



Introduction to Statistics

Exercises 5

Use Excel to solve Problem 11

PROBLEM 1. Two music repositories on the have the same set of tunes. The monthly subscription rates for the two sites are as follows:

Site 1: Fixed subscription of €12 and €3 for downloaded album

Site 2: No fixed subscription and €9 per album download.

- If you download 2 albums per month on average, which site should you choose and what is the cost of getting it wrong?
- Assuming that you have resolved the previous question correctly, how many albums per month would you need to download to make your decision change?

PROBLEM 2. The average, unofficial monthly income (in euros) received by a member of parliament in a certain country is

$$Y = 1000 + 8000 X,$$

where X is equal to the number of builders that the politician meets every month.

- How do you interpret the coefficients (1000 and 8) of this regression equation?
- Given that this is a linear equation, what is the name given to the value 1000 and what is the name given to the value 8000.

PROBLEM 3. The formula for a regression equation based on a sample size of 25 observations is $Y = 2X + 9$. (a) What would be the predicted score for a person scoring 6 on X ? (b) If someone's predicted score was 14, what was this person's score on X ?

PROBLEM 4. In the previous example, suppose that all the scores on X were doubled. What would the new regression line relating X and Y be?

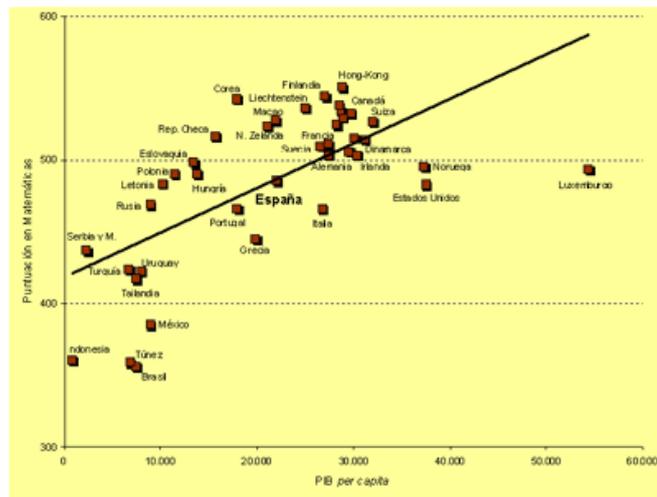
PROBLEM 5. Calculate the regression line of y given x for the data of Problem 2 of Exercises 6.

PROBLEM 6. With the same data, calculate the regression of x given y . Are the lines the same? Why or why not?

PROBLEM 7. Considering the data of Problem 1 of Exercise sheet 6, calculate the regression line of the grade in Journalism (y) given the Politics grade (x).

PROBLEM 8. In the previous problem, estimate the Journalism grade for a student with a Politics grade of 70. Do this again for a student with Politics grade 40. Are these predictions likely to be reliable?

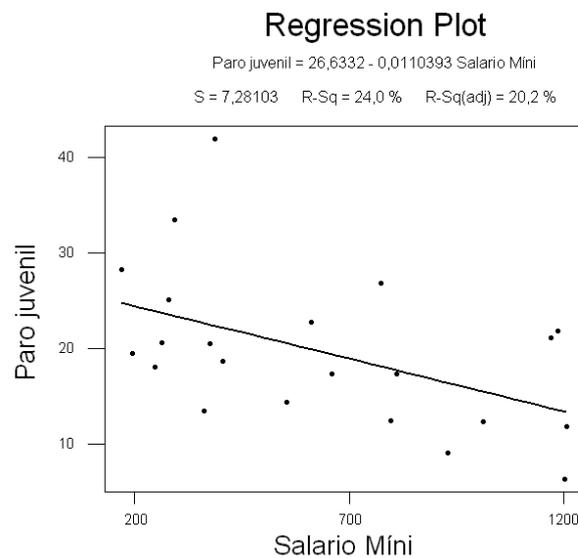
PROBLEM 9. The scatterplot shows the score in mathematics versus GDP per head in various European countries according to the PISA 2003 report.



Mark the correct response:

- a) The slope of the regression line is positive and the correlation between GDP and the score in mathematics is positive
- b) The slope and correlation are negative.
- c) The slope is positive and the correlation is negative.
- d) The slope is negative and the correlation is positive.

PROBLEM 10. The following diagram (Socialdemocracia.org) is a regression of the percentage of juvenile unemployment against minimum wage for 18 EU countries.



If the minimum wage in another EU country is €1000 per month, the predicted proportion of juvenile unemployment is approximately:

- a) -37,67%
- b) 37,67%
- c) 15,59%
- d) -15,59%

PROBLEM 11. The following table gives the grades obtained by 10 students in courses on Politics and Journalism respectively.

Políticas	Periodismo
75	82
80	78
93	86
65	72
87	91
71	80
98	95
68	72
84	89
77	74

- a) Draw a scatterplot for this data.
- b) Calculate the covariance s_{xy}
- c) Calculate the correlation coefficient r
- d) Interpret the result of c).

PROBLEM 12. If the correlation between two variables is equal to -1:

- a) The covariance is equal to -1.
- b) There is an approximate linear increasing relationship between the two variables.
- c) There is an exact, negative linear relationship.
- d) The variances of the two variables are the same.

PROBLEM 13. A simple of voters gave their evaluations for the president both in 2008 and 2011. The mean evaluations are 2 and 8, with standard deviations of 3 and 4 respectively. Supposing that the covariance is 4, the correlation is:

- a) 1/4.
- b) 4
- c) 1/3.
- d) 3.

PROBLEM 14. The table shows the Gross domestic product (GDP) per head in \$US in 2008 and 2009 for the G8 countries.

Country	GDP 2008 (x)	GDP 2009 (y)
Canada	42030	39217
France	45981	42091
Germany	44471	39442
Italy	38309	34955
Japan	38443	39573

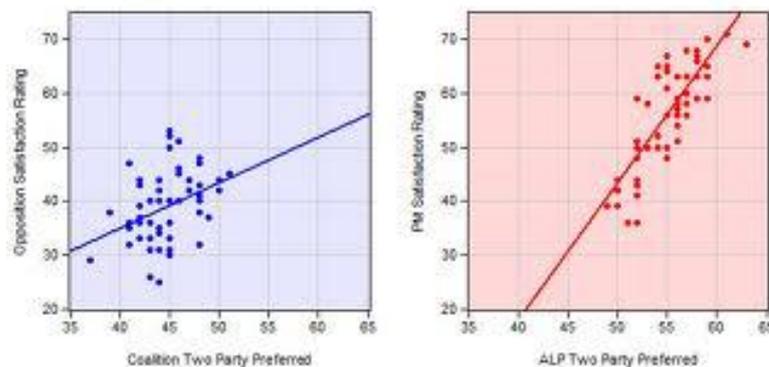
Russia	11339	8874
UK	43088	35728
USA	46716	46443

The covariance between the two variables is 116000000 and the correlation is 0.974. The Libyans prefer to measure GDP in Libyan dinars. The dollar-dinar exchange rate is (approx.) 1 dollar = 2 dinars.

If we measure GDP per head in Libyan dinars, which of the following is the correct solution?

- a) Neither the covariance nor the correlation change.
- b) The correlation is 0.2475 and the covariance does not change.
- c) The covariance is 464000000 and the correlation does not change.
- d) Both covariance and correlation increase to 4 times their previous values.

PROBLEM 15. The following graphs show the levels of satisfaction with the opposition leader (lhs) and the primeminister (rhs) as a function of the preferred vote.



Which of the following is correct?

- a) The correlation between the two variables is higher for the opposition leader.
- b) The two correlations are equal.
- c) The correlation is higher in the case of the primeminister.
- d) Both correlations are negative.