Lecture 1: OLS revisited

Econometric Methods

Andrzej Torój

SGH Warsaw School of Economics – Institute of Econometrics

Outline

Outline

Course information

- lecturers: Andrzej Torój & Bartosz Olesiński
- my website: http://e-web.sgh.waw.pl/atoroj/ (lecture slides, exercise files, literature, contact)
- final grade: details on the website

Course information

- lecturers: Andrzej Torój & Bartosz Olesiński
- my website: http://e-web.sgh.waw.pl/atoroj/ (lecture slides, exercise files, literature, contact)
- final grade: details on the website

Course information

- lecturers: Andrzej Torój & Bartosz Olesiński
- my website: `http://e-web.sgh.waw.pl/atoroj/` (lecture slides, exercise files, literature, contact)
- final grade: details on the website

Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios

Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios

Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios

Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios

Why econometrics?

- investigation of relationships
- finding parameter values in economic models (e.g. elasticities)
- confronting economic theories with data
- forecasting
- simulating policy scenarios

Example (1/4)

Student satisfaction survey

Master students of Applied Econometrics at Warsaw School of Economics in Winter semester 2016/2017 were asked about their satisfaction from studying to be evaluated from 0 to 100. In addition, their average note from previous studies and their sex were registered.

1. What kind of data is this? Cross-section, time series, panel? Frequency? Micro- or macroeconomic?
2. How can we quickly visualise a hypothesised causality from average note to satisfaction from studying?
3. Does such a relationship seem to be there?
4. How can sex of the respondent potentially affect the satisfaction from studies or the relationship in question? How can we visualise this?
5. Bottom line, what is the right specification of the linear regression model?

Outline

Linear regression model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \varepsilon_i =$$

$$\begin{bmatrix} 1 & x_{1,i} & x_{2,i} & \ldots & x_{k,i} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon_i = \mathrm{x}_i \boldsymbol{\beta} + \varepsilon_i$$

Vector of parameters $\begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \ldots & \beta_k \end{bmatrix}^T$ is unknown.
Minimize the dispersion of ε_i around zero, as measured e.g. by $\sum_{t=1}^{n} \varepsilon_i^2$.

Ordinary Least Squares (OLS)

$$S=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1x_{1,i}-\beta_2x_{2,i}-\ldots-\beta_kx_{k,i}\right)^2\to\min_{\beta_0,\beta_1,\ldots}$$

$$\text{FOC: } \frac{\partial S}{\partial \boldsymbol{\beta}}=0$$

$$\text{Denote: } \mathrm{y}=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_n \end{bmatrix},\ \mathrm{X}=\begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \ldots & x_{k,1}\\ 1 & x_{1,2} & x_{2,2} & \ldots & x_{k,2}\\ \vdots & \vdots & \vdots & \ldots & \vdots\\ 1 & x_{1,n} & x_{2,n} & \ldots & x_{k,n} \end{bmatrix},\ \boldsymbol{\beta}=\begin{bmatrix} \beta_0\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_k \end{bmatrix}$$

and obtain:

$$\boldsymbol{\beta}=\left(\mathrm{X}^T\mathrm{X}\right)^{-1}\mathrm{X}^T\mathrm{y}$$

Proof

$$
\begin{aligned}
S &= \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathrm{y} - \mathrm{X}\boldsymbol{\beta})^T (\mathrm{y} - \mathrm{X}\boldsymbol{\beta}) = \\
&= \mathrm{y}^T \mathrm{y} - \boldsymbol{\beta}^T \mathrm{X}^T \mathrm{y} - \mathrm{y}^T \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathrm{X}^T \mathrm{X}\boldsymbol{\beta} = \\
&= \mathrm{y}^T \mathrm{y} - 2\mathrm{y}^T \mathrm{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathrm{X}^T \mathrm{X}\boldsymbol{\beta}
\end{aligned}
$$

(2. and 3. component were transposed scalars, so they were equal)

$\frac{\partial S}{\partial \boldsymbol{\beta}} = 0 \Longleftrightarrow \frac{\partial \mathrm{y}^T \mathrm{y}}{\partial \boldsymbol{\beta}} - \frac{2\mathrm{y}^T \mathrm{X}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} + \frac{\boldsymbol{\beta}^T \mathrm{X}^T \mathrm{X}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = 0$

According to the rules of matrix calculus:

$-2\mathrm{y}^T \mathrm{X} + \boldsymbol{\beta}^T \left(2\mathrm{X}^T \mathrm{X}\right) = 0 \dashrightarrow \boldsymbol{\beta}^T \left(\mathrm{X}^T \mathrm{X}\right) = \mathrm{y}^T \mathrm{X} \dashrightarrow \left(\mathrm{X}^T \mathrm{X}\right) \boldsymbol{\beta} = \mathrm{X}^T \mathrm{y}$

$\boldsymbol{\beta} = \left(\mathrm{X}^T \mathrm{X}\right)^{-1} \mathrm{X}^T \mathrm{y}$

Example (2/4)

Student satisfaction survey

1. Run the regression model with an automated command in R.
2. Write the equation and try to interpret the parameters. Be careful – it's tricky! (Why?)
3. Manually replicate the parameter estimates.

Estimator as a random variable

- $\hat{\beta}$ is an **estimator** of the true parameter value β (function of the random sample choice)
- samples, and hence the values of $\hat{\beta}$, can be different
- estimator as a *(vector) random variable* has its **variance(-covariance matrix)**

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \qquad Var\left(\hat{\beta}\right) =$$

$$\begin{bmatrix} var\left(\hat{\beta}_0\right) & cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & \cdots & \\ cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & var\left(\hat{\beta}_1\right) & cov\left(\hat{\beta}_1, \hat{\beta}_2\right) & \cdots & \\ cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & cov\left(\hat{\beta}_1, \hat{\beta}_2\right) & var\left(\hat{\beta}_2\right) & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ & & & & var\left(\hat{\beta}_k\right) \end{bmatrix}$$

Estimator as a random variable

- $\hat{\beta}$ is an **estimator** of the true parameter value β (function of the random sample choice)
- samples, and hence the values of $\hat{\beta}$, can be different
- estimator as a *(vector) random variable* has its **variance(-covariance matrix)**

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \qquad Var\left(\hat{\beta}\right) =$$

$$\begin{bmatrix} var\left(\hat{\beta}_0\right) & cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & \cdots & \\ cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & var\left(\hat{\beta}_1\right) & cov\left(\hat{\beta}_1, \hat{\beta}_2\right) & \cdots & \\ cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & cov\left(\hat{\beta}_1, \hat{\beta}_2\right) & var\left(\hat{\beta}_2\right) & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ & & & & var\left(\hat{\beta}_k\right) \end{bmatrix}$$

Estimator as a random variable

- $\hat{\beta}$ is an **estimator** of the true parameter value β (function of the random sample choice)
- samples, and hence the values of $\hat{\beta}$, can be different
- estimator as a *(vector) random variable* has its **variance(-covariance matrix)**

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad Var\left(\hat{\beta}\right) =$$

$$\begin{bmatrix} var\left(\hat{\beta}_0\right) & cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & \cdots & \\ cov\left(\hat{\beta}_0, \hat{\beta}_1\right) & var\left(\hat{\beta}_1\right) & cov\left(\hat{\beta}_1, \hat{\beta}_2\right) & \cdots & \\ cov\left(\hat{\beta}_0, \hat{\beta}_2\right) & cov\left(\hat{\beta}_1, \hat{\beta}_2\right) & var\left(\hat{\beta}_2\right) & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ & & & & var\left(\hat{\beta}_k\right) \end{bmatrix}$$

Variance-covariance matrix of a random vector

- Definition:
  $Var(\beta) = E\left\{[\beta - E(\beta)][\beta - E(\beta)]^T\right\}$
- For a centered variable, i.e. $E(\varepsilon) = 0$, this definition simplifies:
  $Var(\varepsilon) = E\left(\varepsilon\varepsilon^T\right)$

OLS estimator: properties

$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^T\mathrm{X}\right)^{-1}\mathrm{X}^T\mathrm{y}$ is an **estimator** (function of the sample) of the "true", unknown values $\boldsymbol{\beta}$ (population / data generating process). Under certain conditions (i.a. $E(\boldsymbol{X}^T\varepsilon) = 0$ $E(\varepsilon\varepsilon^T) = \sigma^2\boldsymbol{I}$), the OLS estimator is:

- **unbiased**: $E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$
- **consistent**: $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ with growing n
- **efficient**: least possible estimator variance (i.e. highest precision)

OLS estimator: properties

$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^T\mathrm{X}\right)^{-1}\mathrm{X}^T\mathrm{y}$ is an **estimator** (function of the sample) of the "true", unknown values $\boldsymbol{\beta}$ (population / data generating process). Under certain conditions (i.a. $E(\boldsymbol{X}^T\varepsilon) = 0$ $E(\varepsilon\varepsilon^T) = \sigma^2\boldsymbol{I}$), the OLS estimator is:

- **unbiased**: $E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$
- **consistent**: $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ with growing n
- **efficient**: least possible estimator variance (i.e. highest precision)

OLS estimator: properties

$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^T\mathrm{X}\right)^{-1}\mathrm{X}^T\mathrm{y}$ is an **estimator** (function of the sample) of the "true", unknown values $\boldsymbol{\beta}$ (population / data generating process). Under certain conditions (i.a. $E(\boldsymbol{X}^T\varepsilon) = 0$ $E(\varepsilon\varepsilon^T) = \sigma^2\boldsymbol{I}$), the OLS estimator is:

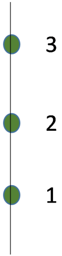
- **unbiased**: $E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$
- **consistent**: $\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ with growing n
- **efficient**: least possible estimator variance (i.e. highest precision)

Variance of the error term (1)

- **1. Variance of the error term** (scalar): $\hat{\sigma}^2 = \frac{1}{n-(k+1)} \sum_{i=1}^{n} \varepsilon_i^2$
- Why such a formula if the general formula is
  $Var(X) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$?
  - First of all note that $\bar{\varepsilon} = 0$ (prove it on your own).
  - Second, we need to know why 1 turned into $(k+1)$ in the denominator.

Variance of the error term (2)

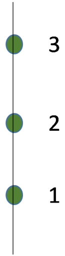
- **By Your intuition, what is the standard deviation in the following dataset of 3 observation?**



- Without a correction in denominator:
$\sqrt{Var} = \sqrt{\frac{1}{3}\left[(3-2)^2 + (2-2)^2 + (1-2)^2\right]} = \sqrt{\frac{2}{3}} \neq 1$

Variance of the error term (2)

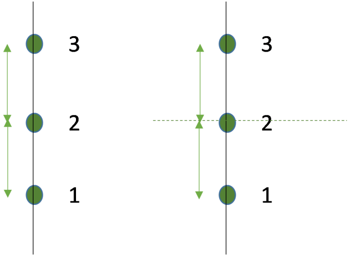
- **By Your intuition, what is the standard deviation in the following dataset of 3 observation?**



- Without a correction in denominator:
$$\sqrt{Var} = \sqrt{\frac{1}{3}\left[(3-2)^2 + (2-2)^2 + (1-2)^2\right]} = \sqrt{\frac{2}{3}} \neq 1$$

Variance of the error term (3)

- The intuition behind the standard deviation of 1 is build upon an implicit, graphical calibration of mean based on the data sample.

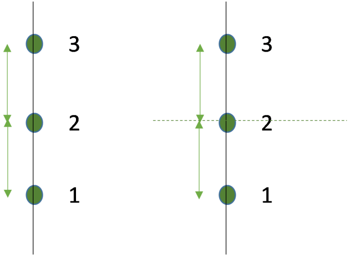


- With an adequate correction for thatin denominator:

$$\sqrt{Var} = \sqrt{\frac{1}{3-1}\left[(3-2)^2 + (2-2)^2 + (1-2)^2\right]} = \sqrt{\frac{2}{2}} = 1$$

Variance of the error term (3)

- The intuition behind the standard deviation of 1 is build upon an implicit, graphical calibration of mean based on the data sample.



- With an adequate correction for thatin denominator:

$$\sqrt{Var} = \sqrt{\frac{1}{3-1}\left[(3-2)^2 + (2-2)^2 + (1-2)^2\right]} = \sqrt{\frac{2}{2}} = 1$$

Variance of the error term (4)

- When X is directly observed, the terms like $(x_i - \bar{x})$ consume one degree of freedom (there is one $\bar{x}$ estimated before).
- When ε is not observed, the terms $\varepsilon_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - ... - \hat{\beta}_k x_{ki}$ consume $k + 1$) degrees of freedom (there are $k + 1$ elements in vector $\hat{\boldsymbol{\beta}}$ estimated before).

Variance-covariance matrix of the estimator

$$
\begin{aligned}
Var\left(\hat{\boldsymbol{\beta}}\right) &= E\left[\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T\right] = \\
&= E\left\{\left[\left(X^TX\right)^{-1}X^T\mathbf{y}-\boldsymbol{\beta}\right]\cdot\left[\left(X^TX\right)^{-1}X^T\mathbf{y}-\boldsymbol{\beta}\right]^T\right\} = \\
&= E\left\{\left[\left(X^TX\right)^{-1}X^T\left(X\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)-\boldsymbol{\beta}\right]\cdot\left[\left(X^TX\right)^{-1}X^T\left(X\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)-\boldsymbol{\beta}\right]^T\right\} = \\
&= E\left[\left(X^TX\right)^{-1}X^T\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}^TX\left(X^TX\right)^{-1}\right] = \\
&= \left(X^TX\right)^{-1}X^TE\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)X\left(X^TX\right)^{-1} = \\
&= \left(X^TX\right)^{-1}X^T\sigma^2IX\left(X^TX\right)^{-1} = \\
&= \sigma^2\left(X^TX\right)^{-1}X^TX\left(X^TX\right)^{-1} = \\
&= \sigma^2\left(X^TX\right)^{-1}
\end{aligned}
$$

Empirical counterpart: $Var\left(\hat{\boldsymbol{\beta}}\right)=\hat{\sigma}^2\left(X^TX\right)^{-1}\equiv\left[d_{i,j}\right]_{(k+1)\times(k+1)}$

Standard errors of estimation

- **Standard errors of estimation** (vector – for each parameter): $s\left(\hat{\beta}_0\right) = \sqrt{d_{1,1}} \quad s\left(\hat{\beta}_1\right) = \sqrt{d_{2,2}} \quad s\left(\hat{\beta}_2\right) = \sqrt{d_{3,3}} \ldots$

Calculating S.E.

1. estimate parameters, 2. compute the empirical error terms, 3. estimate their variance, 4. compute the variance-covariance matrix of the OLS estimator, 5. compute the SE as a square root of its diagonal elements.

Standard errors of estimation

- **Standard errors of estimation** (vector – for each parameter):
  $s\left(\hat{\beta}_0\right) = \sqrt{d_{1,1}} \quad s\left(\hat{\beta}_1\right) = \sqrt{d_{2,2}} \quad s\left(\hat{\beta}_2\right) = \sqrt{d_{3,3}} \ldots$

Calculating S.E.

1. estimate parameters, 2. compute the empirical error terms, 3. estimate their variance, 4. compute the variance-covariance matrix of the OLS estimator, 5. compute the SE as a square root of its diagonal elements.

t-tests for variable significance

t-Student test

$H_0 : \beta_i = 0$, i.e. i-th explanatory variable does not significantly influence y

$H_1 : \beta_i \neq 0$, i.e. i-th explanatory variable does not significantly influence y

Test statistic: $t = \frac{\hat{\beta}_i}{s(\hat{\beta}_1)}$ is distributed as $t\left(n-k-1\right)$.

p-value$<\alpha^*$ – reject H_0

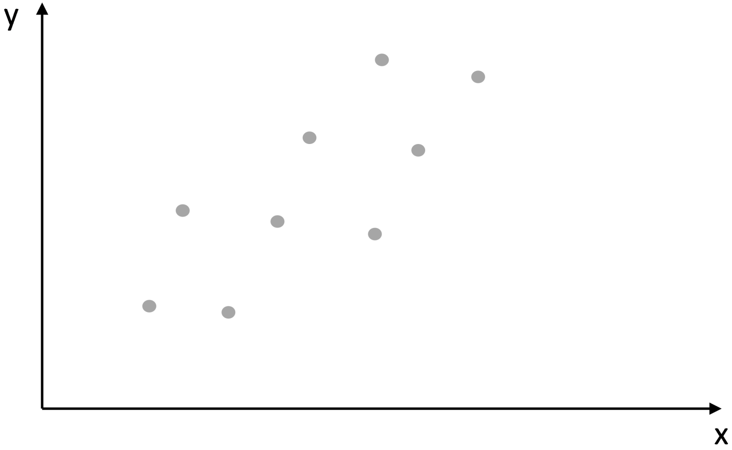
p-value$>\alpha^*$ – **do not reject** H_0

Example (3/4)

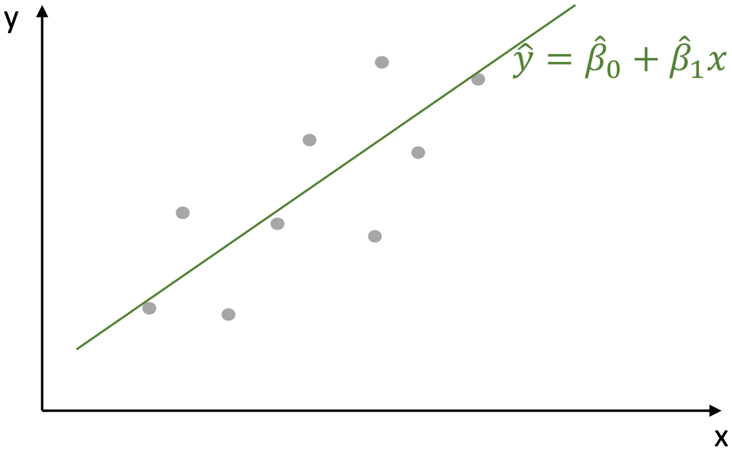
Student satisfaction survey

1. Compute the fitted values and the error terms.
2. Use this result to estimate the variance of the error term.
3. Estimate the variance-covariance matrix of the $\hat{\beta}$ estimates.
4. Derive the standard errors from this matrix.
5. Replicate and interpret the t statistics and the p-values.

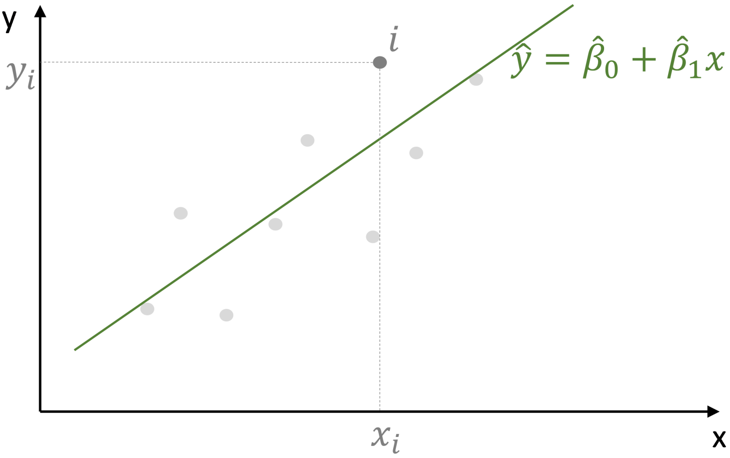
R-squared (1)



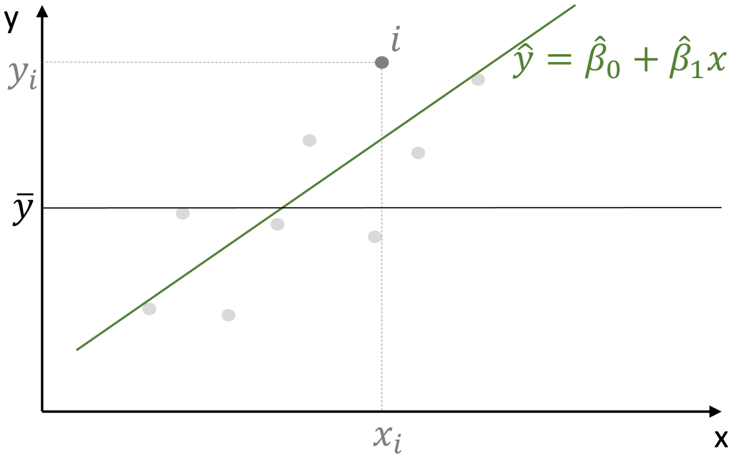
R-squared (2)



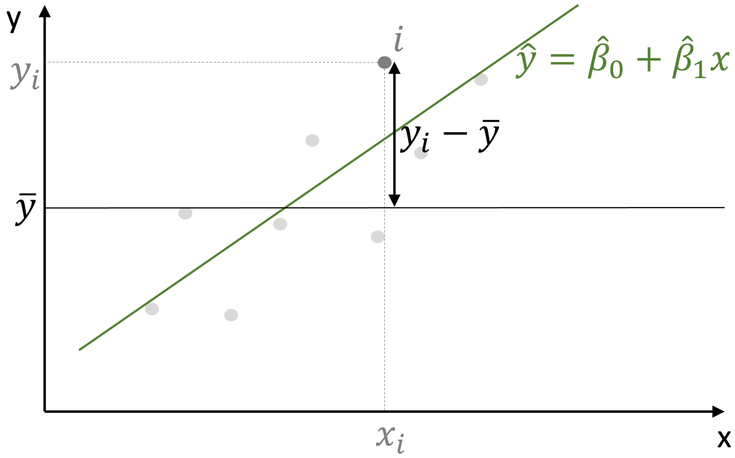
R-squared (3)



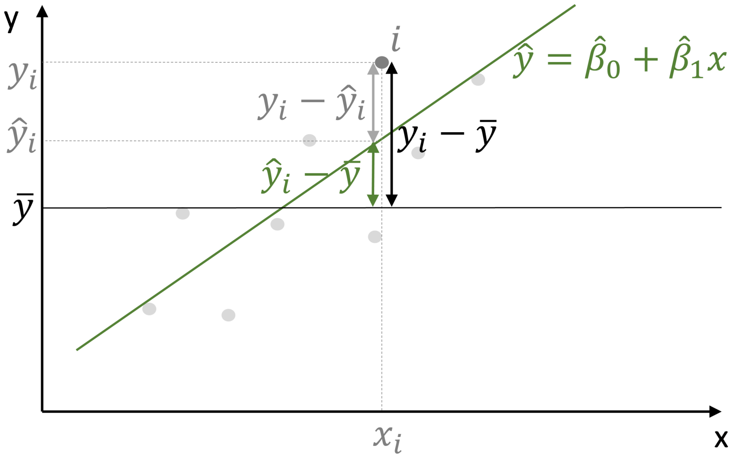
R-squared (4)



R-squared (5)



R-squared (6)


y
i
y_i
$\hat{y}_i$
$\bar{y}$
$y_i - \hat{y}_i$
$y_i - \bar{y}$
$\hat{y}_i - \bar{y}$
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
x_i
x

R-squared (7)

$R^2 \in [0;1]$ is a share of y_t volatility **explained by the model** in total y_t volatility:

$$\sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2 + \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \qquad R^2 = \frac{\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Standard goodness-of-fit measure in OLS regressions with a constant.

Wald test statistic

Wald test

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$, i.e. no explanatory variable influences y

$H_1 : \exists_i \qquad \beta_i \neq 0$, at least 1 explanatory variable influences y

Test statistic: $F = \frac{R^2/k}{(1-R^2)/(T-k-1)}$ distributed as $F\left(k,\ T-k-1\right)$.

Adjusted R-squared

$$\bar{R}^2 = \underbrace{R^2}_{fit} - \underbrace{\frac{k}{T-(k+1)}\left(1-R^2\right)}_{penalty\ for\ overparametrization}$$

Example (4/4)

Student satisfaction survey

1. Interpret the F-test result.
2. Replicate the F statistic and its p-value manually.
3. Interpret the R-squared.
4. Replicate the R-squared and adjusted R-squared manually.