

## EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

Pharmacy/Biotechnology 1st year

Version A

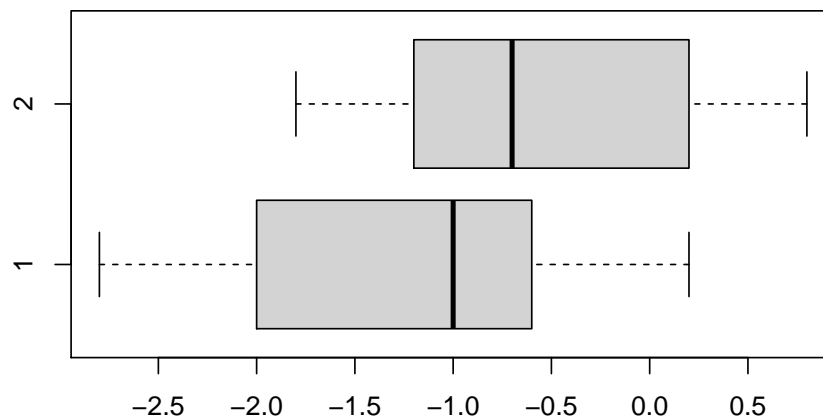
January, 18 2021

**Duration:** 1 hour.

- (5 pts.) 1. The table below contains the differences between the grades in the final school exam and the entrance exam in a sample of public high schools ( $X$ ) and private high schools ( $Y$ ):

Public schools	-1.2	-0.7	-0.4	-0.9	-1.6	0.5	0.2	-1.8	0.8
Private schools	-2.1	-0.5	-0.7	-1.9	0.2	-2.8	-1		

- (a) Which of the following box plots corresponds to each variable? Compare the central dispersion of the two variables according to the box plots. In which variable is smaller the median?



- (b) In which type of schools is more representative the mean of grades?  
 (c) In which type of schools is more symmetric the distribution of grades?  
 (d) In which type of schools is more peaked the distribution of grades?  
 (e) Which difference is relatively smaller,  $-0.5$  points in a public high school or  $-1$  points in a private high school?

Use the following sums for the computations:

Public:  $\sum x_i = -5.1$ ,  $\sum x_i^2 = 9.63$ ,  $\sum (x_i - \bar{x})^3 = 0.95$  and  $\sum (x_i - \bar{x})^4 = 8.76$ .

Private:  $\sum y_i = -8.8$ ,  $\sum y_i^2 = 17.64$ ,  $\sum (y_i - \bar{y})^3 = -0.82$  and  $\sum (y_i - \bar{y})^4 = 11.28$ .

### Solution

- (a) The box plot 1 corresponds to private schools and the box plot 2 to public schools. The central dispersion is pretty similar in both variables. The median is smaller in private schools.

- (b) Public schools:  $\bar{x} = -0.5667$ ,  $s^2 = 0.7489$ ,  $s = 0.8654$  y  $cv = 1.5271$ .  
 Private schools:  $\bar{y} = -1.2571$ ,  $s^2 = 0.9396$ ,  $s = 0.9693$  y  $cv = 0.7711$ .  
 Thus, the mean of the grade is more representative in private schools.
- (c)  $g_{1x} = 0.1626$  and  $g_{1y} = -0.1285$ . Thus, the distribution of grades in private schools is more symmetric as the coefficient of skewness is closer to 0.
- (d)  $g_{2x} = -1.2651$  and  $g_{2y} = -1.1748$ . Thus, the distribution of grades in private schools is more peaked.
- (e) Public schools:  $z(-0.5) = 0.077$ .  
 Private schools:  $z(-1) = 0.2653$ .  
 Thus, a difference of grades -0.5 in a public schools is relatively smaller than a difference of -1 in a private school.

- (4 pts.) 2. An auditor is studying the relationship between the salary and the number of absences of a hospital warden. The following table shows the salary in thousands of euros ( $X$ ) and the annual average of absences with that salary ( $Y$ ).

Salary	20.0	22.5	25	27.5	30.0	32.5	35.0	37.5	40.0
Absences	2.3	2.0	2	1.8	2.2	1.5	1.2	1.3	0.6

- (a) Compute the regression line that best explains the absences as a function of the salary.
- (b) What is the expected number of absences that will have a warden with a salary of 29000€? Is this prediction reliable?
- (c) How much will the number of absences increase or decrease for every increment of 1000€ in the salary?

Use the following sums for the computations:

$$\sum x_i = 270 \cdot 10^3 \text{€}, \sum y_i = 14.9 \text{ absences},$$

$$\sum x_i^2 = 8475 (10^3 \text{€})^2, \sum y_i^2 = 27.11 \text{ absences}^2,$$

$$\sum x_i y_i = 420 \cdot 10^3 \text{€ absences}.$$

### Solution

- (a)  $\bar{x} = 30 \cdot 10^3 \text{€}$ ,  $s_x^2 = 41.6667 (10^3 \text{€})^2$ ,  
 $\bar{y} = 1.6556 \text{ absences}$ ,  $s_y^2 = 0.2714 \text{ absences}^2$ ,  
 $s_{xy} = -3 \cdot 10^3 \text{€ absences}$ .  
 Regression line of absences on salary:  $y = 3.8156 - 0.072x$ .
- (b)  $y(29) = 1.7276 \text{ absences}$ .  
 $r^2 = 0.796$ , thus the model fits well as the coefficient of determination is not far from 1, but the sample size is too small to be reliable the prediction.
- (c) The number of absences will decrease 0.072 for every increment of 1000€ in the salary.

- (1 pts.) 3. In a regression study it is known that the regression line of  $Y$  on  $X$  is  $y + 2x - 10 = 0$  and the regression line of  $X$  on  $Y$  is  $y + 3x - 14 = 0$ .

- (a) Compute the means of  $X$  and  $Y$ .
- (b) Compute the linear correlation coefficient and interpret it.

### Solution

- (a)  $\bar{x} = 4$  and  $\bar{y} = 2$ .
- (b)  $r = -0.8165$ . The linear correlation coefficient is near -1 so there is a strong inverse relation between  $X$  and  $Y$ .