APPLIED ECONOMICS

Introduction: Linear Regression

Material from Stock and Watson (ch.4-6), Wooldridge (ch.3-5)

Outline

Introduction to the Simple Linear Regression Model 1/2

- Much of applied econometric analysis begins with the following premise: Y and X are two variables, and we are interested in "explaining Y in terms of X", or in "studying how Y varies with changes in X".

- In order to do that, we must consider three issues:
  1. there is never an exact relationship between two variables
  2. we need to specify the functional relationship between Y and X
  3. how can we be sure we are capturing a ceteris paribus relationship between Y and X (if that is a desired goal)?

Introduction to the Simple Regression Model 2/2

- In the Simple Linear Regression Model the relationship between the variable we want to explain (Y) and the explanatory variable (X) is expressed as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where ε represents the error term (unobservable) and β_0 and β_1 the population parameters.

- ε represents all the factors other than X that affect Y. In the simple regression model we assume that all these other factors are unobserved.

Example of Simple Regression Model

Example

$$wage = \beta_0 + \beta_1 educ + \varepsilon$$

- The error term (ε) can include different variables like: experience, capacity or ability, other personal characteristics like age, sex, ...
- Under certain assumptions, we will capture a ceteris paribus relationship between *wage* and *educ*

Assumptions of the Simple Regression Model

1. Linear in parameters.

   Not very restrictive, we are able to model non linear relationships in
   variables.

2. Zero Conditional Mean: $E(\varepsilon|X) = 0$
   This assumption implies:
   1. $E(\varepsilon) = 0$ and
   2. $C(h(X), \varepsilon) = 0$, where $h(X)$ is any function of X. In particular implies
      $C(X, \varepsilon) = 0$

3. Observations are drawn from random distributions.

   Too strong in some applications. For example, if we want to study

   female wages and we observe the salaries of working women.

Example

$$wage = \beta_0 + \beta_1 educ + \varepsilon$$

- In order to simplify the discussion we can assume that ε is innate ability.
- $E(ability|educ = 0)$ denotes the average ability for the group of people with no education, and $E(ability|educ = 12)$ denotes the average ability among people in the population with 12 years of education.
- Then the zero conditional mean assumption implies that $E(ability|educ = 0) = E(ability|educ = 12)$. In fact, the average level of ability is the same for all levels of education.
- Since we do not observe innate ability, there is no way of knowing whether or not average ability is the same for all levels of education.

Potential Violations of Conditional Mean Independence

Model proposed is wrongly specified

- Example: *wages* do not depend linearly on *educ*

Explanatory variables are potentially correlated with missing variables

- Example: We do not control for ability and ability correlated with *educ*

Measurement error in explanatory variables

- Example: a proxy to ability (*IQ*) is included, but measured with error

Explanatory variables simultaneously determined with *Y*

- Example: The level of education chosen depends on the expected returns to education

Interpretation of the coefficients in the Simple Regression Model

- Given the two first assumptions (Linearity and Zero Conditional Mean), it is true that the population regression function (PRF) is a linear function of X:

$$E(Y|X) = \beta_0 + \beta_1 X$$

- Therefore the PRF is exactly the linear projection of Y given X and the slope β_1 has a **causal interpretation**.

Interpretation of the coefficients in the Simple Regression Model

- The slope parameter:

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta X}$$

when X changes in one unit, Y changes on average β_1 units, holding the other factors in ε fixed.

- The constant term β_0 (intercept parameter) is interpreted as the average value for Y when $X = 0$: $E(Y|X = 0)$. In practice, this term does not always have a useful interpretation. Nevertheless we always include a constant in our models.

- Note that if $E(\varepsilon|X) \neq 0$ the parameters do not have a causal interpretation.

Example

$$\hat{wage} = -0.90 + 0.54educ$$

where the hourly wage is measured in euros and education in years.

- The slope measures the change in hourly wage given another year of education, holding all other factors fixed. In this case, one more year of education (for any level), increases the hourly wage in 54 cents.

- With respect to the constant term, our estimation says that a worker with no education would earn an average hourly wage of -90 cents. This nonsense is explained by the fact that in our sample there is no worker with no education, the minimum value for *educ* is 8 years.

Estimation: Analogy Principle

- Our goal is to estimate the population parameters of the model. To do that we use a sample of the population. For that random sample we can write:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

for each i, where $E(\varepsilon_i|X_i) = 0$ and $V(\varepsilon_i|X_i) = \sigma^2$, $\forall i = 1, ..., n$

- One way to obtain estimators for the parameters is using the analogy principle. The idea is to use the corresponding sample quantity as an estimate of the population quantity.

Estimation: Analogy Principle

- Under the assumptions of the simple regression model if we minimize the variance of ε:

$$E(\varepsilon^2) = E[(Y - \beta_0 - \beta_1 X)^2]$$

we get the formulas for the population parameters:

$$\beta_0 = E(Y) - \beta_1 E(X)$$
$$\beta_1 = \frac{C(Y, X)}{V(X)}$$

Estimation: Analogy Principle

- We have then the population parameters in terms of population moments (Expected values, Covariances, Variances).

- Applying the analogy principle we obtain estimators for β_0 and β_1.

$$\hat{\beta_0} = \bar{Y} - \hat{\beta_1}\bar{X}$$
$$\hat{\beta_1} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$$

Estimation: OLS Criteria

- The same estimators can be obtained minimizing the variance of the sample analog of the errors: the residuals.

- The residuals are the difference between the observed value and the predicted value:
$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i \right).$$

- If we minimize $\sum \hat{\varepsilon}_i^2$, we get the same formulas as using the analogy principle. These estimators then can be interpreted as those estimators that minimize the sum of the squared residuals: the Ordinary Least Squares (OLS) estimators.

Estimation: OLS Criteria

- The first order conditions are:

$$\sum \hat{\varepsilon}_i = 0$$
$$\sum \hat{\varepsilon}_i X_i = 0$$

- Note that these conditions are the sample analog of

$$E(\varepsilon) = 0$$
$$C(X, \varepsilon) = 0$$

- Once we have determined the OLS intercept and slope estimates, we form the OLS regression line: $\hat{Y} = \left(\hat{\beta}_0 + \hat{\beta}_1 X\right)$

Multiple Linear Regression Model

- Clearly most economic relations involves more than two variables, therefore we are going to include as many variables as necessary:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_K X_K + \varepsilon$$

- Multiple regression models are more adequate for ceteris paribus analysis because they allow us to explicitly control for many other factors which simultaneously affect the dependent variable.

Assumptions 1/2

1. Linear in parameters.

2. Zero Conditional Mean: $E(\varepsilon|X_1, X_2, ..., X_K) = 0$
   This assumption implies:
   1. $E(\varepsilon) = 0$ and
   2. $C(h(X_j), \varepsilon) = 0$, where $h(X_j)$ is any function of X_j. In particular, $C(X_j, \varepsilon) = 0$

Assumptions 2/2

3. Observations are drawn from random distributions.

4. We need an additional assumption: No Perfect Multicollinearity:
   $(X_1, X_2, ..., X_K)$ do not produce a perfect linear combination.
   Technical: Invertibility of Variance-covariance Matrix

Interpretation

- Now the coefficients have partial effect, or ceteris paribus, interpretations:

$$\beta_j = \frac{\Delta E(Y|X)}{\Delta X_j}$$

where X represents the vector of explanatory variables: $X_1, X_2, ..., X_K$

- The coefficient on X_j measures the change in Y due to a one-unit increase in X_j, holding all other independent variables fixed.

Example

$$w\hat{a}ge = -2.87 + 0.60\,educ + 0.02\,exper + 0.17\,tenure$$

where the hourly wage is measured in euros and education, exper (labor market experience), and tenure (years with the current employer) in years.

- Now the coefficients have a ceteris paribus interpretation.

- The coefficient on *educ* measures the change in hourly wage given another year of education, holding experience and tenure fixed. In this case, one more year of education (for any level), increases the hourly wage in 60 cents.

- Alternatively, if we take two people with the same levels of experience and job tenure, the coefficient on *educ* is the proportionate difference in predicted wage when their education levels differ by one year.

A "partialling-out" Interpretation: Multiple Regression

True Model

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + u$

Two-step procedure

1. OLS regress X_1 on constant and $X_2, ..., X_k$, and get residuals $\hat{u}_1$
2. OLS regress Y on $\hat{u}_1$

- $\hat{u}_1$ captures sample changes in X_1 which are uncorrelated with the rest of the regressors (alternatively, $\hat{u}_1$ is X_1 once the effects of the rest of the regressors have been partialled out).
- the estimated slope in the second regression is $\hat{\beta}_1$
- $\hat{\beta}_1$ can be expressed as: $\frac{\left(\sum \hat{u_{1i}} Y_i\right)}{\left(\sum \hat{u_{1i}}^2\right)}$
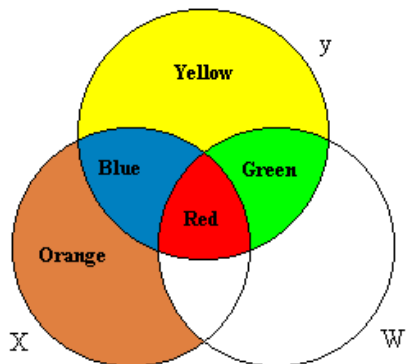
A "partialling-out" Interpretation - Example

True Model

- $wages = \alpha_0 + \alpha_1\,educ + \alpha_2\,IQ + u$

Two-step procedure

1. OLS regress *educ* on constant and *IQ*, and get residuals $\hat{u}_1$
2. OLS regress *wages* on $\hat{u}_1$

- $\hat{u}_1$ captures sample changes in *educ* uncorrelated with changes in *IQ*
- the estimated slope in the second regression is $\hat{\alpha}_1$

Venn Diagrams for Multiple Regression Model[1]

- Each circle represents sample variation in the corresponding variable.
- In a simple model (y on X), OLS uses Blue + Red to estimate β_X
- When y is regressed on X and W: OLS throws away the red area and just uses blue to estimate β_X
- Idea: red area is contaminated (we do not know if the movements in y are due to X or to W).
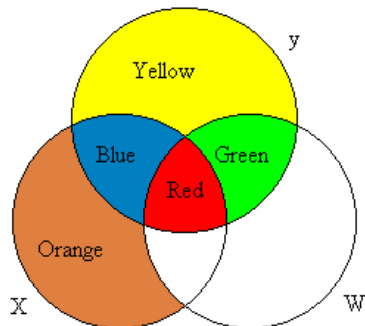- What is the effect of omitting W then?

Multicollinearity

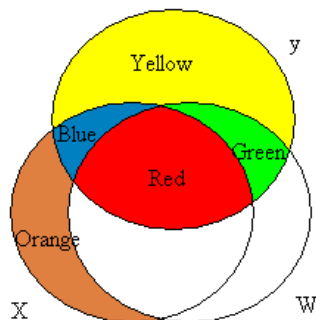

Figure 3a Modest collinearity   Figure 3b Considerable collinearity

In the case of considerable collinearity, OLS are still consistent estimators. However, since we use less information (compare the blue and green areas in both figures), the estimation is less precise.

Goodness of Fit

The Variance Decomposition

- Each observation of Y_i is made up of two components: $Y_i = \hat{Y}_i + \hat{u}_i$
  - the fitted values: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + ... + \hat{\beta}_k X_{ki}$
  - the OLS residual: $\hat{u}_i = Y_i - \hat{Y}_i$

- We can define the total sum of squares (SST), the explained sum of squares (SSE), and the residual sum of squares or sum of squared residuals (SSR), as

  $SST = \sum(Y_i - \bar{Y})^2$

  $SSE = \sum(\hat{Y}_i - \bar{Y})^2$

  $SSR = \sum(\hat{u}_i)^2$

The R^2

$$SST = SSE + SSR$$

- Solving for SSE and dividing by SST:

$$\frac{SSE}{SST} = 1 - \frac{SSR}{SST} \equiv R^2$$

- the R^2 is the proportion of the sample variation in Y_i explained by the regression line: $0 \leq R^2 \leq 1$

- it never decreases when a new regressor is added

Linear in the exogenous variables

$$E(Y|X) = \beta_0 + \beta_1 X,$$

- As we saw before, in this case

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta X}$$

- If X varies in 1 unit, Y varies on average β_1 units of Y .
- What's the elasticity in this case?

$$\eta = \frac{\Delta E(Y|X)/E(Y|X)}{\Delta X/X} = \beta_1 \frac{X}{E(Y|X)}$$

- Note that the elasticity depends on the specific values of X and Y, and therefore it is not constant.

Semilogarithmic models: log in the exogenous variable

$$E(Y|X) = \beta_0 + \beta_1 lnX,$$

$$\beta_1 = \frac{\Delta E(Y|X)}{\Delta lnX} \approx \frac{\Delta E(Y|X)}{\Delta X/X}$$

- To express the impact of the variation in X in percentage terms we divide and multiply by 100:

$$\beta_1/100 \approx \frac{\Delta E(Y|X)}{100\Delta X/X}$$

- If X changes 1%, Y changes on average $\beta_1/100$ units (β_1 is a semi-elasticity).

Semilogarithmic models: log in the dependent variable

$$E(lnY|X) = \beta_0 + \beta_1 X,$$

$$\beta_1 = \frac{\Delta E(lnY|X)}{\Delta X} \approx \frac{E(\Delta Y/Y|X)}{\Delta X}$$

- To express the variation in Y in percentage terms we multiply by 100:

$$100\beta_1 \approx \frac{E((100 \times \Delta Y/Y)|X)}{\Delta X}$$

- If X changes in one unit, Y changes on average $100\beta_1\%$.

Double log model

- This model is used for example when we want to model a constant elasticity between X and Y.

$$E(lnY|X) = \beta_0 + \beta_1 lnX,$$

$$\beta_1 = \frac{\Delta E(lnY|X)}{\Delta lnX} \approx \frac{E((\Delta Y/Y)|X)}{\Delta X/X}$$

- If X changes 1%, Y changes on average β_1 %: β_1 represents the elasticity of Y with respect to X.

Model with quadratic terms

- This model is used for example when modeling increasing or decreasing marginal returns.

$$E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2,$$

Then,

$$\frac{\Delta E(Y|X)}{\Delta X} = \beta_1 + 2\beta_2 X$$

- If X changes one unit, Y changes on average $\beta_1 + 2\beta_2 X$. The marginal effect of X on Y depends on the sign of β_2.

Example

The file gasoline.gdt contains information about the US gasoline market for the years 1960-1995. The variable *pcgas* is the per capita gasoline consumption and *inc* is per capita disposable income measured in dollars. In each point, estimate the model, give an interpretation for β_1 and compute the elasticity of *pcgas* with respect to *inc*.

1. $pcgas = \beta_0 + \beta_1 inc + \varepsilon$

2. $pcgas = \beta_0 + \beta_1 log(inc) + \varepsilon$

3. $log(pcgas) = \beta_0 + \beta_1 log(inc) + \varepsilon$

4. $log(pcgas) = \beta_0 + \beta_1 log(inc) + \beta_2 log(P_g) + \varepsilon$, where P_g is the price of gasoline. Compute the price elasticity

Properties of OLS estimators

- Unbiasedness (under linearity and zero conditional mean assumptions)
  - $E\left(\widehat{\beta}_0\right) = \beta_0$
  - $E\left(\widehat{\beta}_1\right) = \beta_1$
  - Remember that unbiasedness is a feature of the sampling distributions of the estimators, which says nothing about the estimate that we obtain for a given sample. We hope that, if the sample we obtain is somehow "typical", then our estimate should be "near" the population value.

The Notion of Large Sample Properties

- we look at properties as the sample size n gets larger and larger:

  1. how far is $\hat{\beta}_j$ from the true parameter β_j as $n \to \infty$?

  2. how does the distribution of $\hat{\beta}_j$ look as $n \to \infty$?

A Law of Large Numbers (1/2)

- it essentially states that the average of the results obtained from a large number of trials will tend to become closer to the expected value as more trials are performed

- the LLN is important because it "guarantees" stable long-term results for random events

- it does not say that a streak of one value will immediately be "balanced" by the others (the Gambler's fallacy)

A Law of Large Numbers (2/2)

Theorem

For any random variable y with an expected value μ define the average of a sample of size n as $\overline{y}_n$. Then

$$plim(\overline{y}_n) = \mu$$

$$plim(\hat{cov}(y,x)) = cov(y,x)$$

$$plim(\hat{var}(x)) = var(x)$$

plim Properties

Continuous Mapping Theorem

For every continuous function? $g(\cdot)$ and random variable x:

$$plim(g(x)) = g(plim(x))$$

Example 1

$plim(x + y) = plim(x) + plim(y)$

Example 2

$plim\left(\frac{x}{y}\right) = \frac{plim(x)}{plim(y)}$ if $plim(y) \neq 0$

Convergence in Probability

Definition

As the sample size grows, any positive distance between $\hat{\beta}_j$ and β_j becomes arbitrarily unlikely

- β_j is probability limit of $\hat{\beta}_j$
- $\hat{\beta}_j$ converges in probability to β_j
- $plim(\hat{\beta}_j) = \beta_j$
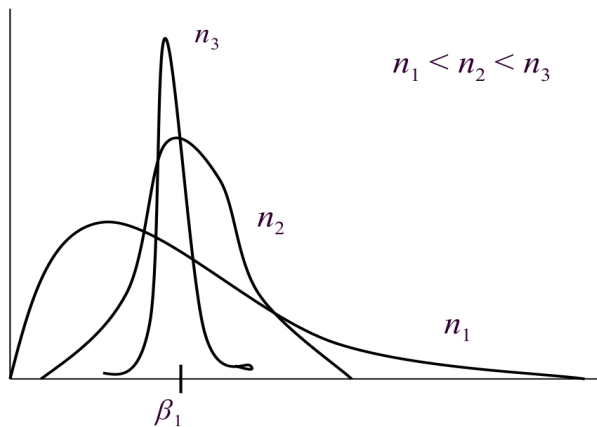
OLS Consistency

Theorem

Under linear regression model assumptions, OLS is consistent

Example: $wages = \beta_0 + \beta_1 educ + u$ with $cov(educ, u) = 0$

- $\hat{\beta}_1 = \beta_1 + \frac{c\hat{o}v(educ_i, u_i)}{v\hat{a}r(educ_i)}$

- $plim\left(\hat{\beta}_1\right) = plim(\beta_1) + \frac{plim(c\hat{o}v(educ_i, u_i))}{plim(v\hat{a}r(educ_i))} = \beta_1 + \frac{cov(educ, u)}{var(educ)}$

- Since $cov(educ, u) = 0 \Rightarrow plim\left(\hat{\beta}_1\right) = \beta_1$

A Graphical Interpretation of Consistency

An Example of Inconsistency

True Model: $wages = \beta_0 + \beta_1 educ + \beta_2 IQ + v$

- $cov(educ, v) = cov(IQ, v) = 0$
- $cov(educ, IQ) \neq 0, \beta_2 \neq 0$

- Estimated equation by OLS: $wages = \hat{\gamma}_0 + \hat{\gamma}_1 educ + \hat{u}_{educ}$

$\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2 \frac{\hat{cov}(educ, IQ)}{\hat{var}(educ)} \Rightarrow plim(\hat{\gamma}_1) = \beta_1 + \beta_2 \frac{cov(educ, IQ)}{var(educ)}$

- $plim(\hat{\gamma}_1) \neq \beta_1$ if
  - intelligence is relevant: $\beta_2 \neq 0$
  - education is correlated to intelligence: $cov(educ, IQ) \neq 0$

The Central Limit Theorem

Theorem

For any random variable y with an expected value μ and variance σ^2 define the average of a sample of size n as $\overline{y}_n$. Then

$$n^{\frac{1}{2}} \frac{\overline{y}_n - \mu}{\sigma} \to N(0,1) \text{ as } n \to \infty$$

- what is remarkable about the CLT is that the distribution of the random variable y is irrelevant

- the crucial issue is the averaging carried out over y

Asymptotic Normality for OLS Estimators

Under the Linear Regression Model Assumptions

$$n^{\frac{1}{2}} \frac{\hat{\beta}_j - \beta_j}{a_j} \to N(0,1) \text{ as } n \to \infty \text{ where } a_j{}^2 = Asy.Var\left(n^{\frac{1}{2}}\hat{\beta}_j\right)$$

(Robust) Standard Errors

we can estimate a_j consistently: $se\left(n^{\frac{1}{2}}\hat{\beta}_j\right)$

- As sample size n increases, the OLS estimators—conveniently scaled up—get as close as we want to a normal distribution...

$$n^{\frac{1}{2}}\hat{\beta}_j \approx N(\beta_j, a_j{}^2)$$

- we can approximate the standard deviation of $\hat{\beta}_j$,
  $sd\left(\hat{\beta}_j\right) \approx \frac{a_j}{n^{1/2}} \approx se\left(\hat{\beta}_j\right)$

(Robust) Standard Errors

- these standard errors are "robust-to-heteroskedasticity"

- one important "detail": in most computer packages, the "standard error" reported is not $se\left(\hat{\beta}_j\right)$ but one which is valid under unrealistic assumptions (homoskedasticity)

- Conditional Homoskedasticity : $V(\varepsilon|X) = \sigma^2$

- In the wage equation, we can interpret this assumption as innate ability having the same dispersion for different levels of education. It implies also that the dispersion of wages does not depend on the level of education. For instance, $V(wage|educ = 0) = V(wage|educ = 12)$.

- What if people with very low levels of education have very few opportunities and must work at the minimum wage?

The t Test

- from the CLT (and a LLN), it can also be shown that

$$t = \frac{\hat{\beta}_j - \beta_j}{se\left(\hat{\beta}_j\right)} \to N(0,1) \text{ as } n \to \infty$$

- this result can be used for example, to test whether a coefficient is significant

The Significance Test

Example: $wages = \beta_0 + \beta_1 educ + \beta_2 IQ + u$

- we want to check that ability significantly affects wages ceteris paribus

- we set up the null we expect to reject
  - null hypothesis: $H_0 : \beta_2 = 0$ (IQ does not affect wages ceteris paribus)
  - alternative hypothesis: $H_1 : \beta_2 \neq 0$ (IQ affects wages ceteris paribus)

Testing Single Parameters

- under the null, $t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \approx N(0,1)$

to reject the null we must find sufficient evidence against it

- "sufficient evidence" would be that t is very large in absolute value: we control the probability that we-reject-the-null-but-should-not (the significance level)

- fix the significance level, $\alpha\%$, and define critical value (c_α), such that

$$\Pr(|t| \geq c_\alpha) = \alpha$$

- if $|t| > c_\alpha$, reject the null at $1 - \alpha\%$ of confidence.

Critical Values

If we fix the significance level at 10%, the critical value is 1.65
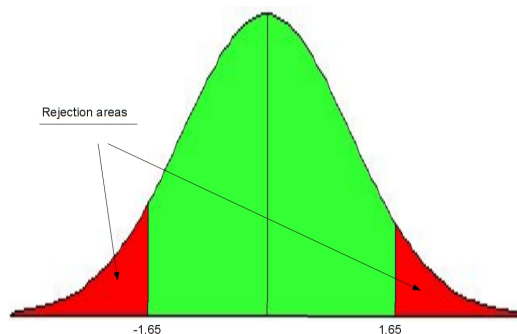
simple rule of thumb for a 5% significance level: if $|t|$ larger than 2, reject the null.

For a one tail test, if we fix the significance level at 10%, the critical value is 1.28. If t is larger than the critical value, we reject the null in favor of the alternative that the parameter is greater than the value under the null.
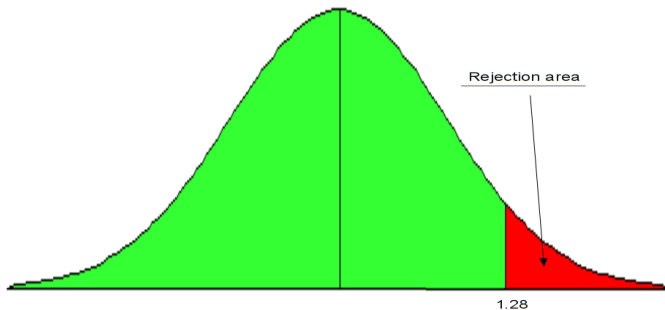
Power of the *t* Test

- we control the probability that we-reject-but-we-should-not by fixing the significance level: this is called the "Type I error"

- in any test, we can incur in a so-called "Type II error": we-do-not-reject-but-we-should

- the probability that we-reject-when-we-should is called "the power of the test" and clearly depends on the alternative

- in practice, if we reject we are ok, if we do not, perhaps the test has no power...

p-values

- an alternative to the classical approach is to ask, "what is the smallest significance level at which the null would be rejected?"

- to answer this question, we first compute the t statistic

- then we look up at the mass probability under the null larger than the actual value we get

- this is the *p*-value

- most computer packages will compute the *p*-value in a two-sided test

Testing a Linear Combination

- Given the following null hypothesis $H_0 : \sum a_j \beta_j = a^0$, we define the following statistic:

$$t = \frac{\sum a_j \hat{\beta}_j - a^0}{s_{\sum a_j \hat{\beta}_j}}$$

- We can show that $t \sim N(0, 1)$. Note that in order to carry out this test, it is necessary to know or estimate the variance of the linear combination. That implies the need of knowing or estimating covariances between estimators.

- Alternatively we can redefine the model in order to obtain directly the statistic we need.

Example: $wages = \beta_0 + \beta_1\, educ + \beta_2\, exper + \beta_3\, IQ + u$

- we want to test if the return to experience is the same as the return to education

- $H_0)\beta_1 = \beta_2$ and $H_0)\beta_1 \neq \beta_2$

- Define a new parameter $\theta = \beta_1 - \beta_2$ and rewrite the model

$$wages = \beta_0 + \theta\, educ + \beta_2(educ + exper) + \beta_3\, IQ + u$$

- now we simply use the t statistic for the significance of $\hat{\theta}$

Testing Several linear restrictions 1/3

- A test that includes several linear restrictions is like this:

$$\beta_j = \beta_k = \beta_l = 0$$

- When we want to test several restrictions, we need to distinguish the unrestricted and restricted models.

- The unrestricted model is the one in which we want to test our hypothesis, the original model. The restricted model is the model once we impose the null hypothesis.

Testing Several linear restrictions 2/3

- For each model we compute the sum of the residuals squared and define the following statistic:

$$W^0 = n\frac{SRR - SRS}{SRS} \overset{\cdot}{\sim} \chi^2_q$$

where SRR represents the sum of the residuals squared in the restricted model, SRS the sum of the residuals squared in the unrestricted model and q the number of restrictions in the null hypothesis.

- An equivalent statistic is:

$$F = n\frac{SRR - SRS}{qSRS}$$

Testing Several linear restrictions 3/3

- W^0 (or F) is never negative, since $SRR \geq SRS$
- Alternatively, when the dependent variable in both models is the same, we can express the statistic in terms of the R^2:

$$W^0 = n \frac{R_S^2 - R_R^2}{1 - R_S^2} \overset{\sim}{\cdot} \chi_q^2$$

where R_R^2 represents the R^2 in the restricted model and R_S^2 the R^2 in the unrestricted model.

Example: $wages = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 IQ + u$

-

$$H_0 : \beta_1 = \beta_2 = 0 \text{ vs } H_1 : H_0 \text{ false}$$

- we cannot reparametrize so that we can get the test we want from the standard output
- we can carry out the W statistic
- Restricted Model: $wages = \beta_0 + \beta_3 IQ + u$, and $q = 2$

Joint Significance

- The most usual case of this tests is the joint significance test. Starting with the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_K X_K + \varepsilon$$

- The null hypothesis, that none of the variables are relevant, includes K linear restrictions:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_K = 0$$

- The restricted model is $Y = \beta_0 + \varepsilon$, and the statistic we need is:

$$W^0 = n\frac{R_S^2}{1 - R_S^2} \overset{\cdot}{\sim} \chi_K^2$$