EXAM OF STATISTICS
(DESCRIPTIVE STATISTICS AND REGRESSION)

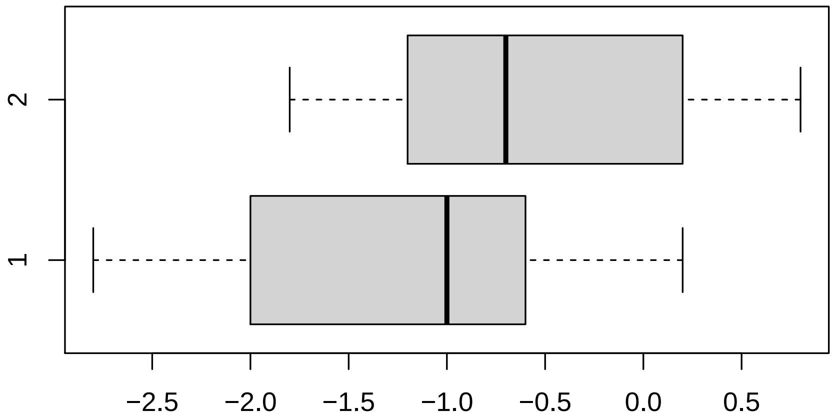
Pharmacy/Biotechnology 1st year Version A **January, 18 2021**

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? In which variable is there a higher central dispersion? 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 relativey 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) Padres: $\bar{x} = -0.5667$, $s^2 = 0.7489$, $s = 0.8654$ y $cv = -1.5271$.
Madres: $\bar{y} = -1.2571$, $s^2 = 0.9396$, $s = 0.9693$ y $cv = -0.7711$.
La estatura media es un poco más representativa en el grupo de las madres.

(b) Las vallas en la muestra de hijos son $f_1 = 10$ y $f_2 = 26$ por lo que no hay estaturas atípicas entre los hijos.

(c) $g_{1y} = -0.1285$ y $g_{2y} = -1.1748$. Como el coeficiente de asimetría y el de apuntamiento están dentro del intervalo de -2 a 2, podemos asumir que la muestra de estaturas de madres proviene de una población normal.

(d) Puntuación típica madres: $z(165) = -83.9682$.
Así pues, una madre de 165 es relativamente más alta que un hijo de 178 .

(e) La respresentatividad de las medias no cambiaría ya que tanto las medias como las desviaciones típicas estarían divididas por 100.

(3 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 per every increment of 1000€ in the salary?

Use the following sums for the computations:
$\sum x_i = 270\ 10^3$€, $\sum y_i = 14.9$ absences,
$\sum x_i^2 = 8475\ (10^3€)^2$, $\sum y_i^2 = 27.11$ absences^2,
$\sum x_i y_j = 420\ 10^3$€ absences.

Solution

(2 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
