Least Squares Method

Econometrics in one slide :)

Questions:

- How does the world work? How does variable x influence on variable y?
- What will happen tomorrow? How to predict the y variable?

Answer:

Model is the formula for the response variable

For example:

- $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

Main types of data:

- Time series
- Cross-sectional data
- Panel data

There are many-many more!

Time series

Data for Russia:

Year	Population	Unemployment
2010	142962	7.4
2011	142914	6.5
2012	143103	5.5
2013	143395	5.5

Cross-sectional sample

2014 Winter Olympics Results:

Country	Gold	Silver	Bronze
Russia	13	11	9
Norway	11	5	10
Canada	10	10	5
USA	9	7	12

Panel data

Combination of the first two: data on several variables for many objects at different time points

Data — denotation

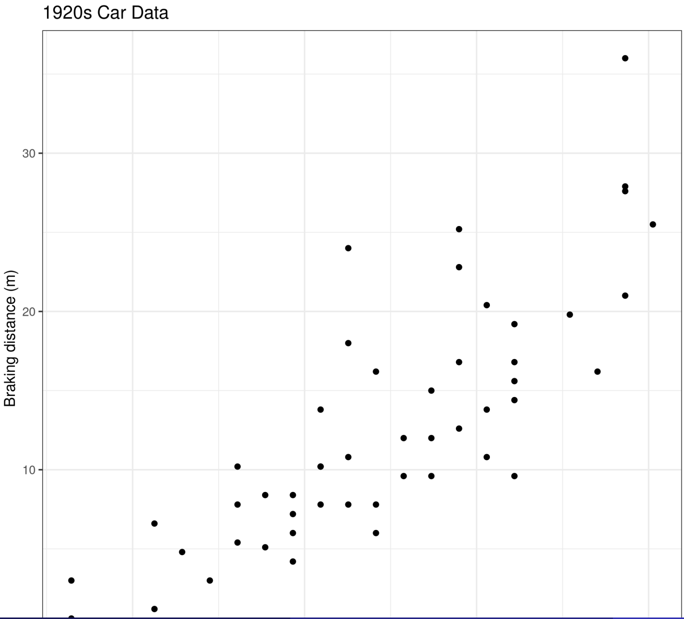
- One dependent, response variable: y
- Several regressors, explanatory variables: x, z, $\ldots$
- n observations for every variable: y_1, y_2, $\ldots$, y_n

Data — example

Historical data from the 1920s :)

Braking distance (m), y_i	Car velocity (km/h), x_i
0.6	6.44
3.0	6.44
1.2	11.27
...	...

Always depict the data!



Model:

Example: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

- Observable variables: y, x
- Unknown parameters: β_1, β_2
- Random component, error: ε

Strategy

- come up with an adequate model
- obtain estimates of unknown parameters: $\hat{\beta}_1$, $\hat{\beta}_2$
- predict, replacing unknown parameters with their estimates:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

Least Squares Method

- A way to obtain estimates of the unknown parameters of the model using real data.

Forecast error: $\hat{\varepsilon}_i = y_i - \hat{y}_i$.

Sum of squared forecast errors:

$$Q(\hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

The essence of LS method: take estimates $\hat{\beta}_1$, $\hat{\beta}_2$ such that the sum of squared forecast errors Q is minimal.

Cars example:

Factual data:

$x_1 = 6.68$, $x_2 = 6.68$, $\dots$,

$y_1 = 0.6$, $y_2 = 3$, $\dots$

Model: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$. Forecast formula: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$

Sum of squared forecast errors: $Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$$Q = (0.6 - \hat{\beta}_1 - \hat{\beta}_2 6.68)^2 + (3 - \hat{\beta}_1 - \hat{\beta}_2 6.68)^2 + ...$$

Minimum point, found in R: $\hat{\beta}_1 = -5.3$, $\hat{\beta}_2 = 0.7$:

Forecast formula: $\hat{y}_i = -5.3 + 0.7 x_i$

Simple example [at the blackboard]

Name	Weight (kg), y_i	Height (cm), x_i
Vasya	60	170
Kolya	70	170
Petya	80	181

Estimate the models:

$y_i = \beta + \varepsilon_i,$

$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

Small preparation: $n\bar{x} = \sum_i x_i = \sum_i \bar{x}$, $\sum_i (x_i - \bar{x}) = 0$.

Final LS formulae. Regression on a constant

In the $y_i = \beta + \varepsilon_i$ model

$$\hat{\beta} = \bar{y}$$

Interpretation:

In a model without explanatory variables the best forecast is the mean of the response variable.

Final LS formulae. Pair regression

In the $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ model

$$\hat{\beta}_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

Interpretation:

The $(\bar{x}, \bar{y})$ point lies on the regression line $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$

Terminology and denotation:

y_i — dependent, response variable

x_i — regressor, explanatory variable

ε_i — error, model error, random component

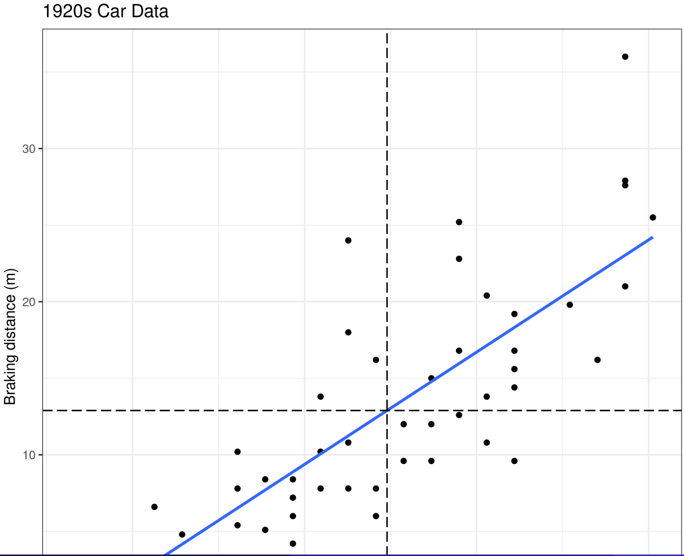
$\hat{y}_i$ — forecast, predicted value

$\hat{\varepsilon}_i = y_i - \hat{y}_i$ — residual, forecast error

$RSS = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$ — residual sum of squares

Regression goes through the midpoint [at the blackboard]

```
## `geom_smooth()` using formula 'y ~ x'
```



Many explanatory variables [at the blackboard]

$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$

Write out the system of equations for the $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ estimates:

$$\begin{cases} \sum \hat{\varepsilon}_i \cdot 1 = 0 \\ \sum \hat{\varepsilon}_i \cdot x_i = 0 \\ \sum \hat{\varepsilon}_i \cdot z_i = 0 \end{cases}$$

Sums of squares

- Residual sum of squares

$$RSS = \sum \hat{\varepsilon}_i^2$$

- Total sum of squares

$$TSS = \sum (y_i - \bar{y})^2$$

- Explained sum of squares

$$ESS = \sum (\hat{y}_i - \bar{y})^2$$

Nuts-and-bolts linear algebra course

Vectors: y, x, $\hat{y}$, ε, ...

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \; x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \; \hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} \; \vec{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

In our model: $\hat{y} = \hat{\beta}_1 \cdot \vec{1} + \hat{\beta}_2 \cdot x + \hat{\beta}_3 \cdot z$

Matrix of all regressors

$$X = \begin{pmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & & \\ 1 & x_n & z_n \end{pmatrix}$$

Vector length

Vector length, $|y| = \sqrt{y_1^2 + y_2^2 + \ldots + y_n^2}$

Vector length squared, $|y|^2 = y_1^2 + y_2^2 + \ldots + y_n^2 = \sum_i y_i^2$

Examples:
$RSS = \sum \hat{\varepsilon}_i^2$ — squared length of the $\hat{\varepsilon}$ vector $TSS = \sum (y_i - \bar{y})^2$ — squared length of the $(y - \bar{y} \cdot \vec{1})$ vector

$$\begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \bar{y} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = y - \bar{y} \cdot \vec{1}$$

Dot product of two vectors:

$$(x, y) = |x| \cdot |y| \cdot cos(x, y)$$

$$(x, y) = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_i x_i y_i$$

Perpendicularity condition:

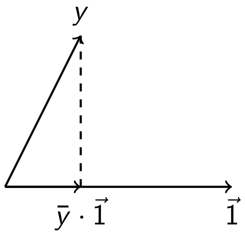
$$x \perp y \Leftrightarrow \sum_i x_i y_i = 0$$

as $cos(90^\circ) = 0$.

Illustration for the regression on a constant [at the blackboard]

Model: $y_i = \beta + \varepsilon_i$

Forecasts: $\hat{y}_i = \hat{\beta} = \bar{y}$



Geometric interpretation of first-order conditions

$$\begin{cases} \sum \hat{\varepsilon}_i \cdot 1 = 0 \\ \sum \hat{\varepsilon}_i \cdot x_i = 0 \\ \sum \hat{\varepsilon}_i \cdot z_i = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} \hat{\varepsilon} \perp \vec{1} \\ \hat{\varepsilon} \perp x \\ \hat{\varepsilon} \perp z \end{cases}$$

Illustration for multivariate regression [at the blackboard]

3dvec-ols.pdf

If a β_1 intercept is included in the regression

If an intercept is included in the regression, $y_i = \beta_1 + \ldots$, and LS-estimates are unique, then:

- $\sum \hat{\varepsilon}_i = 0$
- $\sum y_i = \sum \hat{y}_i$
- $\bar{y} = \bar{\hat{y}}$
- $TSS = RSS + ESS$

Coefficient of determination — a simple quality measure

In the models with an intercept $R^2 = ESS/TSS$

TSS — total dispersion y
ESS — dispersion explained by regressors
R^2 — share of explained dispersion in total dispersion

Theorem. If an intercept is included in the regression, $y_i = \beta_1 + \ldots$, and LS-estimates are unique, then R^2 is equal to the sample correlation between y and $\hat{y}$, i.e.

$$R^2 = (sCorr(y, \hat{y}))^2 = \left(\frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum (y_i - \bar{y})^2} \sqrt{\sum (\hat{y}_i - \bar{y})^2}} \right)^2$$

Explicit formula for coefficient estimates

Model: $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & & \\ 1 & x_n & z_n \end{pmatrix}$$

Linear algebra allows to obtain explicit formulae:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Summing up

HOORAY!!! LS allows us to estimate models!!!

Assuming $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$

we obtain $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$

Questions

- How to choose the structure of the model?
- Will the solution to the minimization problem be unique?
- Will there be a solution to the minimization problem at all?
- Why the residual sum of squares and not, say, modules?
- How accurate are the acquired estimates?
- ...

Sources of wisdom:

- Artamonov N.V., Introduction to Econometrics: chapters 1.1, 1.2, 2.1
- Borzykh D.A., Demeshev B.B., Econometrics in Problems and Exercises: chapter 1
- Katyshev P.K., Peresetskiy A.A., Econometrics. Beginners' Course: chapters 2.1, 2.2, 3.1, 3.2
- Seber G., Linear Regression Analysis: chapters 1.0, 1.1, 1.2, 2.1, 3.1