Testing economic hypotheses. Multiple hypothesis testing. Linear and non-linear hypotheses. Confidence intervals. Delta method.

Jakub Mućk
SGH Warsaw School of Economics

Least squares estimator

- **Least squares estimator** :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \varepsilon \tag{1}$$

where
  - y is the (outcome) dependent variable;
  - $x_1, x_2, \ldots, x_K$ is the set of independent variables;
  - ε is the error term.
- The dependent variable is explained with the components that vary with the **the dependent variable** and **the error term**.
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_K$ are the coefficients (slopes) on $x_1, x_2, \ldots, x_K$.

- **Least squares estimator** :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \varepsilon \quad (1)$$

  where
  - y is the (outcome) dependent variable;
  - $x_1, x_2, \ldots, x_K$ is the set of independent variables;
  - ε is the error term.

- The dependent variable is explained with the components that vary with the **the dependent variable** and **the error term**.
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_K$ are the coefficients (slopes) on $x_1, x_2, \ldots, x_K$.

$\beta_1, \beta_2, \ldots, \beta_K$ measure the effect of change in $x_1, x_2, \ldots, x_K$ upon the expected value of y (*ceteris paribus*).

- **Assumption #1**: true DGP (data generating process):

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon. \tag{2}$$

- **Assumption #2**: the expected value of the error term is zero:

$$\mathbb{E}\left(\varepsilon\right) = 0, \tag{3}$$

and this implies that $\mathbb{E}\left(y\right) = \mathbf{X}\beta$.

- **Assumption #3**: Spherical variance-covariance error matrix.

$$var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon') = I\sigma^2 \tag{4}$$

. In particular:
  - the variance of the error term equals σ:

$$var\left(\varepsilon\right) = \sigma^2 = var\left(y\right). \tag{5}$$

  - the covariance between any pair of ε_i and ε_j is zero"

$$cov\left(\varepsilon_i, \varepsilon_j\right) = 0. \tag{6}$$

- **Assumption #4**: **Exogeneity.** The independent variable are **not random** and therefore they are not correlated with the error term.

$$\mathbb{E}(\mathbf{X}\varepsilon) = 0. \tag{7}$$

- **Assumption #5**: the full rank of matrix of explanatory variables (there is no so-called collinearity):

$$rank(\mathbf{X}) = K + 1 \leq N. \tag{8}$$

- **Assumption #6 (optional)**: the normally distributed error term:

$$\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right). \tag{9}$$

Assumptions of the least squares estimators

Under the assumptions A#1-A#5 of the multiple linear regression model, the least squares estimator $\hat{\beta}^{OLS}$ has the smallest variance of all linear and unbiased estimators of β.

$\hat{\beta}^{OLS}$ is the Best Linear Unbiased Estimators (BLUE) of β.

- **The least squares estimator**

$$\hat{\beta}^{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. \tag{10}$$

- **The variance of the least square estimator**

$$Var(\hat{\beta}^{OLS}) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \tag{11}$$

- **If the (optional) assumption about normal distribution of the error term is satisfied** then

$$\beta \sim \mathcal{N}\left(\hat{\beta}^{OLS}, Var(\hat{\beta}^{OLS})\right). \tag{12}$$

Statistical inference

- **Statistical inference** is the process that using sample data allows to deduce properties of an underlying features of population.
- **Statistical inference** consists of:
  - estimation of underlying parameters,
  - testing hypotheses.

- Hypothesis testing is a comparison of a conjecture we have about a population to the information contained in a sample of data.
- The hypotheses are formed about economic behavior.
- In statistical inference, the hypotheses are then represented as statements about model parameters.
- General procedures:
  1. A null hypothesis $\mathcal{H}_0$,
  2. An alternative hypothesis $\mathcal{H}_1$,
  3. A test statistic
  4. A rejection region
  5. A conclusion

The test statistics and rejection region

SGH

- Based on **the value of a test statistic** we decide either to reject the null hypothesis or not to reject it.
- **The rejection region** consists of values that are unlikely and that have low probability of occurring when the null hypothesis is true.
- **The rejection region** depends on:
  - Distribution of test statistics when the null is true.
  - Alternative hypothesis.
  - Level of significance.

- **Type I error** is a situation, in which we reject the null hypothesis when it is true.
- **Type II error** is a situation, in which we do not reject the null hypothesis when it is false.
- **Significance level α**:

$$P\left(\text{Type I error}\right) = \alpha. \tag{13}$$

- α is usually **arbitrary** chosen to be 0.01, 0.05 or 0.10

Probability value (p-value)

- Standard practice is to use **the probability value (p-value)**. This is the is the smallest significance level at which the null hypothesis could be rejected.
- Given p-value we do not have to compare test statistics with the corresponding critical value.
- If the p-value is lower than the significance level (α) then we are able to reject the null.

Testing simply hypotheses

- Based on the t statistics:

$$t = \frac{\hat{\beta}_i^{LS} - \beta_i}{se\left(\beta_i^{LS}\right)} \sim t_{N-(K+1)}. \tag{14}$$

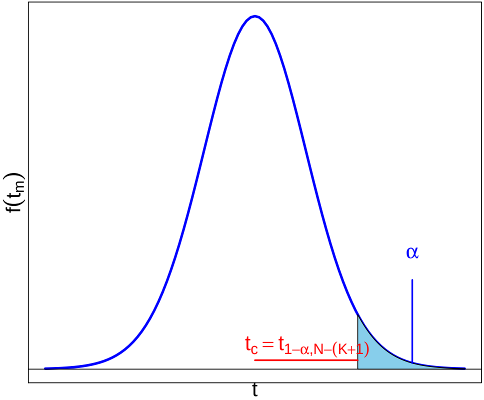
  we can consider the following alternative hypotheses:

  1. $\mathcal{H}_1 : \quad \beta_i \leq c,$
  2. $\mathcal{H}_1 : \quad \beta_i \neq c,$
  3. $\mathcal{H}_1 : \quad \beta_i \geq c.$

- **Test of significance**:
  - The null $\mathcal{H}_1 : \quad \beta_i = 0,$
  - The alternative $\mathcal{H}_0 : \quad \beta_i \neq 0.$
  - t-test statistics:

$$t = \frac{\hat{\beta}_i^{LS}}{se\left(\beta_i^{LS}\right)} \sim t_{N-(K+1)}, \tag{15}$$

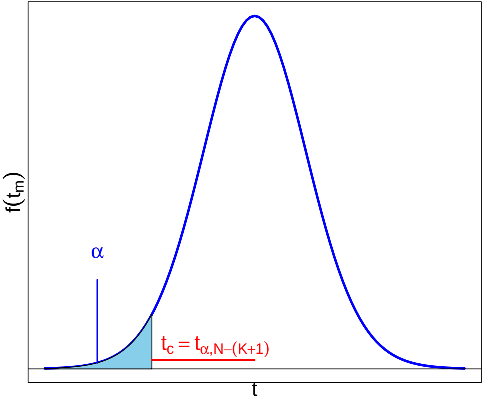
    which is an inverse of relative standard error.



The null and alternative:

$$\mathcal{H}_0 : \quad \beta_i \quad = c$$
$$\mathcal{H}_1 : \quad \beta_i \quad \geq c$$

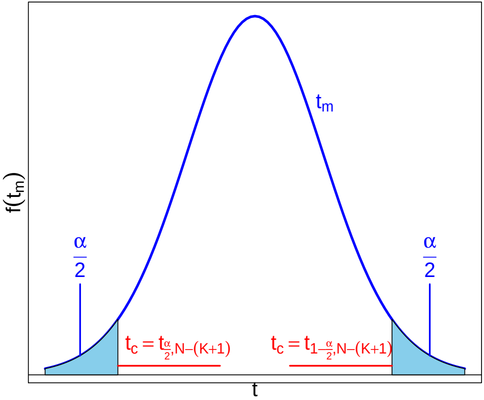
The null hypothesis can be rejected if $t \geq t_{(1-\alpha, N-(K+1))}$



The null and alternative:

$$\begin{aligned} \mathcal{H}_0 : \quad & \beta_i && = c \\ \mathcal{H}_1 : \quad & \beta_i && \leq c \end{aligned}$$

The null hypothesis can be rejected if $t \leq t_{(\alpha, N-(K+1))}$



The null and alternative:

$$\mathcal{H}_0: \quad \beta_i = c$$
$$\mathcal{H}_1: \quad \beta_i \neq c$$

The null hypothesis can be rejected if
$t \leq t_{(\alpha/2, N-(K+1))}$
or $t \geq t_{(1-\alpha/2, N-(K+1))}$

- A linear combination of parameters:

$$\lambda = c_1\beta_1 + c_2\beta_2 \tag{16}$$

where c_1 and c_2 are some constants.

- Under the assumptions #1-#6 (without normality of the error term) the least square estimators $\hat{\beta}_1^{LS}$ and $\hat{\beta}_2^{LS}$ are the best linear unbiased estimators of β_1 and β_2.
- Moreover, the $\hat{\lambda}^{LS} = c_1\hat{\beta}_1^{LS} + c_2\hat{\beta}_2^{LS}$ is also BLUE of λ.
  - The estimator $\hat{\lambda}^{LS}$ is unbiased because:

$$\mathbb{E}(\hat{\lambda}^{LS}) = \mathbb{E}(c_1\hat{\beta}_1^{LS}) + \mathbb{E}(c_2\hat{\beta}_2^{LS}) = c_1\mathbb{E}(\hat{\beta}_1^{LS}) + c_1\mathbb{E}(\hat{\beta}_2^{LS}) = c_1\beta_1 + c_2\beta_2 = \lambda. \tag{17}$$

- The variance of the linear combination of the LS estimates:

$$var(\hat{\lambda}) = var(c_1\hat{\beta}_1^{LS} + c_2\hat{\beta}_2^{LS}) \tag{18}$$

$$= c_1 var(\hat{\beta}_1^{LS}) + c_2 var(\hat{\beta}_2^{LS}) + 2c_1c_2cov(\hat{\beta}_1^{LS}, \hat{\beta}_2^{LS}). \tag{19}$$

therefore we can estimate the variance of the λ by replacing with the (known) estimated variances and covariance.

$$\hat{var}(\hat{\lambda}) = c_1\hat{var}(\hat{\beta}_1^{LS}) + c_2\hat{var}(\hat{\beta}_2^{LS}) + 2c_1c_2\hat{cov}(\hat{\beta}_1^{LS}, \hat{\beta}_2^{LS}) \tag{20}$$

- if the assumption of the error term normality holds or if the sample is large then $\hat{\lambda}$ have normal distribution:

$$\hat{\lambda} = c_1\hat{\beta}_1^{LS} + c_2\hat{\beta}_2^{LS} \sim \mathcal{N}\left(\lambda, var(\hat{\lambda})\right). \tag{21}$$

- The standard t-statistics for the linear combination is:

$$t = \frac{\hat{\lambda} - \lambda}{\sqrt{var(\hat{\lambda})}} = \frac{\hat{\lambda} - \lambda}{se(\hat{\lambda})} \sim t_{N-K}. \tag{22}$$

- Based on the above formulation the variety of hypotheses can be tested. The null is typically:

$$\mathcal{H}_0: \quad \lambda = c_1\beta_1 + c_2\beta_2 = \lambda_0, \tag{23}$$

while the possible alternative hypotheses:

$$\mathcal{H}_1: \quad \lambda = c_1\beta_1 + c_2\beta_2 \neq \lambda_0,$$
$$\mathcal{H}_1: \quad \lambda = c_1\beta_1 + c_2\beta_2 \leq \lambda_0,$$
$$\mathcal{H}_1: \quad \lambda = c_1\beta_1 + c_2\beta_2 \geq \lambda_0.$$

Confidence intervals

Point vs Interval estimation

- **Point estimate** is a single value of the estimator (mean).
- **Interval estimation** provides a range of values in which the true parameter is likely to fall
- **Interval estimation** allows to account for the precision with which the unknown parameter is estimated. The precision is typically measured with variance.

- Under the assumption of normality of the error term the least squares estimator of $\hat{\beta}^{LS}$ is:
$$\hat{\beta}^{LS} \sim \mathcal{N}\left(\beta, \Sigma\right) \tag{24}$$
where Σ is the variance-covariance of the least squares estimator.
- For illustrative purpose we focus on slope parameter in the simple regression model ($\hat{\beta}_1^{LS}$):
$$\hat{\beta}_1^{LS} \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_i^N \left(x_i - \bar{x}\right)^2}\right) \tag{25}$$
- A standardized normal random variable can be obtained from $\hat{\beta}_1^{LS}$ by subtracting its mean and dividing by its standard deviation
$$Z = \frac{\hat{\beta}_1^{LS} - \beta_1}{\sqrt{\sigma^2 \sum_i^N \left(x_i - \bar{x}\right)^2}} \sim \mathcal{N}\left(0, 1\right). \tag{26}$$
- Based on the features of standard normal distribution:
$$P\left(-1.96 \leq Z \leq 1.96\right) = .95, \tag{27}$$

- we can substitute Z

$$P\left(-1.96 \leq \frac{\hat{\beta}_1^{LS}-\beta_1}{\sqrt{\sigma^2 \sum_i^N (x_i-\bar{x})^2}} \leq 1.96\right) = .95, \tag{28}$$

and after manipulations:

$$P\left(\hat{\beta}_1^{LS} - 1.96\sqrt{\sigma^2 \sum_i^N (x_i-\bar{x})^2} \leq \beta_1 \leq \hat{\beta}_1^{LS} + 1.96\sqrt{\sigma^2 \sum_i^N (x_i-\bar{x})^2}\right) = .95. \tag{29}$$

- The two end-points $\hat{\beta}_1^{LS} \pm 1.96\sqrt{\hat{\sigma}^2 \sum_i^N (x_i-\bar{x})^2}$ provide an interval estimator.
- In repeated sampling 95% of the intervals constructed this way will contain the true value of the parameter β_1.
- This easy derivation of an interval estimator is based on the assumption about normality of the error term and that we know the variance of the error term σ^2.

Obtaining interval estimates

- In simply regression, replacing σ^2 by its estimates $\hat{\sigma}^2$ produces a random variable t:

$$t = \frac{\hat{\beta}_1^{LS} - \beta_1}{\sqrt{\hat{\sigma}^2 \sum_i^N (x_i - \bar{x})^2}} = \frac{\hat{\beta}_1^{LS} - \beta_1}{\sqrt{\hat{v}ar\left(\hat{\beta}_1^{LS}\right)}} = \frac{\hat{\beta}_1^{LS} - \beta_1}{se\left(\beta_1^{LS}\right)}. \tag{30}$$

- In multiple regression model, the t ratio, i.e. $t = (\hat{\beta}_j^{LS} - \beta_j)/se(\hat{\beta}_j^{LS})$ has a t-distribution with $N - (K+1)$ degrees of freedoms:

$$t \sim t_{N-(K+1)}, \tag{31}$$

where K is the number of explanatory variables.

- **Critical value** from a t distribution t_c can be found as follows:

$$P(t \geq t_c) = P(t \leq -t_c) = \frac{\alpha}{2}, \tag{32}$$

where α is arbitrary probability (significance level).

- The confidence intervals:

$$P(-t_c \leq t \leq t_c) = 1 - \alpha, \tag{33}$$

and after manipulations (with definition of t random variable):

$$P\left(\hat{\beta}_1^{LS} - t_c se\left(\hat{\beta}_1^{LS}\right) \leq \beta_1 \leq \hat{\beta}_1^{LS} + t_c se\left(\hat{\beta}_1^{LS}\right)\right) = 1 - \alpha. \tag{34}$$

Testing joint hypotheses

- A null hypothesis with multiple conjectures, expressed with more than one equal sign, is called a **joint hypothesis**.
- [Example] Wages (w) and experience ($exper$):

$$w = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \varepsilon. \tag{35}$$

  - Are wages related to experience?
  - To answer the above question we should test jointly $\mathcal{H}_0 : \beta_1 = 0$ and $\mathcal{H}_0 : \beta_2 = 0$.
  - The joint null is $\mathcal{H}_0 : \beta_1 = \beta_2 = 0$.
  - Test of $\mathcal{H}_0$ is a joint test for whether all two conjectures hold simultaneously.

- **The restricted least square estimator** is obtained by minimizing the sum of squares (SSE) subject to **set of restrictions**, which is a function of the unknown parameters, given the data:

$$SSE(\beta_0, \beta_1, \ldots, \beta_K) = \sum_{i=1}^{N} [y_i - \beta_0 - \beta_1 x_1 - \ldots - \beta_K x_K]^2$$

$$\text{subject to} \quad \text{restrictions.}$$

- Examples of restrictions:
  - $\beta_1 = \beta_2$,
  - $\beta_1 = 2$.

Wald test I

- **Wald test** allows to test a set of linear restrictions.
- **The $\mathcal{F}$-statistic** determines what constitutes a large reduction or a small reduction in the sum of squared errors:

$$\mathcal{F} = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}, \tag{36}$$

  where:
  - J is the number of restrictions,
  - N is the number of observations,
  - K is the number of coefficients in the unrestricted model,
  - SSE_R is sum of squared error in **restricted** model,
  - SSE_U is sum of squared error in **unrestricted** model,
- If the null is true then the $\mathcal{F}$-statistic has an $\mathcal{F}$-distribution with J numerator degrees of freedom and $N-K$ denominator degrees of freedom.
- If the null can be rejected then, the differences in sum of squared errors between **restricted model (SSE_R)** and **unrestricted model (SSE_U)** become large.
  - In other words, the imposed restriction significantly reduce the ability of the model to fit the data.
- The $\mathcal{F}$-test can also be used in many application:
  - Testing economic hypotheses.

- Testing the significance of the model.
- Excluding/including a set of explanatory variables.

- Alternatively, the $\mathcal{W}$ statistics can be used which is defined as

$$\mathcal{W} = J\mathcal{F}, \tag{37}$$

and $\mathcal{W}$ is χ^2 distributed with the J degrees of freedom.

- Multiple regression model with K explanatory variable:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_2 x_k + \varepsilon. \quad (38)$$

- **Test of the overall significance of the regression model**. The null hypothesis:

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0, \quad (39)$$

while the alternative is that at least one coefficient is different from 0.

- In this test the restricted model:

$$y = \beta_0, \quad (40)$$

which implies that $SSE_R = SST$.

- Thus, the $\mathcal{F}$ -statistic in the overall significance test can be written as:

$$\mathcal{F} = \frac{(SST - SSE)/K}{SSE/(N - K - 1)}. \quad (41)$$

Linear restriction in matrix form

- **General notation:**

$$\mathbf{R} \times \beta = q, \tag{42}$$

where $\mathbf{R}$ is the $J \times (K+1)$ matrix describing linear restriction and q is the vector of intercepts in each restriction.

- **Example #1:** test of the overall significance of the regression model:

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- **Example #2:** the following restrictions
  1. $\beta_1 = \beta_3$
  2. $\beta_2 = \nu$
  3. $\beta_1 + \beta_4 = \gamma$.

  can be described as

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} 0 \\ \nu \\ \gamma \end{bmatrix}$$

- If a single restriction is considered bot t and $\mathcal{F}$ statistics can be used,
- The results will be identical.
- This is due to an exact relationship between t- and $\mathcal{F}$-distributions. The square of a t random variable with *df* degrees of freedom is an $\mathcal{F}$ random variable with 1 degree of freedom in the numerator and *df* degrees of freedom in the denominator.

Non-sample information

- In many cases we have information over and above the information contained in the sample observation.
- This **non-sample** information can be taken from e.g. economic theory.
- [Example] Production function. Consider the regression of logged output (y) on logged capital (k) and logged labor input (l):

$$y = \beta_0 + \beta_1 k + \beta_2 l + \varepsilon. \tag{43}$$

The natural assumption to verify is constant return to scale (CRS). In this case:

$$\beta_1 + \beta_2 = 1. \tag{44}$$

Delta method

Delta method I

- **Delta method** is the popular strategy for estimating variance for **nonlinear function of the parameters.**
- **Key assumption:** the $g(\beta)$ is the nonlinear continuous function of the parameters.
- Taylor expansion around true value of the parameters, i.e., β:

$$g(\hat{\beta}) = g(\beta) + \left(\frac{\partial g(\beta)}{\partial \beta}\right)' (\hat{\beta} - \beta) + o(||\hat{\beta} - \beta||), \tag{45}$$

where

$$\frac{\partial g(\beta)}{\partial \beta} = \left[\frac{\partial g}{\partial \beta_1}, \frac{\partial g}{\partial \beta_K}, \ldots, \frac{\partial g}{\partial \beta_K}\right]'. \tag{46}$$

- After manipulation

$$g(\hat{\beta}) - g(\beta) = \left(\frac{\partial g(\beta)}{\partial \beta}\right)' (\hat{\beta} - \beta) + o(||\hat{\beta} - \beta||), \tag{47}$$

and taking the variance

$$var\left(g(\hat{\beta}) - g(\beta)\right) = \left(\frac{\partial g(\beta)}{\partial \beta}\right)' var\left(\hat{\beta}\right) \left(\frac{\partial g(\beta)}{\partial \beta}\right). \tag{48}$$