EXAM OF STATISTICS
(PROBABILITY AND RANDOM VARIABLES)

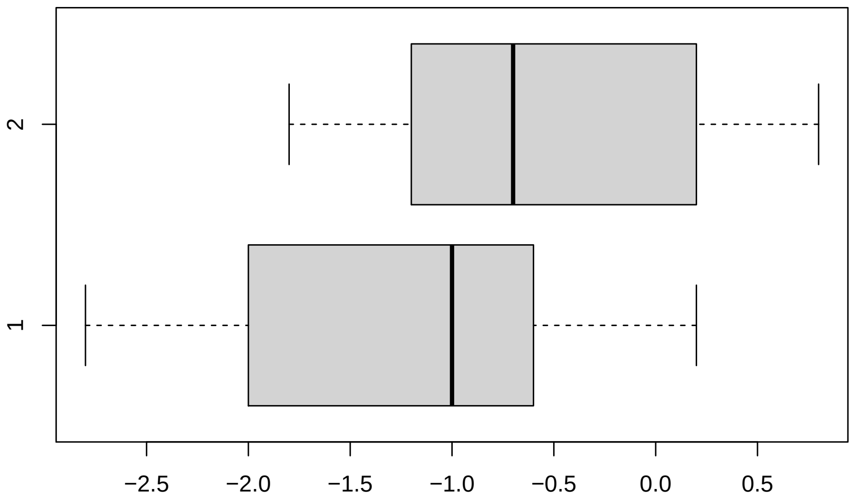
Pharmacy/Biotechnology 1st year Version A **November, 23 2020**

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?

Use the following sums for the computations:
Public: $\sum x_i = -5.1$, $\sum x_i^2 = 9.63$, $\sum(x_i - \bar{x})^3 = 0.95$ y $\sum(x_i - \bar{x})^4 = 8.76$.
Private: $\sum y_i = -8.8$, $\sum y_i^2 = 17.64$, $\sum(y_i - \bar{y})^3 = -0.82$ y $\sum(y_i - \bar{y})^4 = 11.28$.

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.

(5 pts.) 2. In a sample of 10 families with a son older than 20 it has been measured the height of the father (X), the mother (Y) and the son (Z) in centimetres, getting the following results:

$\sum x_i = 1774$ cm, $\sum y_i = 1630$ cm, $\sum z_i = 1795$ cm,
$\sum x_i^2 = 315300$ cm^2, $\sum y_i^2 = 266150$ cm^2, $\sum z_i^2 = 322737$ cm^2,
$\sum x_i y_j = 289364$ cm^2, $\sum x_i z_j = 318958$ cm^2, $\sum y_i z_j = 292757$ cm^2.

(a) On which height does the height of the son depend more linearly, the height of the father or the mother?

(b) Using the best linear regression model, predict the height of a son with a father 181 cm tall and a mother 163 cm tall.

(c) According to the linear model, how much the height of the son will increase for each centimetre that increases the height of the father? And for each centimetre that increases the height of the mother?

(d) How would the reliability of the prediction be if the heights were measured in inches? (An inch is 2.54 cm).

Solution

(a) $\bar{x} = 177.4$ cm, $s_x^2 = 59.24$ cm^2,
$\bar{y} = 163$ cm, $s_y^2 = 46$ cm^2,
$\bar{z} = 179.5$ cm, $s_z^2 = 53.45$ cm^2,
$s_{xz} = 52.5$ cm^2 and $s_{yz} = 17.2$ cm^2.
$r_{xz}^2 = 0.8705$ and $r_{yz}^2 = 0.1203$, thus the height of the son depends linearly more on the height of the father since the $r_{xz}^2 > r_{yz}^2$.

(b) Regression line of Z on X: $z = 22.2836 + 0.8862x$ and $z(181) = 182.6904$.

(c) The height of the son will increase 0.8862 cm per cm of the height of the father and 0.3739 cm per cm of the height of the mother.

(d) The reliability of the prediction will be the same, as after applying the same linear transformation to X and Z, the variances are multiplied by the square of the slope and the covariance is also multiplied by de square of the slope.
