals using cons =
```

$$cons = \hat{cons} \pm t_{\partial/2}^{N-k} * SE(\hat{u})$$

```
 g minconf= chat - (2.04*forcse)
 g maxconf= chat + (2.04*forcse)

/* graph predicted consumption values and confidence intervals */
twoway (scatter cons income if year>89) (line chat income if year>89, lstyle(dot))  (line maxconf income if year>89)
> (line minconf income if year>89)
```



So for most of the out of sample observations the actual value lies *outside* the 95% confidence interval. Hence the predictions of this particular model are not that good.

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect ***all*** the out-of-sample residuals to be close to zero.

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero.
It can be shown that the joint test of all the out-of-sample-residuals being close to zero is given by:

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N-k]$$

where N_o is the number of out-of-sample observations
N is the number of in-sample observations
k is the number of RHS coefficients

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero.
It can be shown that the joint test of all the out-of-sample-residuals being close to zero is given by:

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N-k]$$

where N_o is the number of out-of-sample observations
N is the number of in-sample observations
k is the number of RHS coefficients

Intuitively if the model fits well the RSS from the combined regression should be close to that from the in-sample regression. A “large” difference suggest the RSS are different and so model does not forecast well)

Chow Test of Forecast Stability

If a model forecasts well out of sample then we would expect all the out-of-sample residuals to be close to zero.
It can be shown that the joint test of all the out-of-sample-residuals being close to zero is given by:

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N-k]$$

where N_o is the number of out-of-sample observations
N is the number of in-sample observations
k is the number of RHS coefficients

Intuitively if the model fits well the RSS from the combined regression should be close to that from the in-sample regression. A "large" difference suggest the RSS are different and so model does not forecast well)

Given a null hypothesis that the model is stable out of sample (predicts well) then if

$$\hat{F} > F_{critical}{}^{\alpha}[N_o, N-k]$$

reject null of model stability out-of-sample

Example:
Estimate of consumption function model including out-of-sample observations is

```
. reg cons income

      Source |       SS       df       MS              Number of obs =      45
-------------+------------------------------           F(  1,    43) = 5969.79
       Model |  4.7072e+11     1  4.7072e+11           Prob > F      =  0.0000
    Residual |  3.3905e+09    43  78849774.6           R-squared     =  0.9928
-------------+------------------------------           Adj R-squared =  0.9927
       Total |  4.7411e+11    44  1.0775e+10           Root MSE      =  8879.7
------------------------------------------------------------------------------
        cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .9172948   .0118722    77.26   0.000     .8933523    .9412372
       _cons |   13496.16   4025.456     3.35   0.002      5378.05    21614.26
------------------------------------------------------------------------------
```

Comparing RSS from this with that above can calculate

$$F = \frac{RSS_{in+out} - RSS_{in} / N_o}{RSS_{in} / N - k} \sim F[N_o, N-k] \;=\; \frac{3.3905 - 1.6289/10}{1.6289/35-2} \sim F[10,35-2]$$

$$= 3.57 \sim F[10,33]$$

From tables $F_{critical}^{.05}[10,33] = 2.10$

So $\hat{F} > F_{critical}$ and therefore **reject** null that model predicts well out of sample.