Applied Statistics and Econometrics
Lecture 4

Giuseppe Ragusa

Luiss University

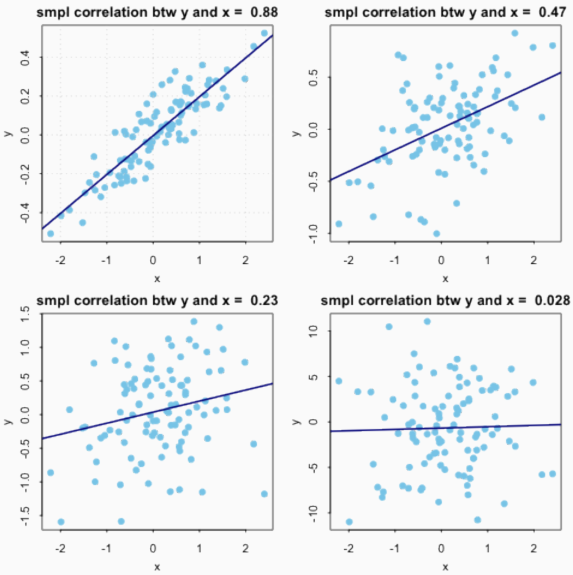
gragusa@luiss.it
http://gragusa.org/

February 19, 2017

Covariance and correlation

- Covariance and correlation are measure of linear dipendence between variables
- The best way of picturing linear association is to immagine to fit a line through the cloud of points in a scatterplot
- The better the could of points be summarized by a line, the stronger is the linear association

Covariance and correlation



The best “line”

In the previous graphics I draw a line through the cloud of points.

Question:

How did I draw this line?

- Put differently: which is the line that better approximate the cloud of points?

The best “line”

In the previous graphics I draw a line through the cloud of points.

Question:

How did I draw this line?

- Put differently: which is the line that better approximate the cloud of points?

Answer:

We use the so called method of least squares.

Method of least squares

Hystory

Carl Friedrich Gauss is credited with developing the fundamentals of the basis for least-squares analysis in 1795 at the age of eighteen. But Adrien Marie Legendre, a french mathematician, was the first to publish the method, however.

- Gauss applied this method to predict the location of Ceres, a dwarf planet discovered by Italian Giuseppe Piazzi in 1801.

Carl Gauss

The ordinary least squares (intuition)

Sample mean

Recall that $\bar{Y}$ was the least squares estimator of μ_Y: $\bar{Y}$ solves,

$$\min_m \sum_{i=1}^{n} (Y_i - m)^2$$

- By taking the first derivative of the objective function and equating to 0, it is easy to see that the solution of $\bar{Y}$

$$-2\sum_{i=1}^{n} Y_i + nm = 0 \implies m = \bar{Y} = \sum_{i=1}^{n} Y_i/n$$

The ordinary least squares (intuition)

Now, we do not have to calculate the mean, but the coefficients of a line:

$$Y = b_0 + b_1 X,$$

where

- b_0 is the intercept
- b_1 is the slope

The ordinary least squares (intuition)

Now, we do not have to calculate the mean, but the coefficients of a line:

$$Y = b_0 + b_1 X,$$

where

- b_0 is the intercept
- b_1 is the slope

Least squares idea

Among all the possible values of (b_0, b_1), we pick the ones that give "predictions" that are closer (in the least squares sense) to the observed data:

$$\hat{Y}_i = b_0 + b_1 X_i, \quad i = 1, \ldots, n$$

The ordinary least squares (intuition)

Assume: $b_0 = -0.6$ and $b_1 = 0.14$

Observed data

y	x
-0.032	-0.626
-0.885	0.184
-7.024	-0.836
-2.677	1.595
1.539	0.330
1.785	-0.820
-5.943	0.487

Predictions

$\hat{Y}_i$
-0.6 $+$ 0.14 $\times$ (-0.626) $=$ -0.68
-0.6 $+$ 0.14 $\times$ (0.184) $=$ -0.57
-0.6 $+$ 0.14 $\times$ (-0.836) $=$ -0.71
-0.6 $+$ 0.14 $\times$ (1.595) $=$ -0.37
-0.6 $+$ 0.14 $\times$ (0.330) $=$ -0.55
-0.6 $+$ 0.14 $\times$ (-0.820) $=$ -0.71
-0.6 $+$ 0.14 $\times$ (-0.487) $=$ -0.63

Least squares (intuition)

Predictions and squares

$\hat{Y}_i$	$(\hat{Y}_i - Y_i)^2$
-0.68	0.431
-0.57	0.096
-0.71	39.776
-0.37	5.291
-0.55	4.379
-0.71	6.249
-0.63	29.28

Sum of squares

The sum of the squares for the line with $b_0 = -0.6$ and $b_1 = 0.14$ is

$$\begin{aligned} SSR =& 0.431 + 0.096 + 39.776 \\ &+ 5.291 + 4.379 + 6.249 + 29.28 \\ =& 85.50 \end{aligned}$$

Least squares

OLS

The least squares ("ordinary least squares" or "OLS") estimators of b_0 and b_1 solve

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

The OLS estimator solves:

$$\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

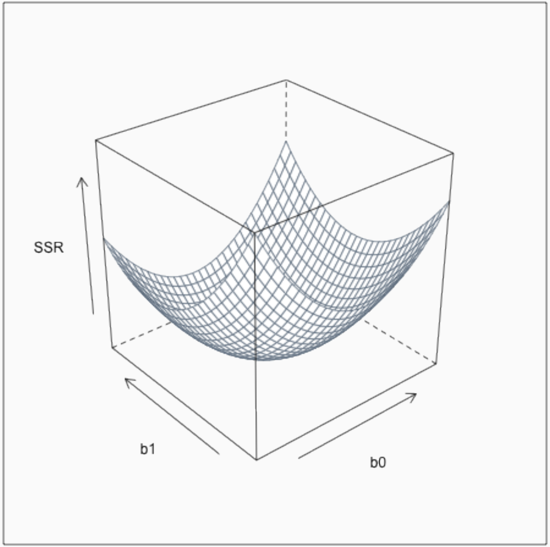
- The OLS estimator minimizes the average squared difference between the actual values of Y_i and the prediction ("predicted value") based on the estimated line $(b_0 + b_1 X_i)$
- We have a multivariate function in (b_0, b_1):

$$SSR(b_0, b_i) = \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

- SSR stands for sum of squared residuals
- The residuals are

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - b_0 - b_1 X_i, \quad i = 1, \ldots, n$$

Sum of squares: graphical representation



Refresh your math

- The minimum value of a function $f(x)$ is denoted

$$\min_x f(x).$$

- The argument for which the function achieves its minimum is denoted

$$\arg\min_x f(x).$$

- If the function is strictly convex, the necessary and sufficient condition for x_0 to be a minimizer of $f(x)$ is

$$\frac{d}{dx} f(x_0) = 0$$

Refresh your math

Example: the math

Let $f(x) = 2 + x^2$. Then,

$$\min_x f(x) = 2$$

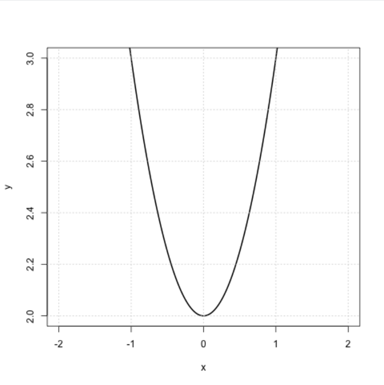
and

$$\arg\min_x f(x) = 0.$$

Also,

$$\frac{d}{dx} f(x) = 2x \implies \frac{d}{dx} f(0) = 0$$

Graph of $2 + x^2$



Least squares first order conditions

$$\min_{b_0,b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

First order conditions

$$\frac{\partial \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2}{\partial b_0} = 0 \implies -2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

$$\frac{\partial \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_1)^2}{\partial b_1} = 0 \implies -2 \sum_{i=1}^{n} X_i (Y_i - b_0 - b_1 X_i) = 0$$

Solving for the first order conditions

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0 \tag{1}$$

$$\sum_{i=1}^{n}X_i(Y_i - b_0 - b_1 X_i) = 0 \tag{2}$$

- Two equations in two unknowns (b_0 and b_1)

Solving for the first order conditions

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0 \tag{1}$$

$$\sum_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) = 0 \tag{2}$$

- Two equations in two unknowns (b_0 and b_1)
- Solve for b_0 in (1), to obtain

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i - b_1 \frac{1}{n}\sum_{i=1}^{n} X_i$$

Solving for the first order conditions

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) = 0 \tag{1}$$

$$\sum_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) = 0 \tag{2}$$

- Two equations in two unknowns (b_0 and b_1)
- Solve for b_0 in (1), to obtain

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i - b_1 \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Substitute b_0 into (2) and solve for b_1

$$b_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Lest squares solution for b_0 and b_1

The convention is to denote the solution of the first order conditions as $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$b_0 = \bar{Y} - b_1 \bar{X}$$

$$b_1 = \frac{s_{xy}}{s_x^2}$$

Remarks

- The underlying assumption for the existence of a solution is that

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 \neq 0$$

(*When does this happen?*)

Remarks

- The underlying assumption for the existence of a solution is that

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 \neq 0$$

  (*When does this happen?*)

- If the FOCs can be solved, the FOCs are sufficient for the minimum (why?)

Remarks

- The least squares method that we just used is usually referred to as "ordinary least squares" (OLS)
- Although b_0 and b_1 are the solutions to a mathematical problem, their are function of sample statistics: sample mean of Y, sample mean X, variance X, and covariance of Y and X
- In due time, we will also see how to solve a more complicated OLS problem, one in which we fit planes or hyper-planes instead of lines.

Conditional expectations as object of interest

We are often interested in the expectation of a r.v. conditional to the value of another.

Example (Wage and education)

- You may (and probably should) what would be your wage after you graduate from college
- If educ denotes years of education and wage the wage rate, you are interested in this quantity

$$E[wage|educ = 16],$$

that is, the expected value of wages for individuals with 16 years of education, exactly the same number of years of education you will have

Conditional expectations as object of interest

Example (Wage and education, ctd.)

- Suppose now you want to know whether to continue your education with a master degree
- To see whether completing a master is worth your money you are concerned about this quantity

$$E[wage|educ = 17] - E[wage|educ = 16],$$

that is, the difference in wage due to the extra year of education

Conditional expectations as object of interest

Policy problem: testscore and str

A school superintend must decide whether to hire new 10 new teachers

- she faces a trade-off because hiring 40 new teachers:
  - will reduce the student-per-teacher (STR) ratio by 2, from 22 to 20
  - will increase expenditures by $1,800,000.

  If she reduces the student-teacher ratio by 2, what will the effect be on standardized test scores in her district?

Policy answer

$$E[testscore|str = 20] - E[testscore|str = 22]$$

A linear model for conditional expectations

- The conditional expectation is a population quantity and so we need to estimate its value
- Unfortunately, estimating conditional expectation is quite difficult is we are not willing to make some assumptions on its form
- A very powerful assumption is the linear one:

Linear assumption

Given a conditional expectation btw Y and X, we assume that this expectation is linear in X, that is,

$$E[Y|X] = \beta_0 + \beta_1 X,$$

where β_0 and β_1 are two parameters

Linear model and conditional expectations

$$E[testscore|str] = \beta_0 + \beta_1 str$$

Linear model and conditional expectations

$$E[testscore|str] = \beta_0 + \beta_1 str$$

- What is the expected value of testscore in a district with $str = 19$?

$$E[testscore|str = 19] = \beta_0 + \beta_1 \times 19$$

Linear model and conditional expectations

$$E[testscore|str] = \beta_0 + \beta_1 str$$

- What is the expected value of testscore in a district with $str = 19$?

$$E[testscore|str = 19] = \beta_0 + \beta_1 \times 19$$

- What is the expected value of testscore in a district with $str = 20$?

$$E[testscore|str = 20] = \beta_0 + \beta_1 \times 20$$

Linear model and conditional expectations

$$E[testscore|str] = \beta_0 + \beta_1 str$$

- What is the expected value of testscore in a district with $str = 19$?

$$E[testscore|str = 19] = \beta_0 + \beta_1 \times 19$$

- What is the expected value of testscore in a district with $str = 20$?

$$E[testscore|str = 20] = \beta_0 + \beta_1 \times 20$$

- What is the expected value of testscore in a district with $str = 23$?

$$E[testscore|str = 23] = \beta_0 + \beta_1 \times 23$$

Linear model and conditional expectations

Question: what is the effect of increasing str by 1?

$$E[testscore|str = 20] - E[testscore|str = 19] = (\beta_0 + \beta_1 20) - (\beta_0 + \beta_1 19)$$
$$= \beta_1$$

Interpretation of β_1

Under the linear assumption, β_1 is the change in the CE of `testscore` when `str` is increased by 1 unit

Linear model and conditional expectations

Question: what is the effect on the CE of increasing str by Δ?

$$E[testscore|str = 20 + \Delta] - E[testscore|str = 20] = \beta_1 \times \Delta$$

Interpretation of β_1

Under the linear assumption, $\beta_1 \times \Delta$ is the change in the CE of testscore when *str* is increased by Δ unit

Linear model and conditional expectations

$$E[Y_i|X_i] = \beta_0 + \beta_1 X_i$$

Interpretation of β_1

$$\beta_1 = \frac{\partial E[Y_i|X_i]}{\partial X_i}$$

that is, β_1 is the effect on the CE of Y when X_i is increased by 1 unit

Linear model and conditional expectations

Interpretation of β_0

$$\beta_0 = E[Y_i|X_i = 0]$$

that is, the CE of Y when $X_i = 0$

The linear assumption

Linear assumption

If

$$E[Y|X] = \beta_0 + \beta_1 X$$

then we can write

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

where

$$E[u|X] = 0$$

.

- u is called the regression error
- it consists of omitted factors, or possibly errors in the measurement of Y.
  - In general, these omitted factors are other variables that influence Y, other than the X

The linear assumption

Implications of the linear assumption

- Since $E[u|X] = 0$ implies that $\text{cov}(u, X) = 0$, the linearity assumption can be interpreted as saying that we are ruling out linear relationship between u and X (that is, the correlation between u and X is zero).
- We will see that this assumption is usually questionable, but for now, we will stick to it

The population linear regression model - general notation

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n$$

- X is the independent variable or regressor
- Y is the dependent variable
- β_0 = intercept
- β_1 = slope
- u_i = the regression error

Terminology in a picture

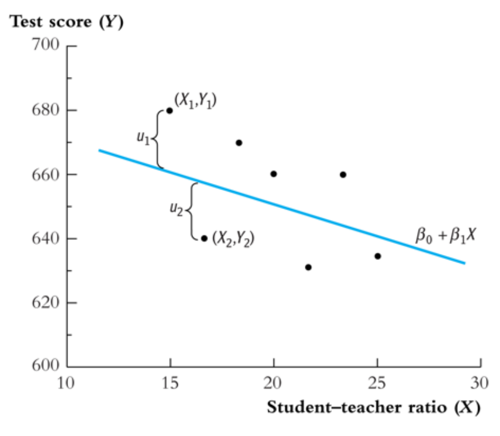


Figure 3: Observations on Y and X; the population regression line; and the regression error

Challenges ahead

The problems of statistical inference for linear regression are, at a general level, the same as for estimation of the mean or of the differences between two means. Statistical, or econometric, inference about the slope entails:

Challenges ahead

The problems of statistical inference for linear regression are, at a general level, the same as for estimation of the mean or of the differences between two means. Statistical, or econometric, inference about the slope entails:

- Estimation
  - How should we estimate β_0 and β_1? (answer: ordinary least squares). What are advantages and disadvantages of OLS?

Challenges ahead

The problems of statistical inference for linear regression are, at a general level, the same as for estimation of the mean or of the differences between two means. Statistical, or econometric, inference about the slope entails:

- Estimation
  - How should we estimate β_0 and β_1? (answer: ordinary least squares). What are advantages and disadvantages of OLS?
- Hypothesis testing
  - How to test if β_1 (or β_0) is zero?

Challenges ahead

The problems of statistical inference for linear regression are, at a general level, the same as for estimation of the mean or of the differences between two means. Statistical, or econometric, inference about the slope entails:

- Estimation
  - How should we estimate β_0 and β_1? (answer: ordinary least squares). What are advantages and disadvantages of OLS?
- Hypothesis testing
  - How to test if β_1 (or β_0) is zero?
- Confidence intervals
  - How to construct a confidence interval for β_1?

Estimation of β_0 and β_1

We already have a way of estimating the intercept and slope using the least squares method:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{s_{YX}}{s_X^2}$$

Implementing OLS

$\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained by using R

```
lm1 <- lm(testscore ~ str, data = CASchools)
lm1
```

```
##
## Call:
## lm(formula = testscore ~ str, data = CASchools)
##
## Coefficients:
## (Intercept)          str
##      698.93        -2.28
```

Implementing OLS

Often the output of the regression is written as:

$$\widehat{testscore} = 698.933 - 2.280 \times str$$

- Estimated slope: $\hat{\beta}_1 = -2.2798$

Implementing OLS

Often the output of the regression is written as:

$$\widehat{testscore} = 698.933 - 2.280 \times str$$

- Estimated slope: $\hat{\beta}_1 = -2.2798$
- Estimated intercept: $\hat{\beta}_0 = 698.9329$

Implementing OLS

Often the output of the regression is written as:

$$\widehat{testscore} = 698.933 - 2.280 \times str$$

- Estimated slope: $\hat{\beta}_1 = -2.2798$
- Estimated intercept: $\hat{\beta}_0 = 698.9329$
- $\widehat{TestScore}$ denotes the estimated regression line

Graphic representation

Interpretation of estimated intercept and slope

- Districts with one more student per teacher on average have test scores that are 2.28 points lower;
- That is,

$$\frac{\Delta TestScore}{\Delta STR} = -2.28$$

- The intercept (taken literally) means that, according to this estimated line, districts with zero students per teacher would have a (predicted) test score of 698.9
- This interpretation of the intercept makes no sense ??? it extrapolates the line outside the range of the data ??? here, the intercept is not economically meaningful