ECON 343
Lecture 2 : Review of the basic multiple regression model

Jad Chaaban

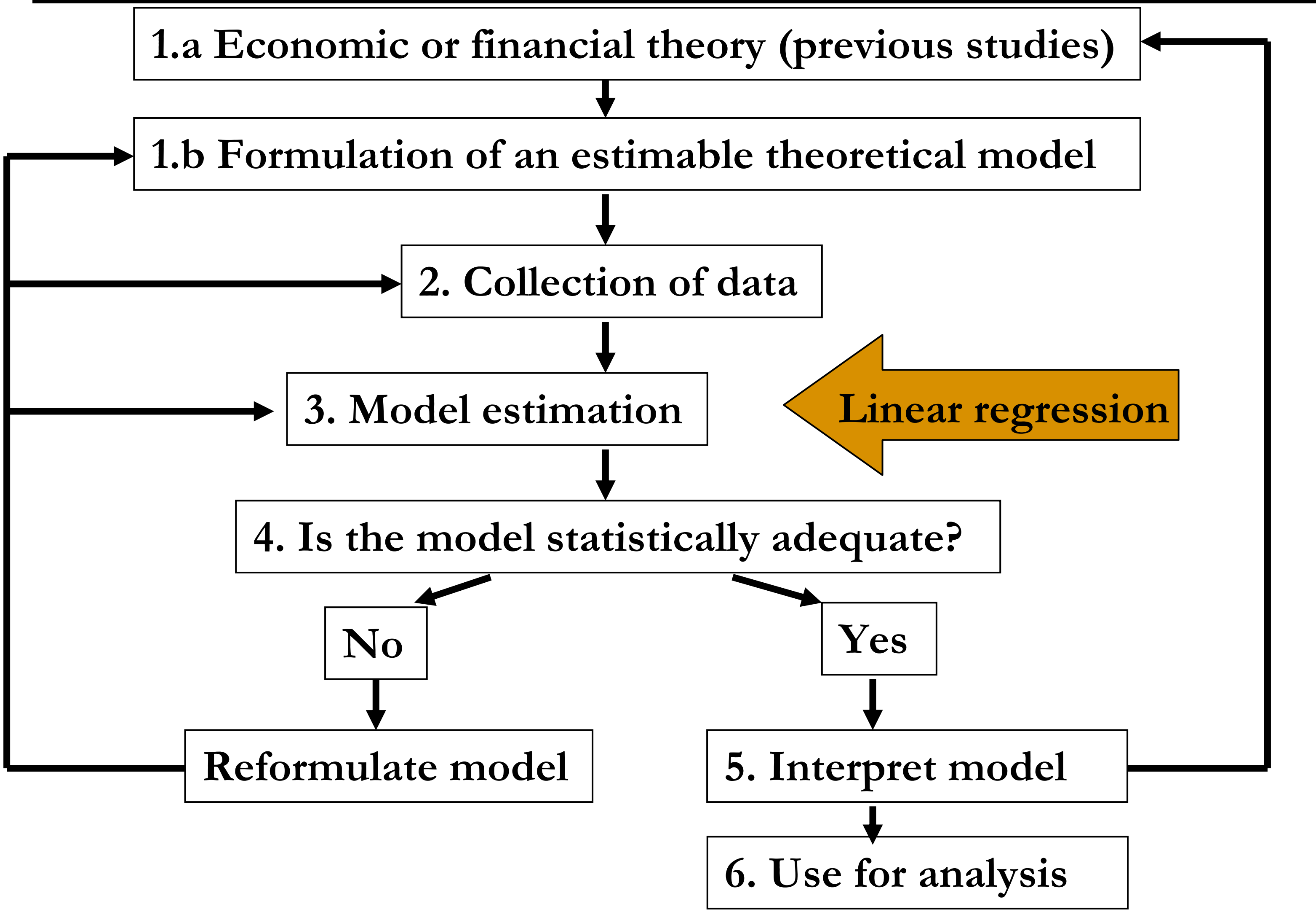
Spring 2005-2006

Outline – Lecture 2

- PART I
  - The classical linear regression model
  - Assumptions underlying OLS estimation
  - Properties of OLS estimators
  - Statistical inference and hypothesis testing
- PART II
  - Goodness of fit and regression diagnostics
  - Regression diagnostics
  - Example: investment and GDP

PART I

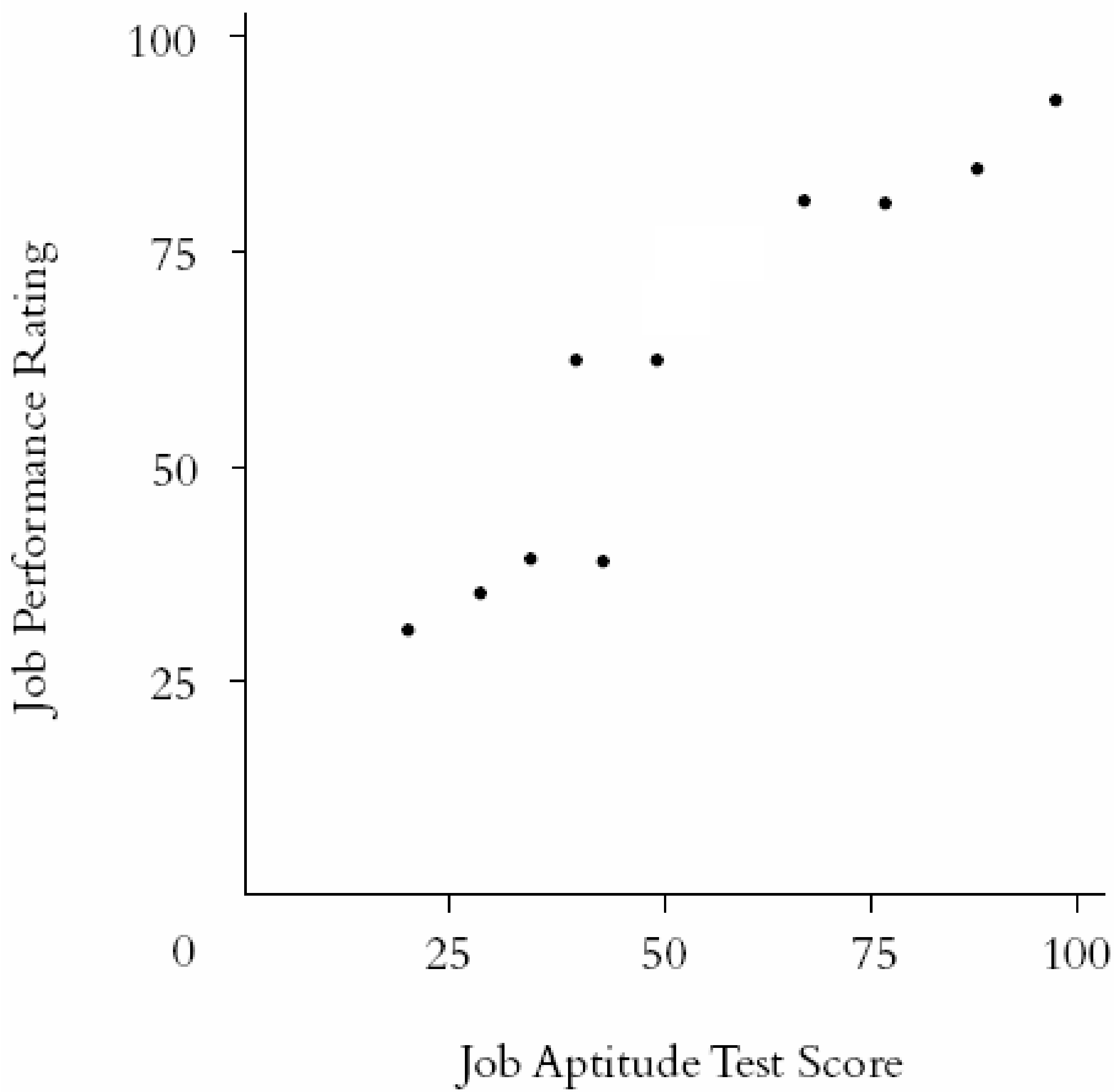
Steps in an econometric model

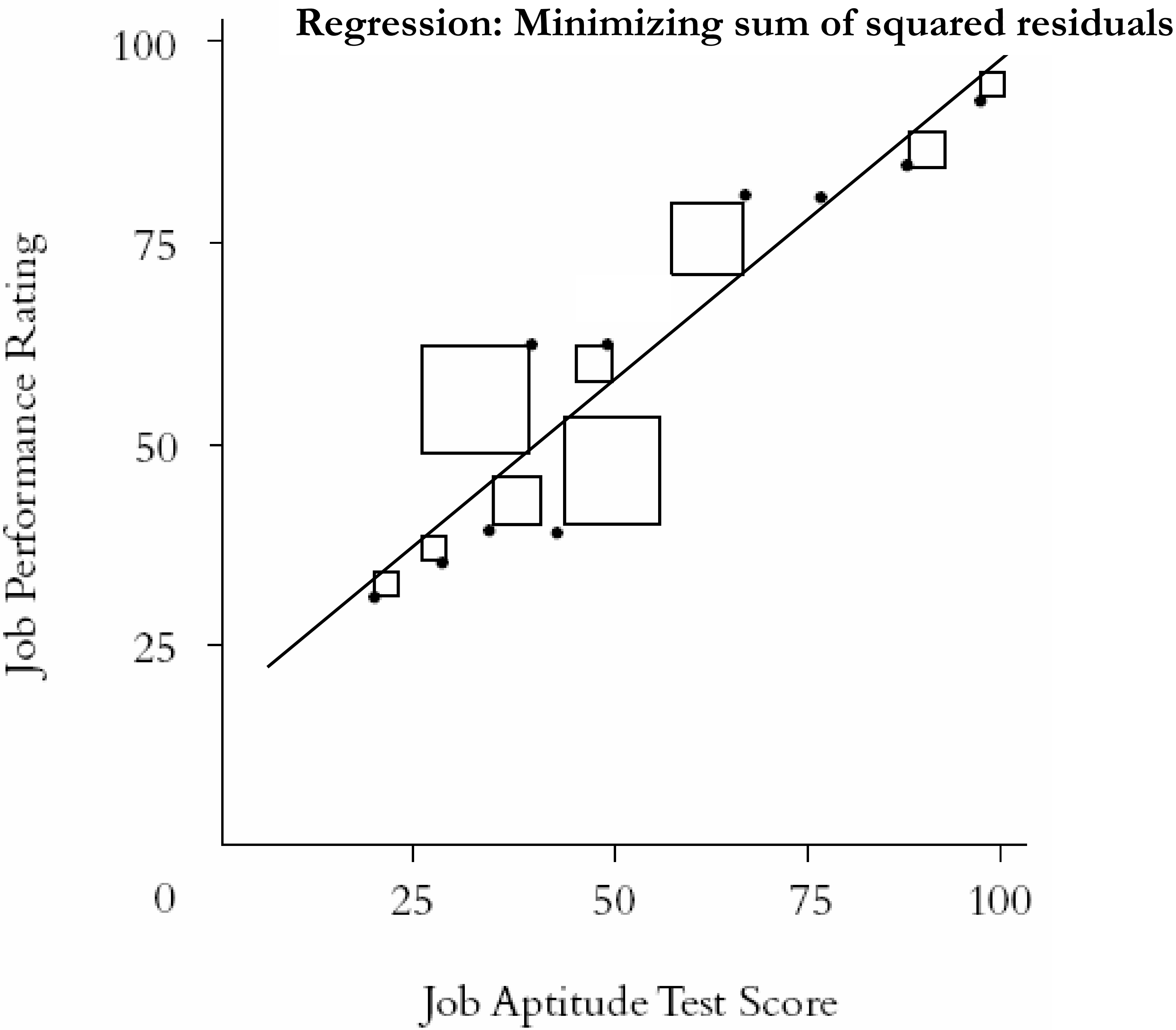


Ordinary Least Squares OLS

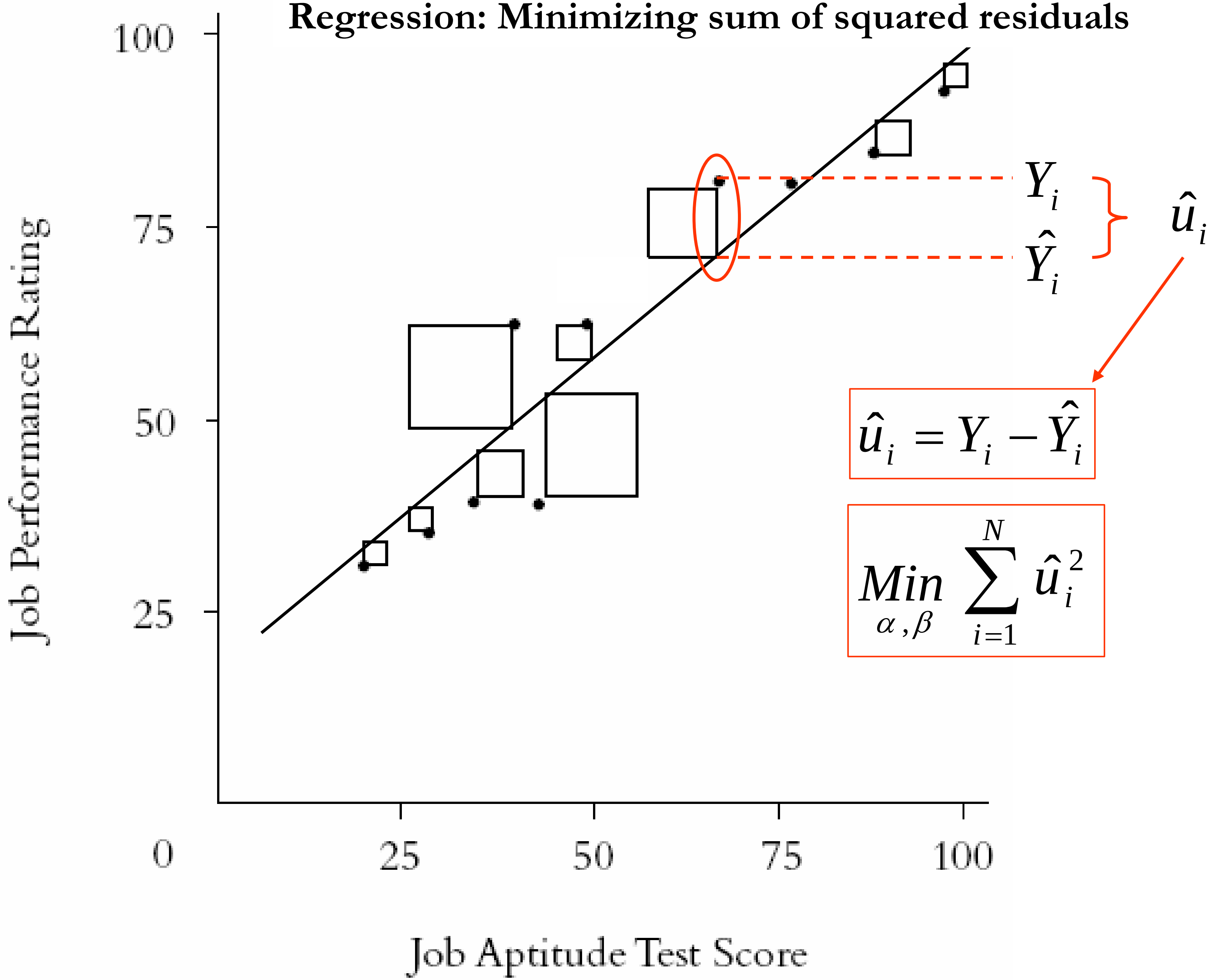
- Method in which a regression line is used to relate the average of one variable, the *dependent variable*, to the values of other *explanatory variables*
- *Y* dependent random variable, *X* independent variables
- Simple model: $$Y_i = \alpha + \beta X_i + u_i \qquad \forall i = \{1, \ldots, N\}$$
- Disturbance term *u* captures:
  - Omitted explanatory variables
  - Measurement error in *Y*
  - Unexplained randomness
- Parameter are chosen to minimize collectively the vertical distances from the data points to the fitted line

Illustration: Scatter plot, aptitude test scores and job performance




Regression: Minimizing sum of squared residuals
100
75
50
25
0
25
50
75
100
Job Performance Rating
Job Aptitude Test Score

Regression: Minimizing sum of squared residuals



$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$\underset{\alpha,\beta}{Min} \sum_{i=1}^{N} \hat{u}_i^2$$

Ordinary Least Squares OLS

- Formulae for estimators in the 2 variables model:

Slope $$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$ $$\bar{Y} = \frac{\sum Y_i}{N}$$

Intercept $$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$ $$\bar{X} = \frac{\sum X_i}{N}$$

- The difference between estimators and estimates
  - Estimators: formulae to calculate the coefficients
  - Estimates: actual numerical values for the coefficients

Assumptions underlying OLS

- 1. The errors have zero mean

$$E(u_i) = 0$$

- 2. The variance of the errors is constant and finite over all values of x_t

$$\text{var}(u_i) = \sigma^2 < \infty$$

- 3. The errors are statistically independent of one another

$$\text{cov}(u_i, u_j)$$

- 4. There is no relationship between the error and corresponding x variable

$$\text{cov}(u_i, x_i)$$

- 5. The errors are normally distributed

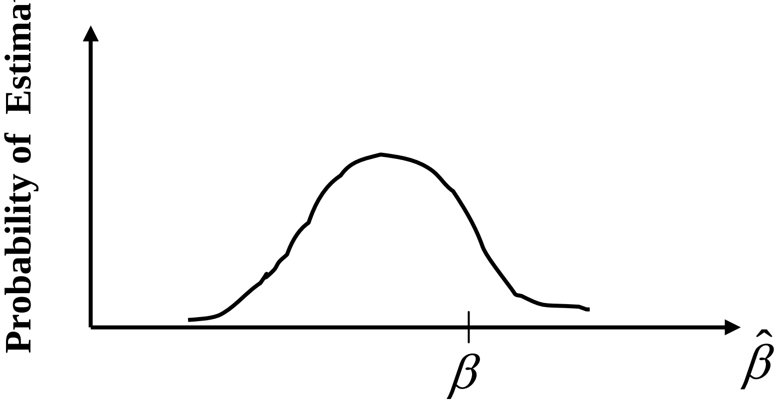
$$u_i \to N(0, \sigma^2)$$

Properties of the OLS estimator

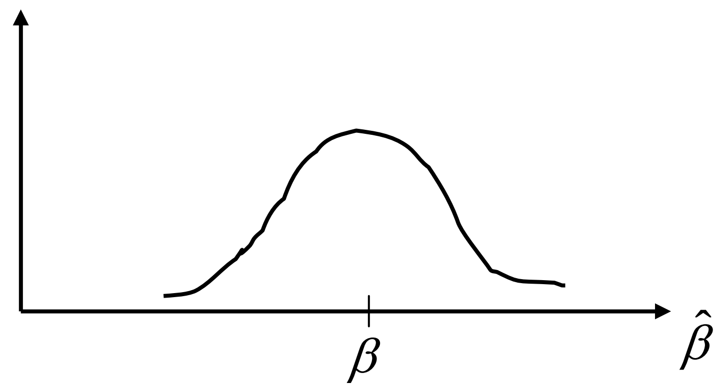
- Consistency: the estimates will converge to their true value as the sample size increases to infinity
- Unbiasedness: On average, the estimated values for the coefficients will be equal to their true values
- Efficiency: an estimator is said to be efficient if no other estimator has a smaller variance

- OLS is BLUE: Best Linear Unbiased Estimator

BIAS

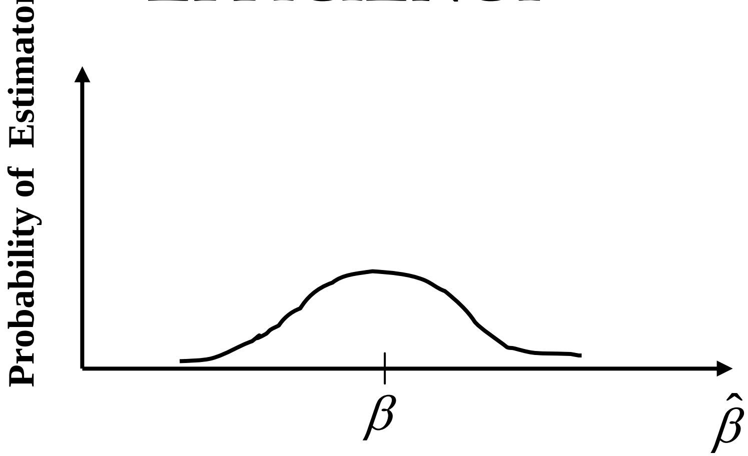


Biased Estimator

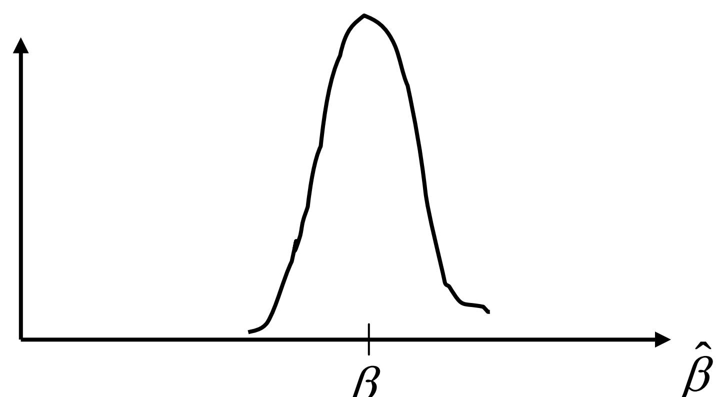


Unbiased Estimator

EFFICIENCY



Inefficient Estimator



Efficient Estimator

Statistical inference: t-statistic

- The coefficient standard error measures the reliability of the estimate
  - A larger standard error indicates a wider confidence interval in which the true value of the coefficient can be expected

- The Student t-statistic tests whether the coefficient is found to be significantly different from zero
  - The reported probability tells us the probability of finding a coefficient larger (when positive, smaller when negative) than the estimated coefficient when in fact the true coefficient is zero
  - Any probability value smaller than the critical significance level, for example 0.05 for a 5% significance level, leads us to reject the null hypothesis that the coefficient is zero (in this case, in a one-sided test)

Statistical inference: t-statistic

- The t-statistic allows one to test hypotheses on the value of single coefficients:

$$\frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \rightarrow t_{N-K}$$

  - The estimated regression parameters, normalized by subtracting the mean and dividing by the *estimated* standard error, follow the t distribution with N-K degrees of freedom

- Test null hypothesis H_0 versus the alternative hypothesis H_1
- Example:

$$H_0 : \beta_i = 1$$

$$H_1 : \beta_i \neq 1$$

Statistical inference: F-statistic

- **The F-test allows one to test multiple hypotheses on the coefficient estimates:**

$$statistic = \frac{RRSS - URSS}{URSS} \times \frac{N - K}{m}$$

which follows a F-distribution

- ***URSS*: residual sum of squares from unrestricted regression**
- ***RRSS*: residual sum of squares from restricted regression**
- ***m*: number of restrictions**
- ***N*: number of observations**
- ***K*: number of variables in unrestricted regression**

Example:

Unrestricted: $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

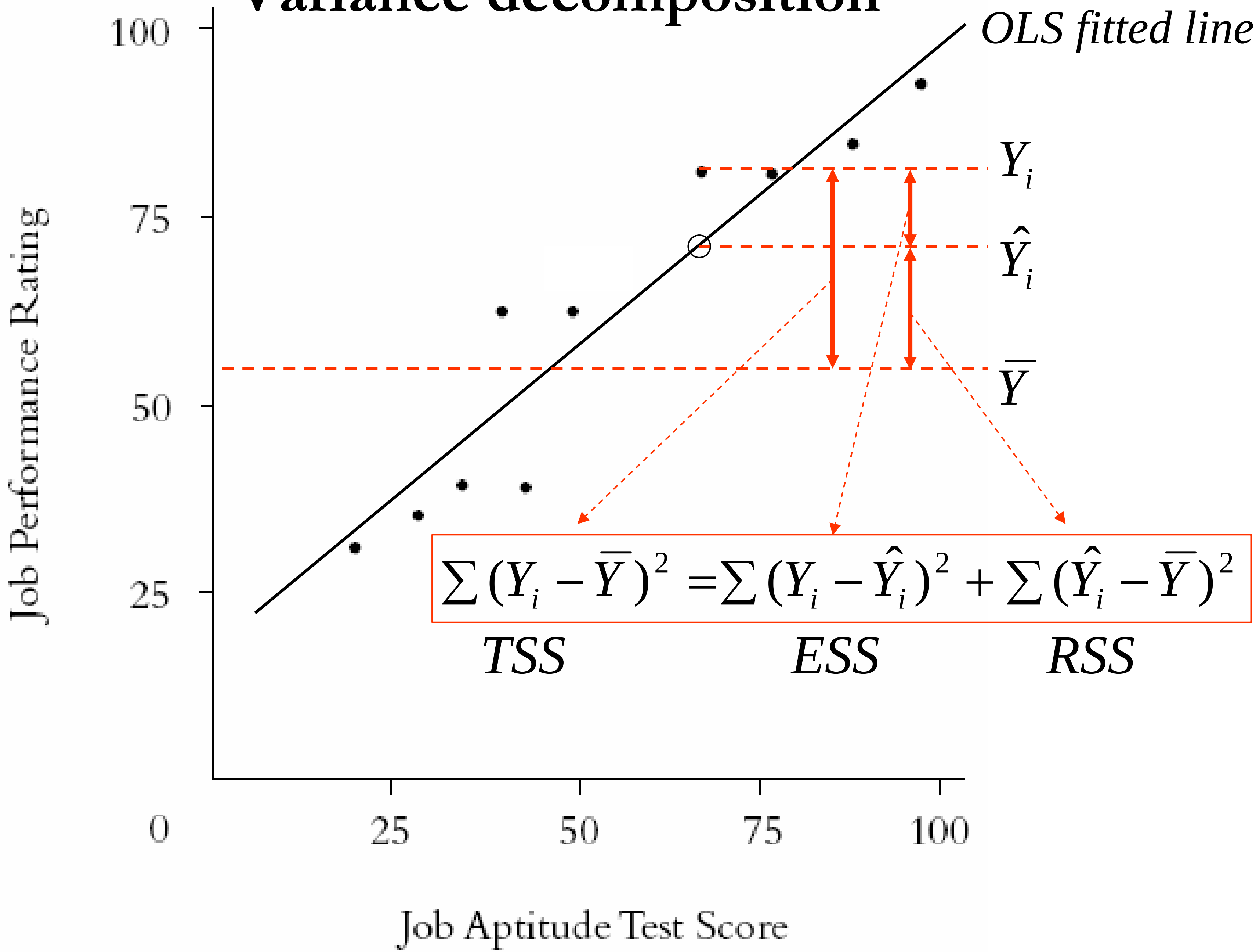
Restricted: $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

Subject to $\beta_1 + \beta_2 = 1$

One restriction

PART II

Variance decomposition



R-squared

- **Definition of R-squared:**

$$R^2 = \frac{RSS}{TSS}$$

- **When all explanatory variables are perfectly useless and are assigned coefficients of 0, the regression amounts to the linear equation**

$$Y_i = \alpha + u_i$$

- **In this case the RSS in R^2 is equal to 0 and the value of R^2 becomes 0**

Adjusted R-squared

- This is a corrected value of R2, which is computed as:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{(n-1)}{(n-k-1)}$$

with n: number of observations and k: number of variables

- The rationale for the correction is that the uncorrected value of R^2 always increases if we add one more explanatory variable to a regression model, although the added variable may in fact be useless
- Therefore, it is not possible to choose between equations based on R^2
- The adjusted-R^2 only increases when "useful" additional explanatory variables are added to the equation, where useful is defined a having a t-value greater in absolute value than 1

Goodness of fit

- The R-squared for the regression, which by construction is always between 0 and 1 inclusive, indicates the degree of 'fit' of the regression
  - A value of 0 indicates that the regression is perfectly useless in explaining the dependent variable
  - A value of 1 signifies a perfect fit between the dependent variable and the linear combination of the explanatory variables
- The goal is frequently (but not always) to get a high R^2
- In financial economics we often do not achieve values higher than 0.15 and often desire values close to zero due to the theory of efficient markets
  - Note that a value of 1 indicates a wrong design of the regression because a perfect fit could be realized only if a linear combination of the explanatory variables were identical to the dependent variable by definition

Regression diagnostics

- Examine the plausibility of estimated coefficients:
  - What is the associated economic theory?
  - Are sign and size of the coefficients consistent with the theory and/or other studies?
- Check significance of each coefficient using t-tests
- Check goodness of fit using the Adjusted R-squared
- Perform joint hypothesis testing using the Fisher F-statistic

The F-test

- The formula for this test is:

$$F = \frac{R^2 / k}{(1 - R^2)/(n - k - 1)}$$

  where k equals the number of explanatory variables (excl. the constant) and n equals the number of usable observations, so that (n-k-1) corresponds to the number of degrees of freedom
- This test statistic follows a F-distribution with k degrees of freedom for the numerator and (n-k-1) degrees of freedom for the denominator (see standard tabulations)
- EViews provides the probability of finding a larger F-test value assuming that the regression equation is in fact useless (the null hypothesis)
- Usually, we accept that the regression model is useful when the Prob(F-statistic) is smaller than the desired significance level, for example 0.05 (for 5% significance level)

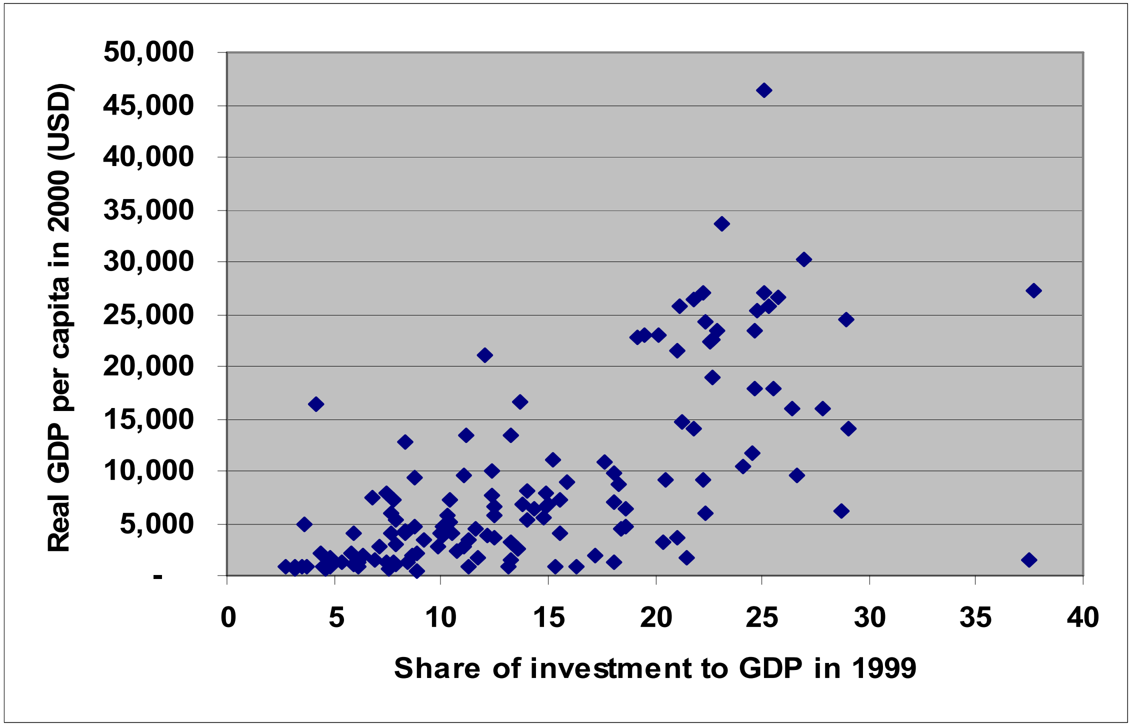
Choosing among different regressions

- Akaike Information Criterion AIC is used to test the relative value of different competing models (i.e. alternative regression equations for a dependent variable). When one equation has a lower AIC value than another equation, the equation with the lower AIC is preferred

- The Shwarz Information Criterion SIC is an alternative to the AIC. When one equation has a lower SIC value than another equation, the equation with the lower SIC is preferred

Example: Investment and GDP

- Does higher investment lead to higher GDP?
- Cross-section of 134 countries, Penn World Tables
- Investment variable lagged one year (to avoid simultaneity)

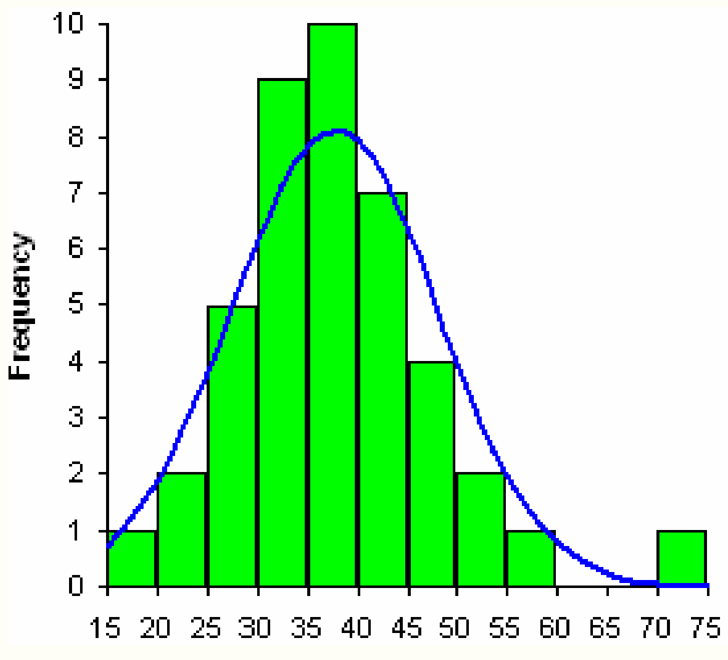


A digression: is world GDP normally distributed?

- Interesting in order to assess world inequality
- In general, 3 ways to check the distribution of a continuous variable:
  - Frequency histogram graph
  - Skewness
  - Kurtosis

Frequency Histogram

A frequency histogram shows the distribution of the observations of a sample. It is usually used to visually assess the scatter and whether the observations are normally distributed.



If the observations are normally distributed the heights of the columns should be roughly shaped like the Normal distribution curve (the superimposed blue line).

Skewness

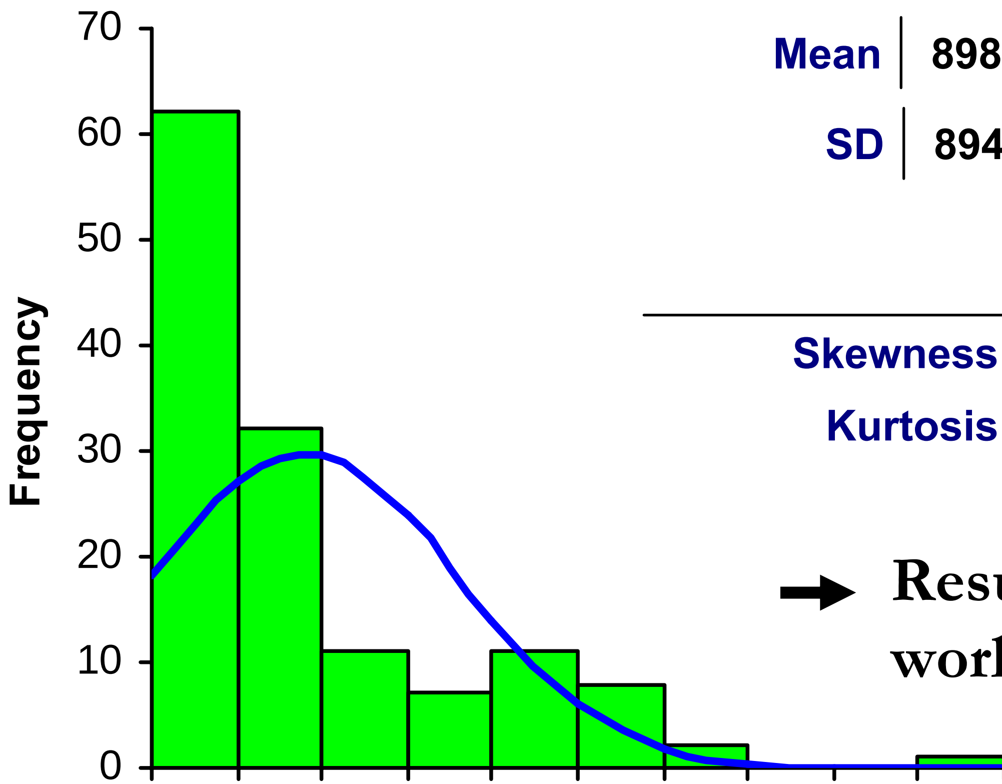
- Skewness is a measure of the symmetry of a distribution around its mean
  - = 0 Distribution is symmetrical about the mean
  - > 0 Distribution has a right-tail skew - more observations in the left-tail than normal. Consider using a log transformation to rectify the skew
  - < 0 Distribution has a left-tail skew - more observations in the right-tail than normal. Consider using an exp (inverse log) transformation to rectify the skew
- A low p-value indicates the skewness is significantly non-normal

Kurtosis

- Kurtosis is a measure of the peakiness of a distribution
  - = 0 Distribution is normal
  - > 0 Distribution is more peaked than normal - more observations are clustered around the mean, with fewer in the tails, than normal.
  - < 0 Distribution is squatter than normal - more observations are in the tails, with less clustered around the mean, than normal.
- A low p-value indicates the kurtosis is significantly non-normal.

Normality check for World GDP

n	134
Mean	8985.173
SD	8945.2286

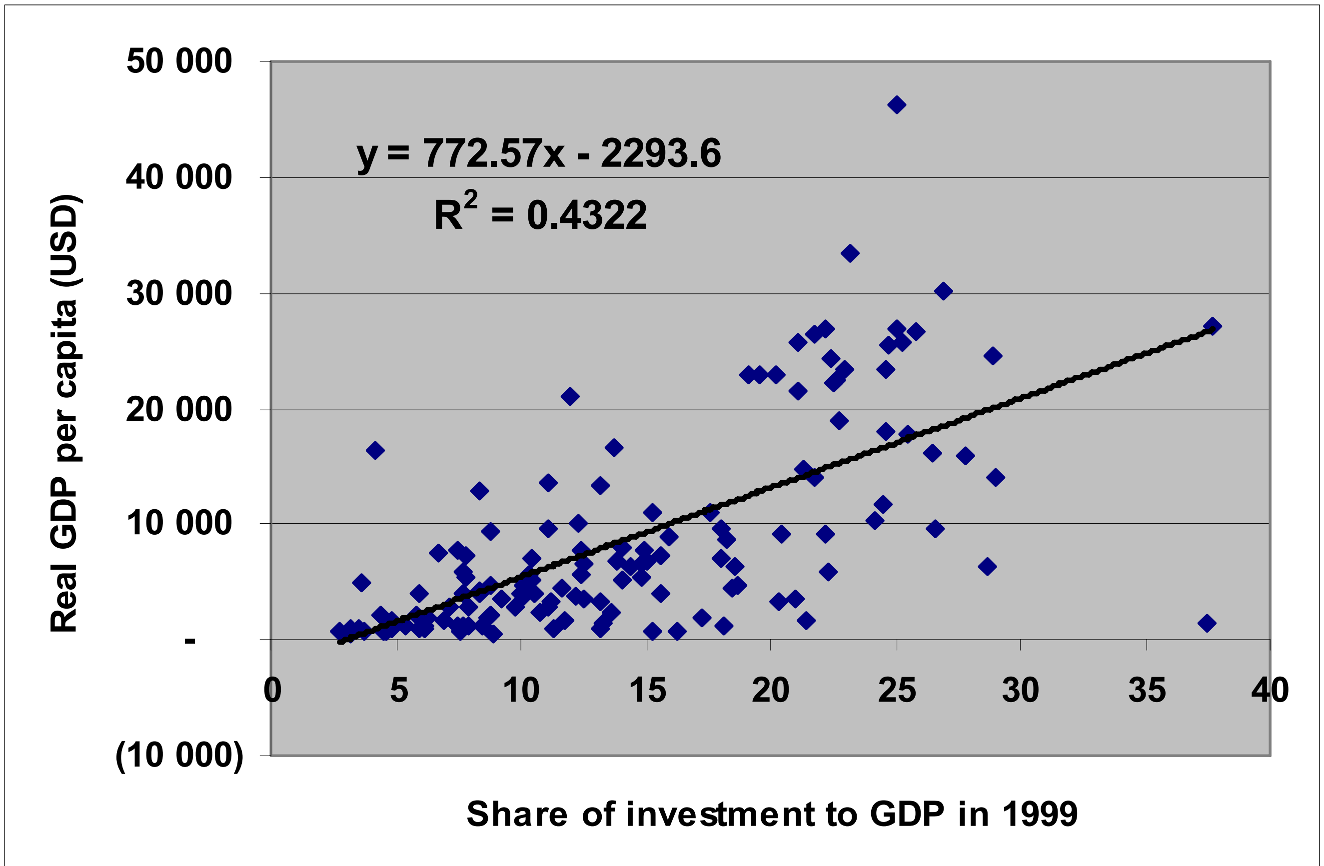


	Coefficient	p
Skewness	1.4056	<0.0001
Kurtosis	1.6567	0.0070

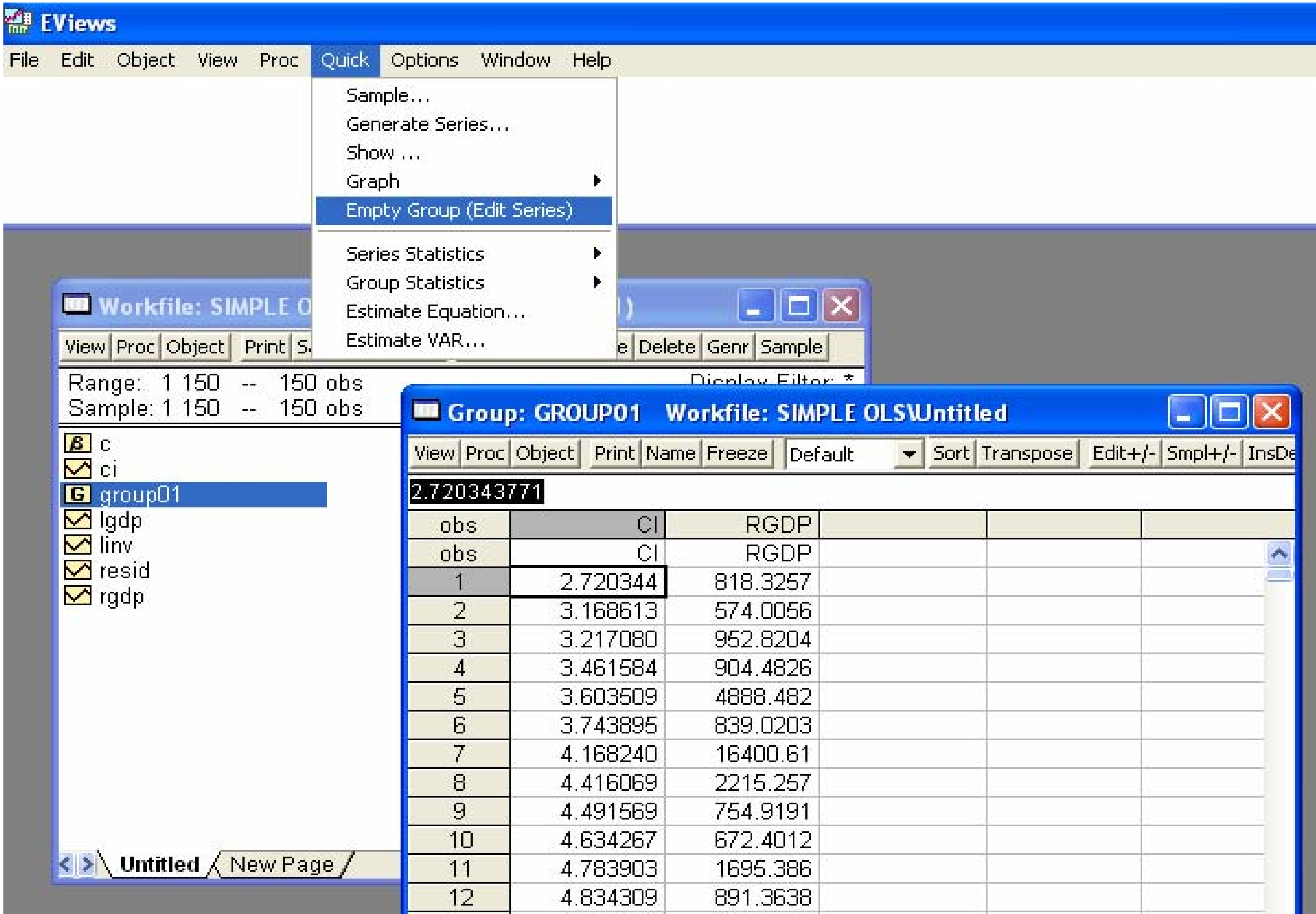
→ Result: high inequality in the world

Regression: GDP vs. Investment

- y=real GDP per capita ; x=share of investment to GDP
- $y_i = \alpha + \beta x_i + u_i$ for i=1…134



Regression: GDP vs. Investment



Regression: GDP vs. Investment

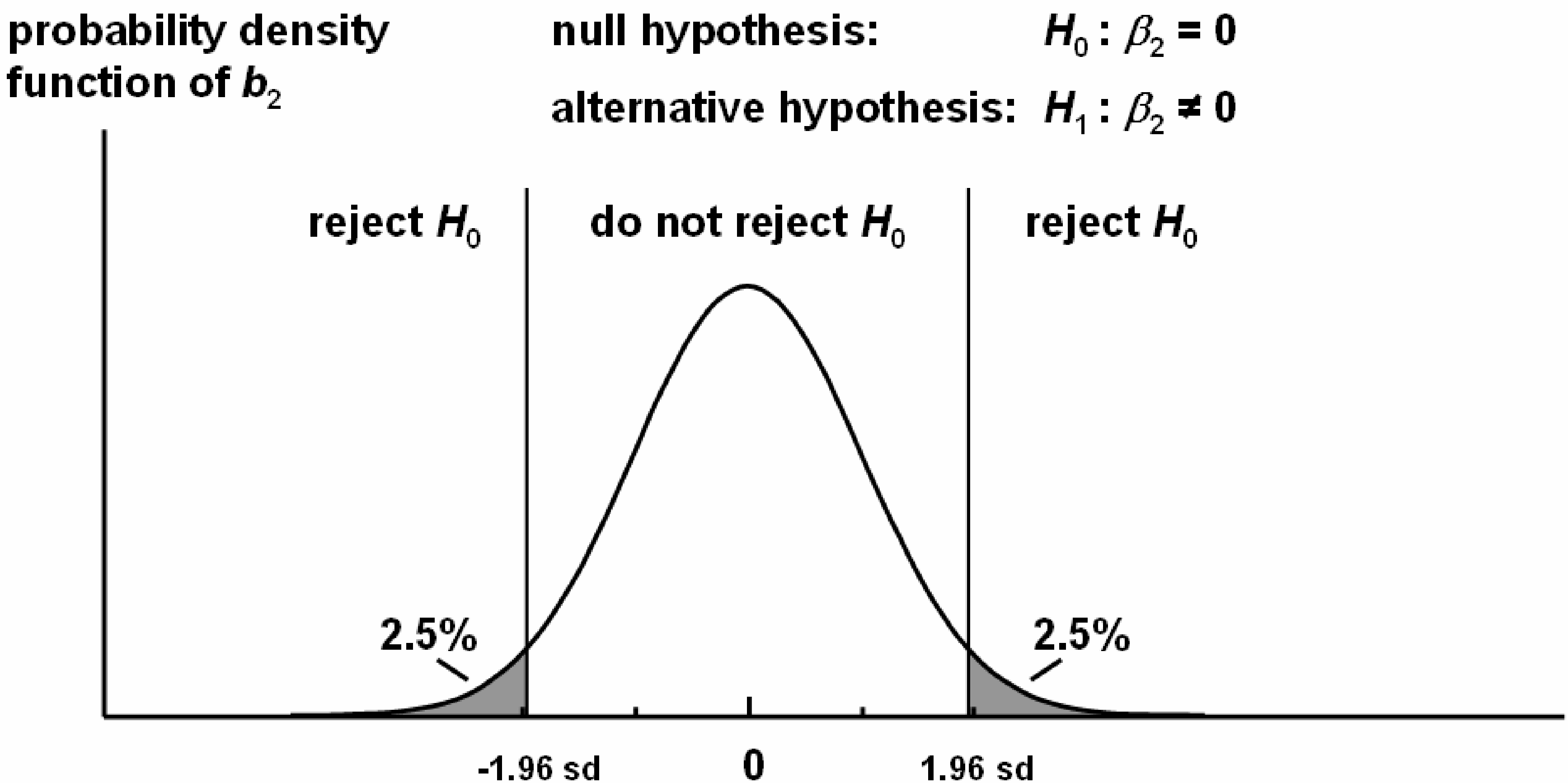
Equation: UNTITLED   Workfile: SIMPLE OLS\Untitled

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RGDP
Method: Least Squares
Date: 02/07/06   Time: 16:48
Sample (adjusted): 1 134
Included observations: 134 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2293.564	1267.900	-1.808946	0.0727
CI	772.5703	77.07033	10.02422	0.0000

R-squared	0.432222	Mean dependent var	8985.173
Adjusted R-squared	0.427920	S.D. dependent var	8945.229
S.E. of regression	6765.809	Akaike info criterion	20.49196
Sum squared resid	6.04E+09	Schwarz criterion	20.53522
Log likelihood	-1370.962	F-statistic	100.4851
Durbin-Watson stat	1.768304	Prob(F-statistic)	0.000000



If you use a two-tailed 5% significance test, your estimate must be 1.96 standard deviations above or below 0 if you are to reject H_0.

s.d. of b_2 known

discrepancy between hypothetical value and sample estimate, in terms of s.d.:

$$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$$

5% significance test:

reject H_0: $\beta_2 = \beta_2^0$ if

$z > 1.96$ or $z < -1.96$

The diagram summarizes the procedure for performing a 5% significance test on the slope coefficient of a regression under the assumption that we know its standard deviation.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

s.d. of b_2 known	s.d. of b_2 not known

discrepancy between hypothetical value and sample estimate, in terms of s.d.:

$$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$$

5% significance test:

reject H_0: $\beta_2 = \beta_2^0$ if

$z > 1.96$ or $z < -1.96$

discrepancy between hypothetical value and sample estimate, in terms of s.e.:

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$$

This is a very unrealistic assumption. We usually have to estimate it with the standard error, and we use this in the test statistic instead of the standard deviation.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

s.d. of b_2 known

discrepancy between hypothetical value and sample estimate, in terms of s.d.:

$$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$$

5% significance test:

reject H_0: $\beta_2 = \beta_2^0$ if

$z > 1.96$ or $z < -1.96$

s.d. of b_2 not known

discrepancy between hypothetical value and sample estimate, in terms of s.e.:

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$$

Because we have replaced the standard deviation in its denominator with the standard error, the test statistic has a *t* distribution instead of a normal distribution.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

s.d. of b_2 known	s.d. of b_2 not known
discrepancy between hypothetical value and sample estimate, in terms of s.d.:	discrepancy between hypothetical value and sample estimate, in terms of s.e.:
$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$
5% significance test:	5% significance test:
reject H_0: $\beta_2 = \beta_2^0$ if	reject H_0: $\beta_2 = \beta_2^0$ if
$z > 1.96$ or $z < -1.96$	$t > t_{crit}$ or $t < -t_{crit}$

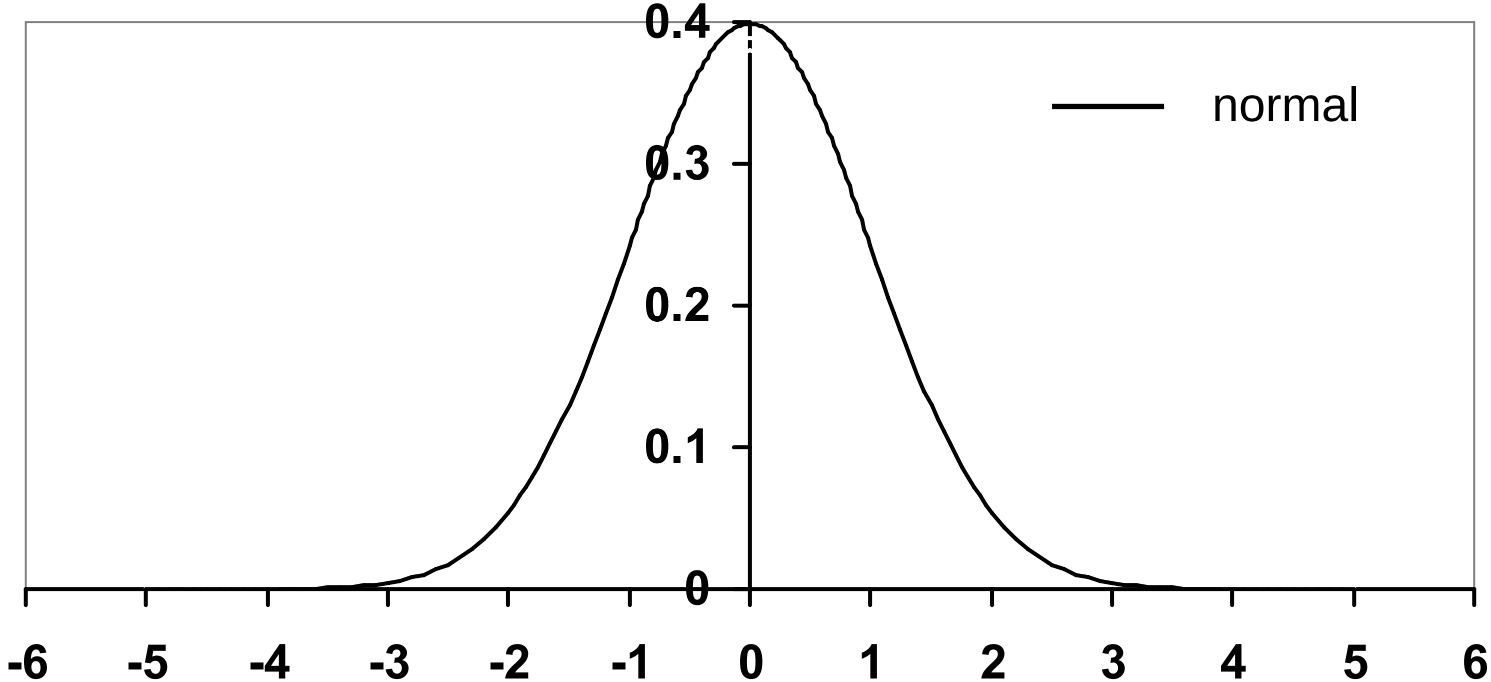
Accordingly, we refer to the test statistic as a *t* statistic. In other respects the test procedure is much the same.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

s.d. of b_2 known	s.d. of b_2 not known
discrepancy between hypothetical value and sample estimate, in terms of s.d.:	discrepancy between hypothetical value and sample estimate, in terms of s.e.:
$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$
5% significance test:	5% significance test:
reject H_0: $\beta_2 = \beta_2^0$ if	reject H_0: $\beta_2 = \beta_2^0$ if
$z > 1.96$ or $z < -1.96$	$t > t_{crit}$ or $t < -t_{crit}$

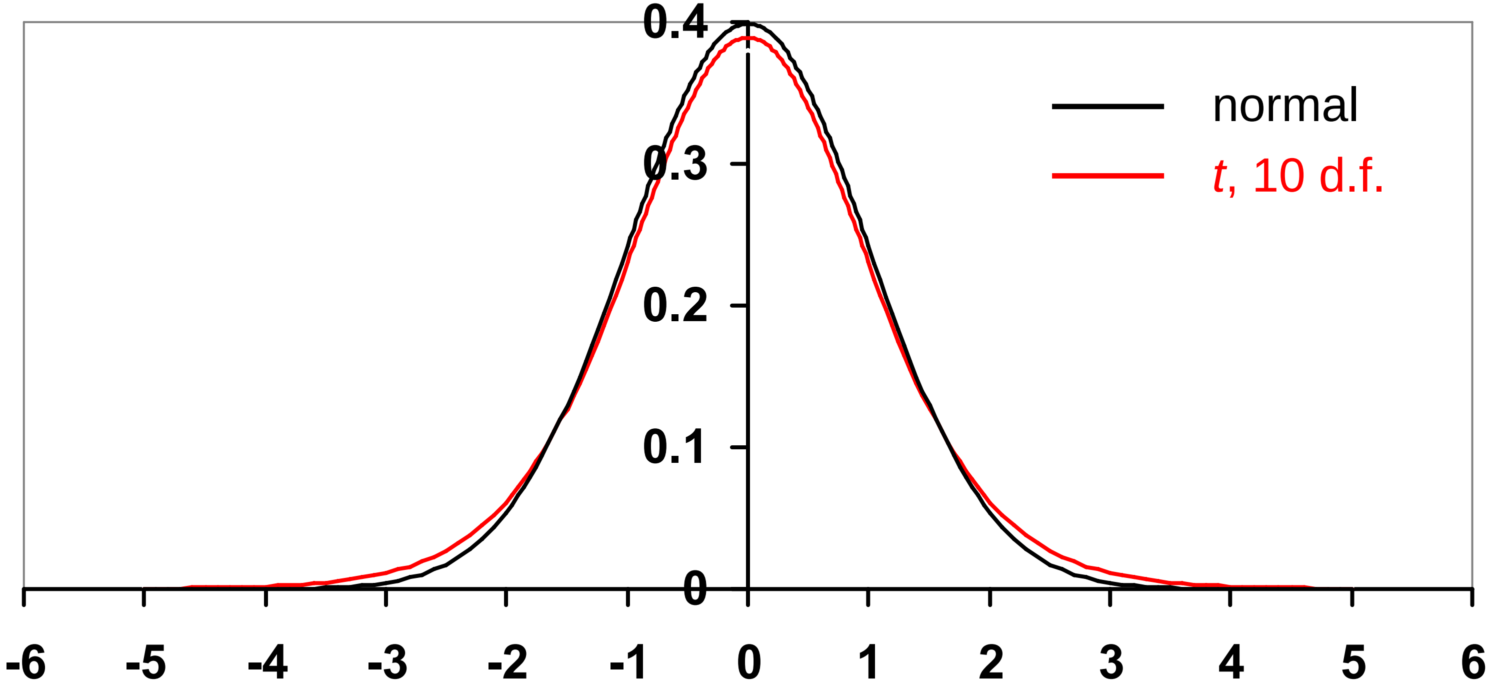
We look up the critical value of t and if the t statistic is greater than it, positive or negative, we reject the null hypothesis. If it is not, we do not.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



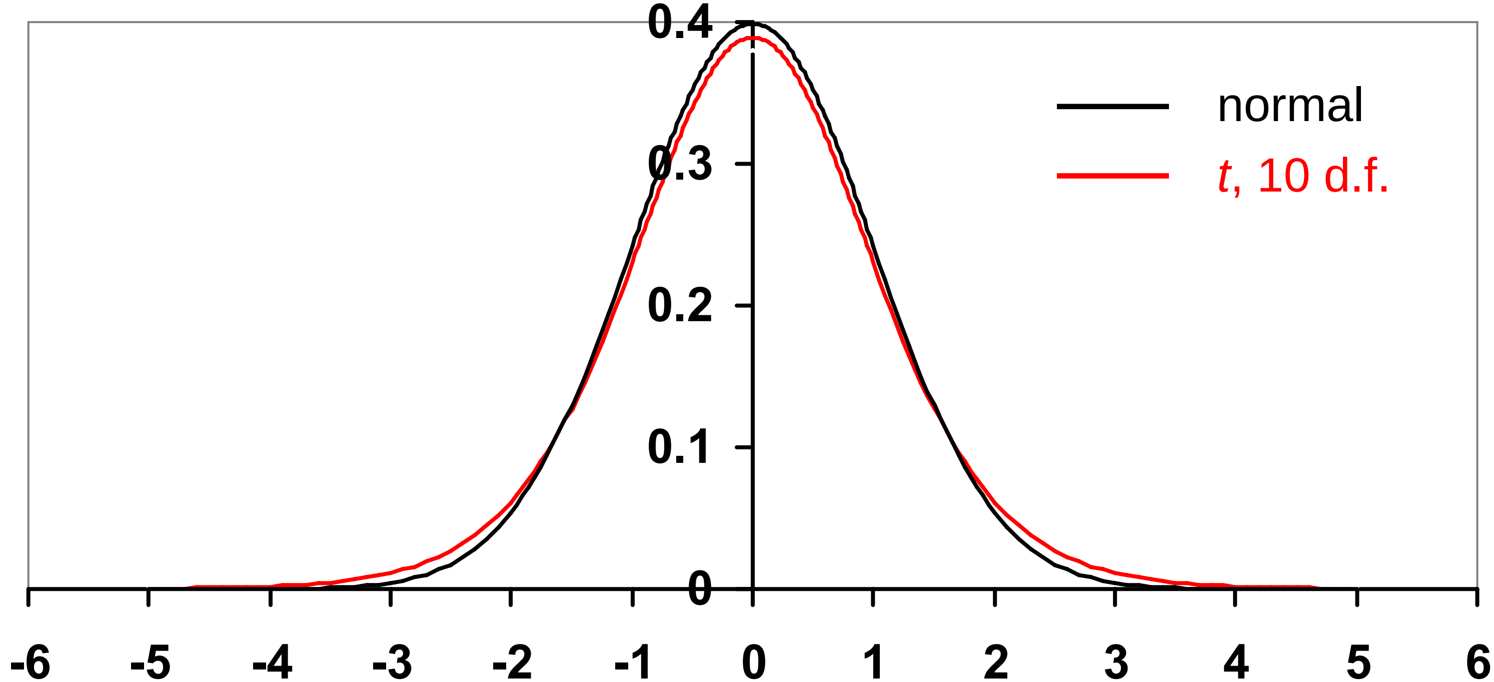
Here is a graph of a normal distribution with zero mean and unit variance

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



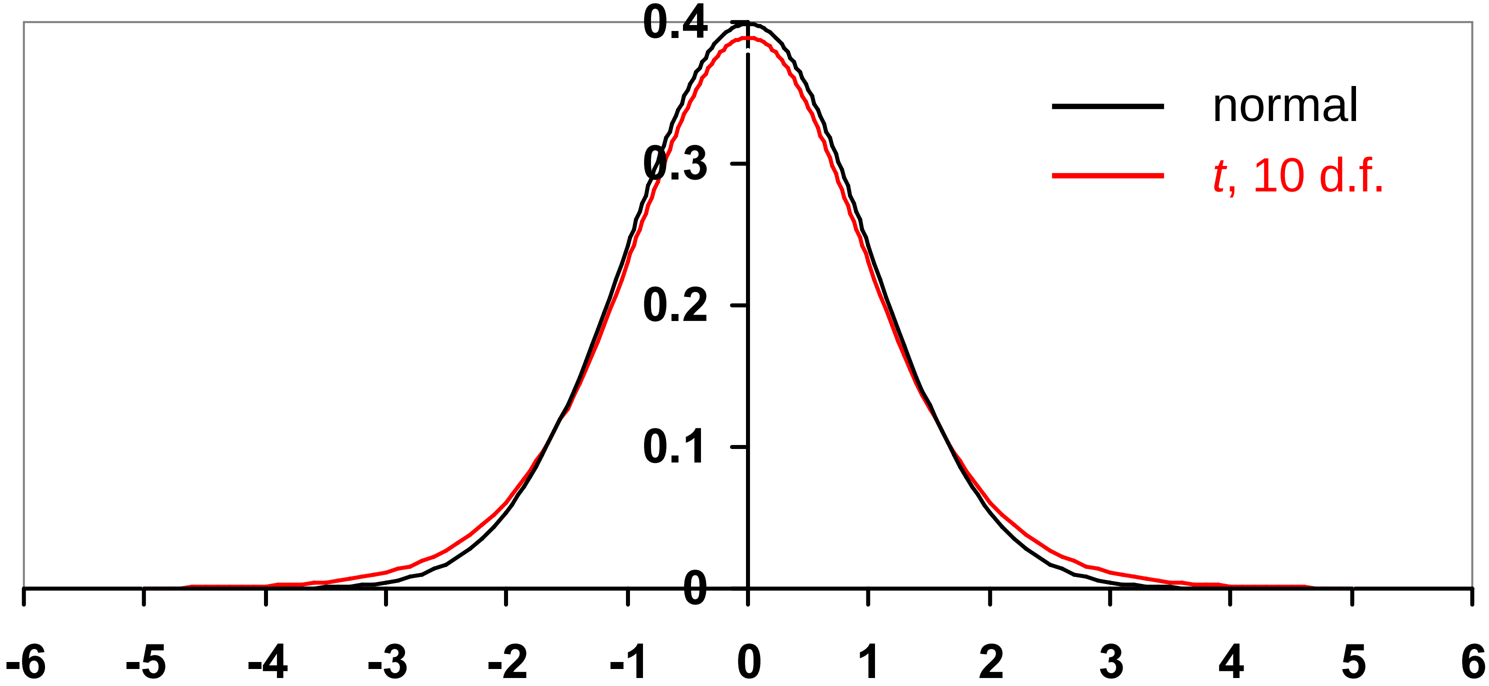
A graph of a *t* distribution with 10 degrees of freedom (this term will be defined in a moment) has been added.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



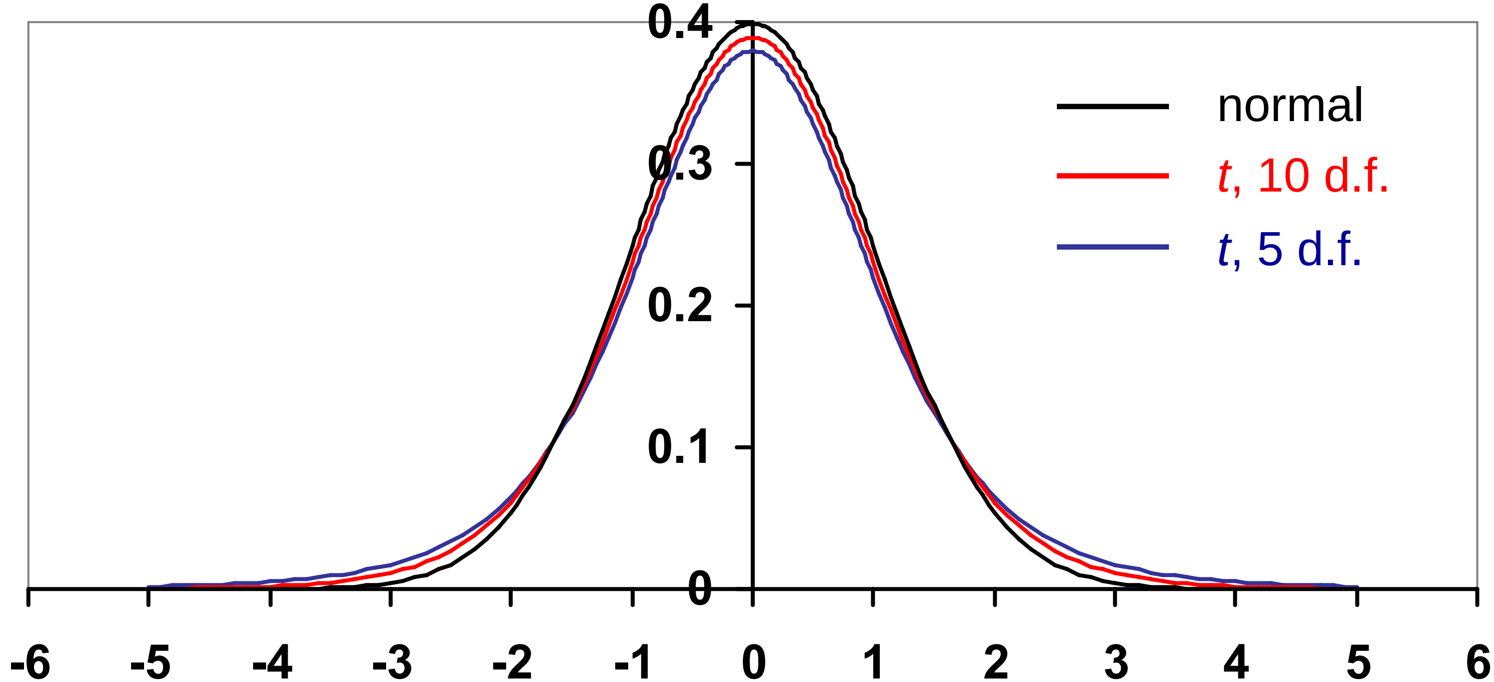
When the number of degrees of freedom is large, the *t* distribution looks very much like a normal distribution (and as the number increases, it converges on one).

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



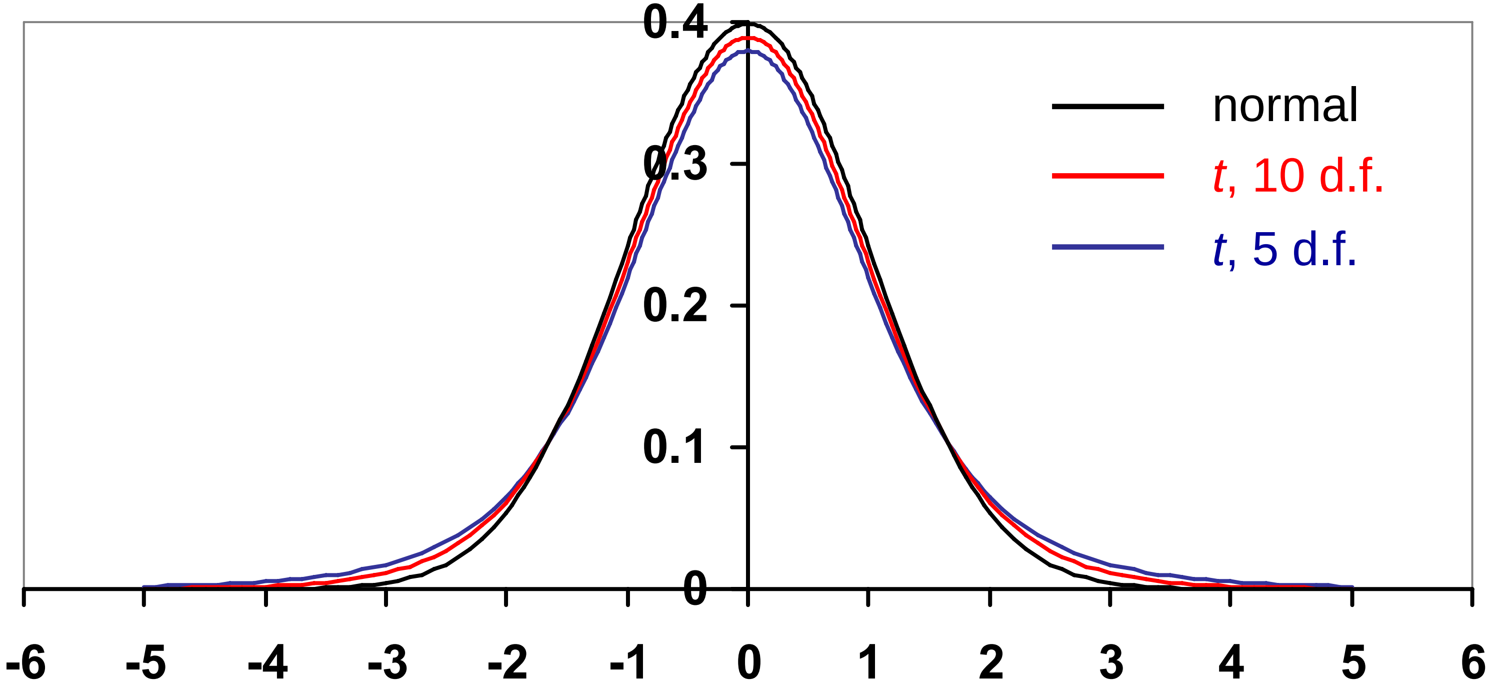
Even when the number of degrees of freedom is small, as in this case, the distributions are very similar.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

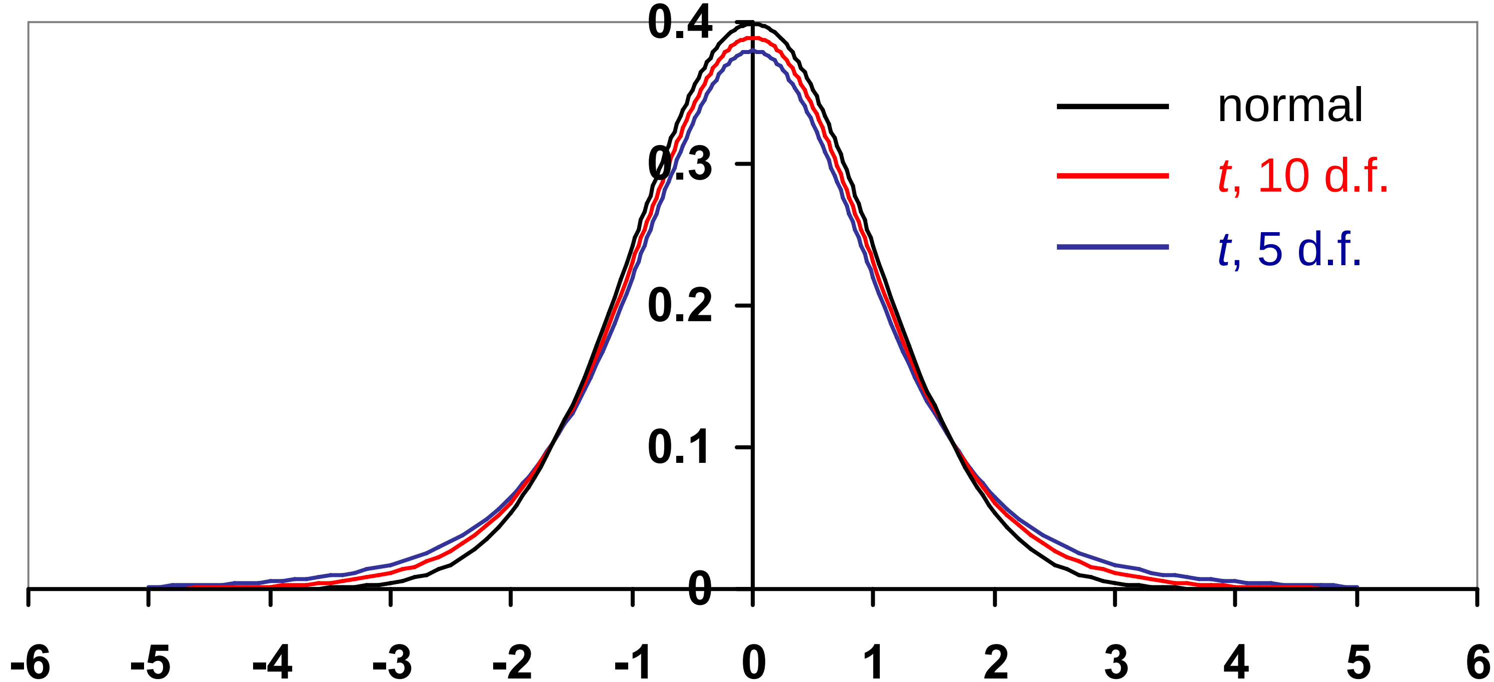


Here is another *t* distribution, this time with only 5 degrees of freedom. It is still very similar to a normal distribution.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

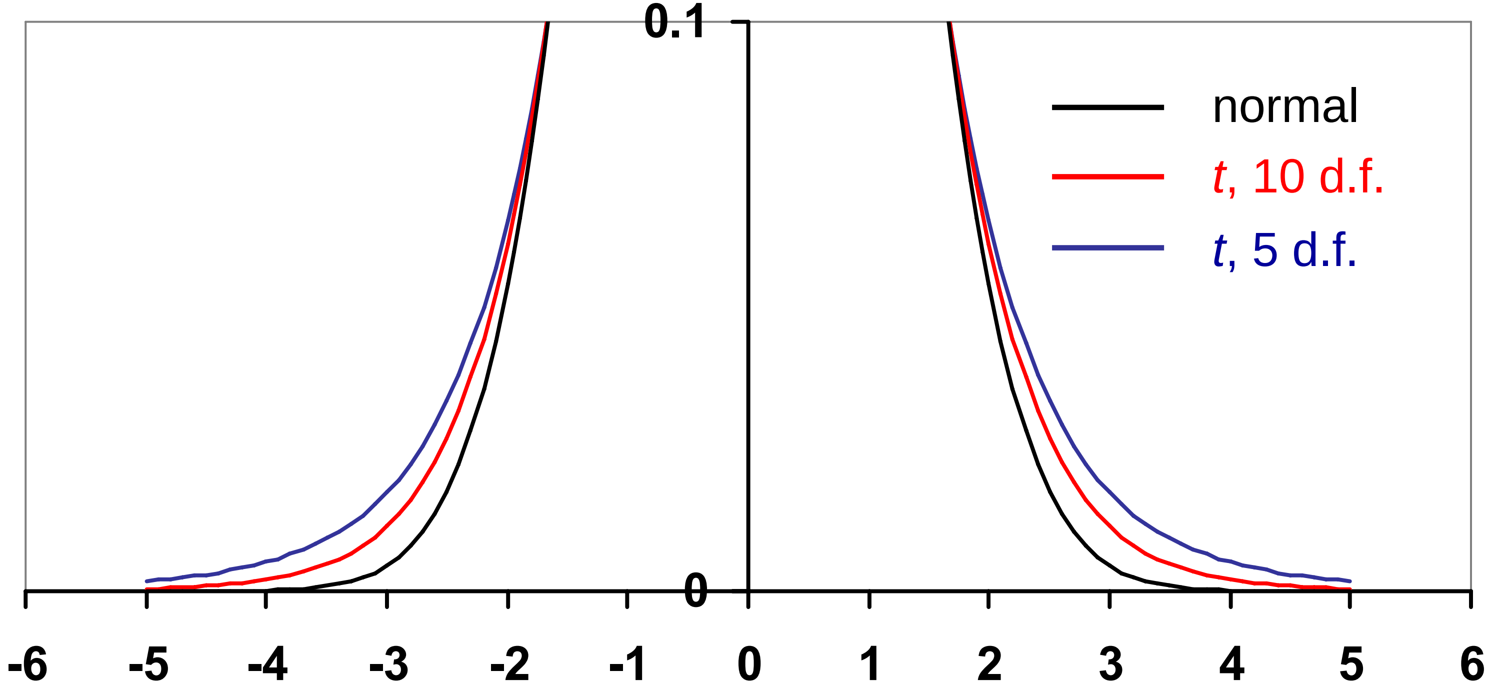


So why do we make such a fuss about referring to the *t* distribution rather than the normal distribution? Would it really matter if we always used 1.96 for the 5% test and 2.58 for the 1% test?



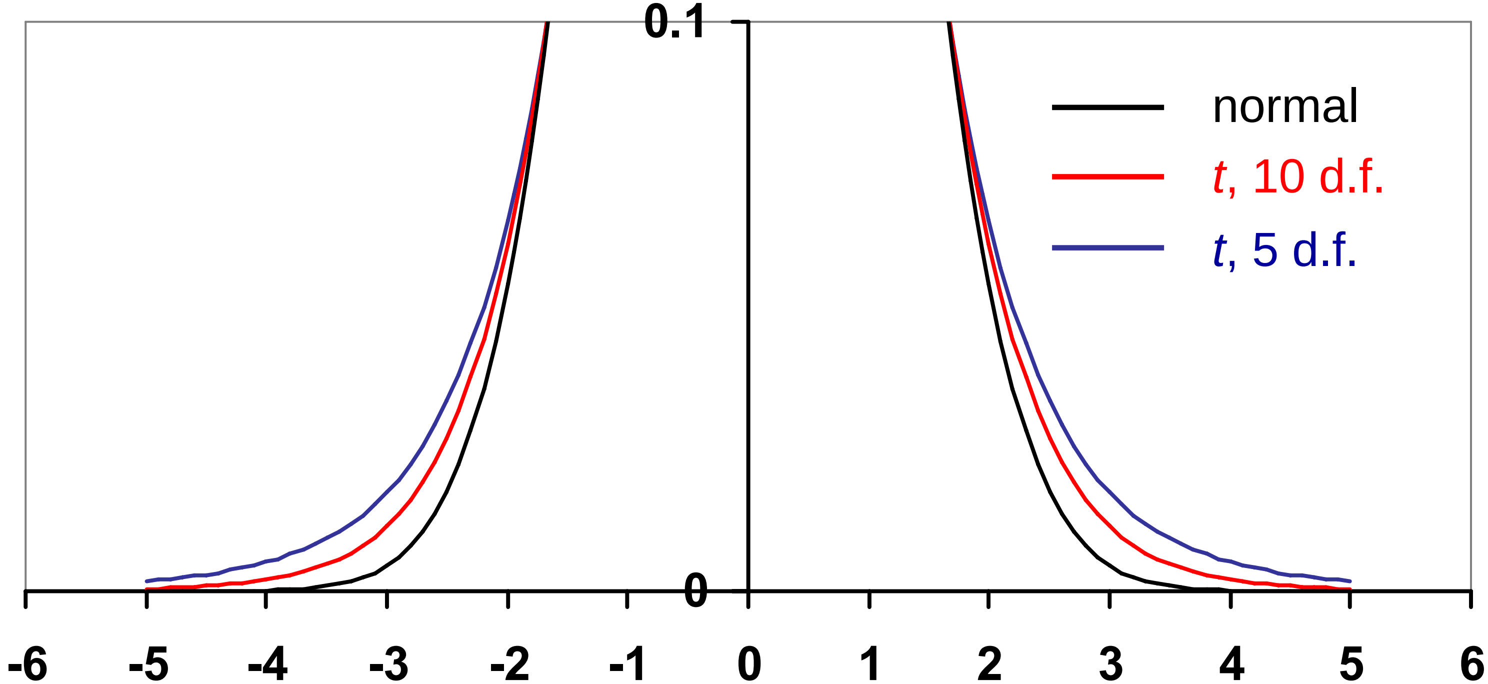
The answer is that it does make a difference. Although the distributions are generally quite similar, the *t* distribution has longer tails than the normal distribution, the difference being the greater, the smaller the number of degrees of freedom.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



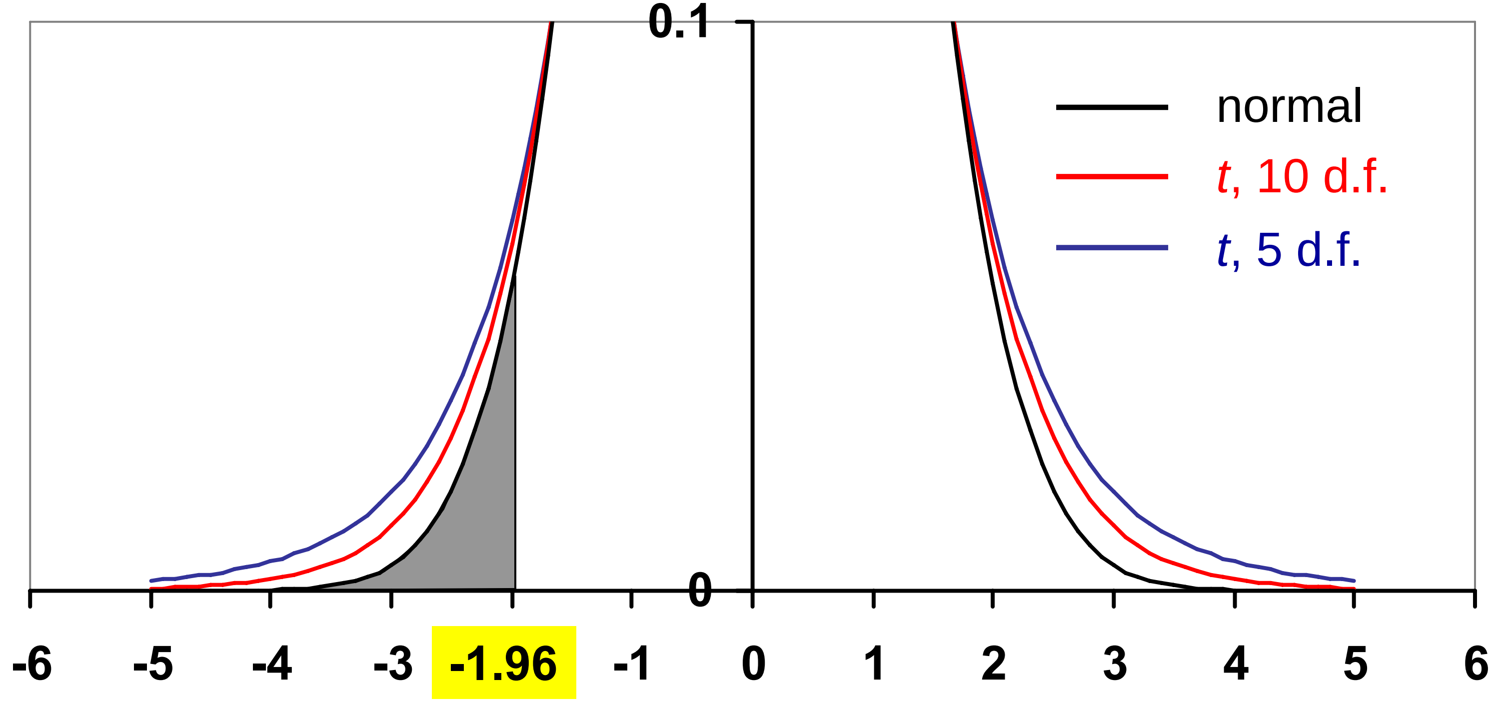
As a consequence, the probability of obtaining a high test statistic on a pure chance basis is greater with a t distribution than with a normal distribution.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



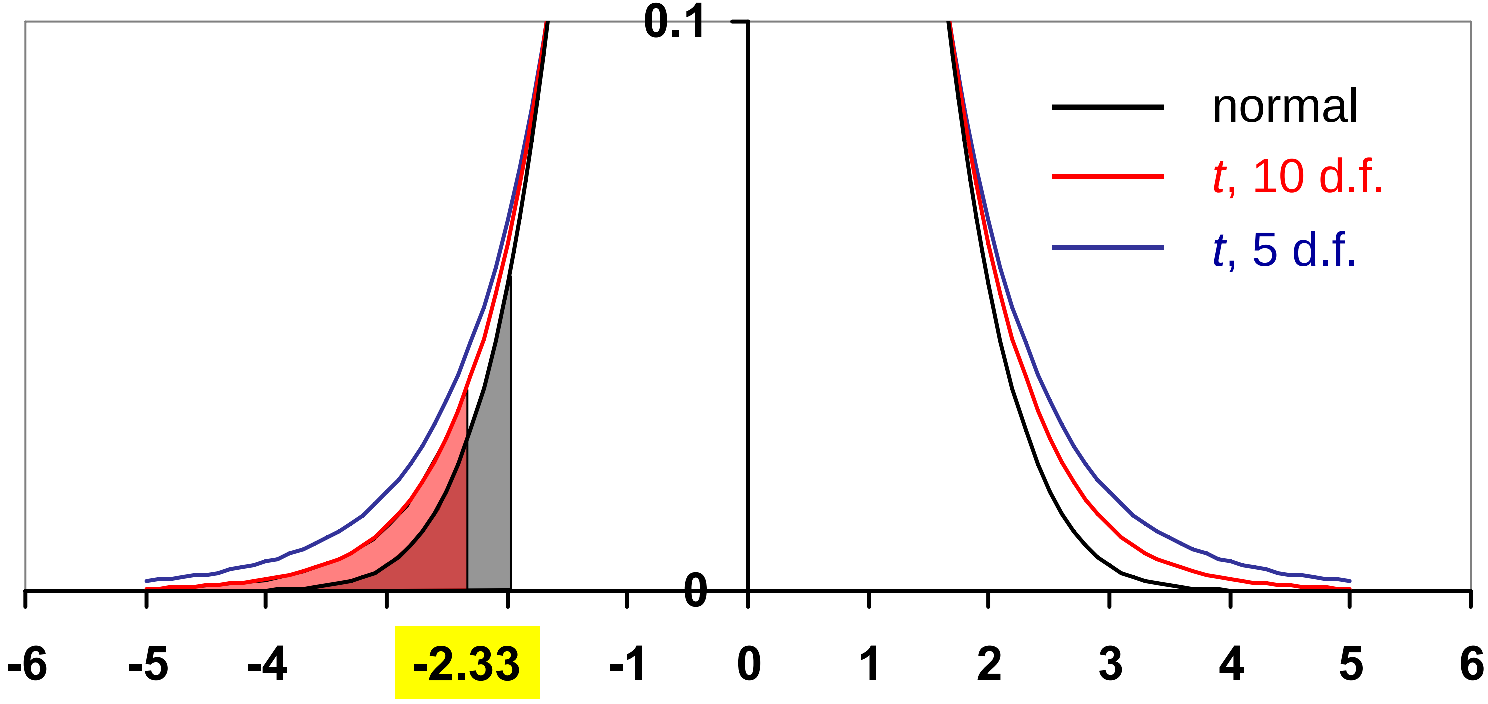
This means that the rejection regions have to start more standard deviations away from zero for a *t* distribution than for a normal distribution.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



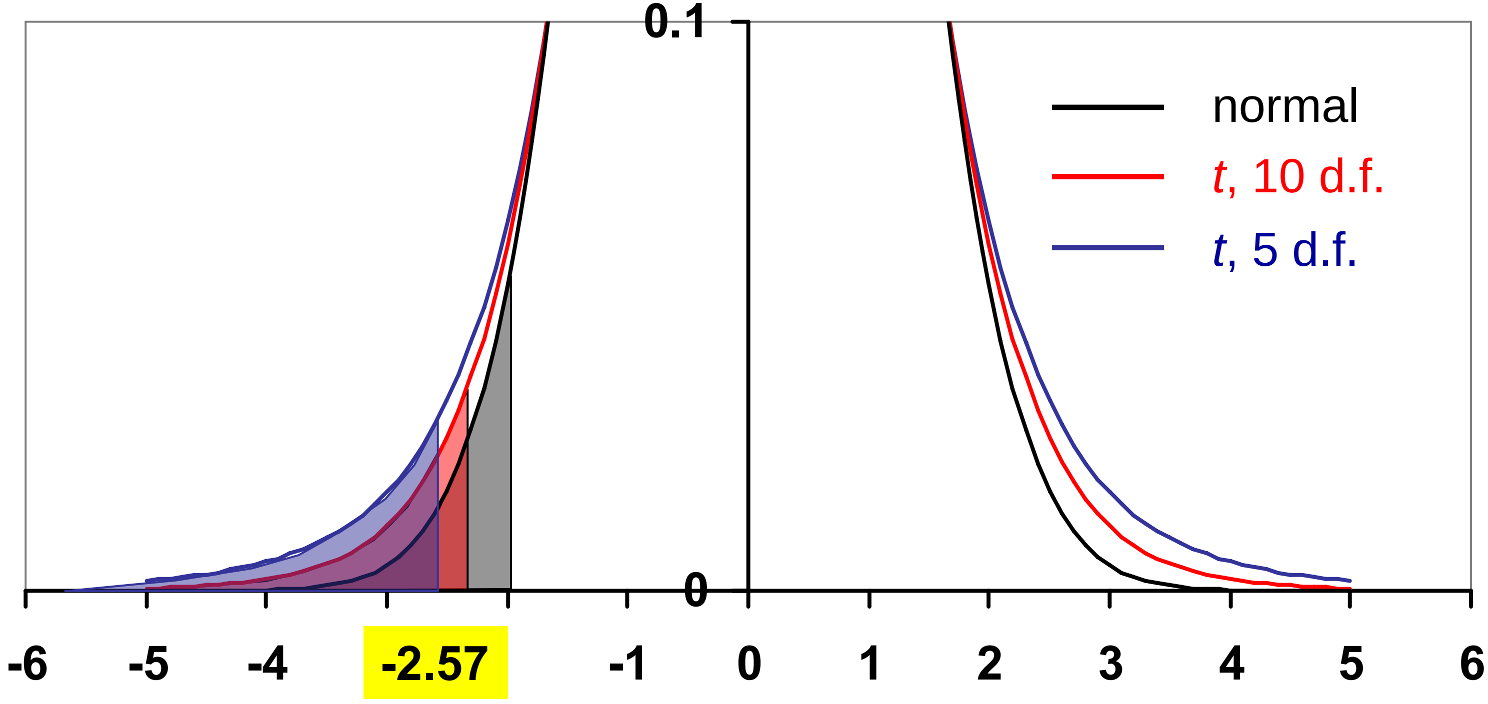
The 2.5% tail of a normal distribution starts 1.96 standard deviations from its mean.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



The 2.5% tail of a *t* distribution with 10 degrees of freedom starts 2.33 standard deviations from its mean.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT



That for a *t* distribution with 5 degrees of freedom starts 2.57 standard deviations from its mean.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

For this reason we need to refer to a table of critical values of *t* when performing significance tests on the coefficients of a regression equation.

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
...		...	...	...	...	...	...
...		...	...	...	...	...	...
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
...		...	...	...	...	...	...
...		...	...	...	...	...	...
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

At the top of the table are listed possible significance levels for a test. For the time being we will be performing two-tailed tests, so ignore the line for one-tailed tests.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

Hence if we are performing a (two-tailed) 5% significance test, we should use the column thus indicated in the table.

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
…							
…							
18		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

Number of degrees of freedom in a regression

= number of observations - number of parameters estimated.

The left hand vertical column lists degrees of freedom. The number of degrees of freedom in a regression is defined to be the number of observations minus the number of parameters estimated.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

In a simple regression, we estimate just two parameters, the constant and the slope coefficient, so the number of degrees of freedom is n - 2 if there are n observations.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

If we were performing a regression with 20 observations, as in the price inflation/wage inflation example, the number of degrees of freedom would be 18 and the critical value of *t* for a 5% test would be 2.101.

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

Note that as the number of degrees of freedom becomes large, the critical value converges on 1.96, the critical value for the normal distribution. This is because the *t* distribution converges on the normal distribution.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

s.d. of b_2 known

discrepancy between hypothetical value and sample estimate, in terms of s.d.:

$$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$$

5% significance test:

reject H_0: $\beta_2 = \beta_2^0$ if

$z > 1.96$ or $z < -1.96$

s.d. of b_2 not known

discrepancy between hypothetical value and sample estimate, in terms of s.e.:

$$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$$

5% significance test:

reject H_0: $\beta_2 = \beta_2^0$ if

$t > t_{crit}$ or $t < -t_{crit}$

Hence, referring back to the summary of the test procedure,

s.d. of b_2 known	s.d. of b_2 not known
discrepancy between hypothetical value and sample estimate, in terms of s.d.:	discrepancy between hypothetical value and sample estimate, in terms of s.e.:
$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$
5% significance test:	5% significance test:
reject H_0: $\beta_2 = \beta_2^0$ if	reject H_0: $\beta_2 = \beta_2^0$ if
$z > 1.96$ or $z < -1.96$	$t > 2.101$ or $t < -2.101$

we should reject the null hypothesis if the absolute value of *t* is greater than 2.101.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

t Distribution: Critical values of *t*

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
…		…	…	…	…	…	…
…		…	…	…	…	…	…
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
…		…	…	…	…	…	…
…		…	…	…	…	…	…
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

If instead we wished to perform a 1% significance test, we would use the column indicated above. Note that as the number of degrees of freedom becomes large, the critical value converges to 2.58, the critical value for the normal distribution.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

t Distribution: Critical values of t

Degrees of freedom	Two-tailed test / One-tailed test	10% / 5%	5% / 2.5%	2% / 1%	1% / 0.5%	0.2% / 0.1%	0.1% / 0.05%
1		6.314	12.706	31.821	63.657	318.31	636.62
2		2.920	4.303	6.965	9.925	22.327	31.598
3		2.353	3.182	4.541	5.841	10.214	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
...		...	...	...	...	...	...
...		...	...	...	...	...	...
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
...		...	...	...	...	...	...
...		...	...	...	...	...	...
120		1.658	1.980	2.358	2.617	3.160	3.373
∞		1.645	1.960	2.326	2.576	3.090	3.291

For a simple regression with 20 observations, the critical value of *t* at the 1% level is 2.878.

t TEST OF A HYPOTHESIS RELATING TO A REGRESSION COEFFICIENT

s.d. of b_2 known	s.d. of b_2 not known
discrepancy between hypothetical value and sample estimate, in terms of s.d.:	discrepancy between hypothetical value and sample estimate, in terms of s.e.:
$z = \frac{b_2 - \beta_2^0}{\text{s.d.}}$	$t = \frac{b_2 - \beta_2^0}{\text{s.e.}}$
5% significance test:	1% significance test:
reject H_0: $\beta_2 = \beta_2^0$ if	reject H_0: $\beta_2 = \beta_2^0$ if
$z > 1.96$ or $z < -1.96$	$t > 2.878$ or $t < -2.878$

So we should this figure in the test procedure for a 1% test.