EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

Pharmacy/Biotechnology 1st year Version A **November, 17 2020**

Duration: 1 hour.

(5 pts.) 1. 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:

Father's height	Mother's hegiht	Son's height
175	164	177
182	175	180
190	165	193
165	160	172
172	155	173
183	172	188
187	160	185
174	151	177
168	165	168
178	163	182

(a) In which sample is the mean more representative, in the sample of fathers or mothers?

(b) Are there outliers in the sample of sons?

(c) According to the shape of the distribution, can the sample of mothers come from a normal population?

(d) Who is higher in her/his sample, a mother 165 cm tall or a son 178 cm tall?

(e) If we had measured the heights in meters, how the representativity of the mean would have been affected?

Use the following sums for the computations:
Father's height: $\sum x_i = 1774$ cm, $\sum x_i^2 = 315300$ cm^2, $\sum(x_i - \bar{x})^3 = 210.48$ cm^3 y $\sum(x_i - \bar{x})^4 = 67596.27$ cm^4.
Mother's height: $\sum y_i = 1630$ cm, $\sum y_i^2 = 266150$ cm^2, $\sum(y_i - \bar{y})^3 = 180$ cm^3 y $\sum(y_i - \bar{y})^4 = 52324$ cm^4.
Son's height: $\sum z_i = 1795$ cm, $\sum z_i^2 = 322737$ cm^2, $\sum(z_i - \bar{z})^3 = 1008$ cm^3 y $\sum(z_i - \bar{z})^4 = 61906.62$ cm^4.

Solution

(a) Padres: $\bar{x} = 177.4$ cm, $s^2 = 59.24$ cm^2, $s = 7.6968$ cm y $cv = 0.0434$.
Madres: $\bar{y} = 163$ cm, $s^2 = 46$ cm^2, $s = 6.7823$ cm y $cv = 0.0416$.
La estatura media es un poco más representativa en el grupo de las madres.

(b) Las vallas en la muestra de hijos son $f_1 = 155$ cm y $f_2 = 203$ cm por lo que no hay estaturas atípicas entre los hijos.

(c) $g_{1y} = 0.0577$ y $g_{2y} = -0.5272$. 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: $z(165) = 0.2949$.
Puntuación típica hijos: $z(178) = -0.2052$.
Así pues, una madre de 165 cm es relativamente más alta que un hijo de 178 cm.

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

(5 pts.) 2. A variable that is usually used to diagnose the open angle glaucoma is the Bruch's membrane opening minimum rim width (X) of the retina, but it is known that it depends of the age (Y) and the membrane opening area (Z). These variables have been measured in 1000 patients with the following results:
$\sum x_i = 346337.03$ μm, $\sum y_i = 47212.1$ years, $\sum z_i = 2002.384$ mm^2,
$\sum x_i^2 = 123828243.48$ μm^2, $\sum y_i^2 = 2601264.99$ years^2, $\sum z_i^2 = 4175.89$ mm^4,
$\sum x_i y_j = 15855138.59$ μm·years, $\sum x_i z_j = 686623.65$ μm·mm^2, $\sum y_i z_j = 94144.37$ years·mm^2.

(a) Compute the regression lines of the Bruch's membrane opening minimum rim width on the age, and the Bruch's membrane opening minimum rim width on the membrane opening area.

(b) Acording to the linear model, how much the Bruch's membrane opening minimum rim width will increase or decrease for each aditional year of the patient?

(c) What percentage of the variabilitity of Bruch's membrane opening minimum rim width explains the two previous linear models?

(d) Using the best of the two previous linear models, predict the Bruch's membrane opening minimum rim width of a patient 60 years old with a membrane opening area of 2 mm^2. How is the reliability of this prediction?

Solution

(a) $\bar{x} = 346.337$ μm, $s_x^2 = 3878.9051$ μm^2,
$\bar{y} = 47.2121$ years, $s_y^2 = 372.2826$ years^2,
$\bar{z} = 2.0024$ mm^2, $s_z^2 = 0.1664$ mm^4,
$s_{xy} = -496.1599$ μm·years y $s_{xz} = -6.8761$ μm·mm^2.
Recta de regresión de X sobre Y: $x = 409.259 + -1.3328y$.
Recta de regresión de X sobre Z: $x = 429.1056 + -41.335z$.

(b) La distancia mínima al borde de la abertura de la membrana de Bruch disminuye 1.3328 μm por cada año más del paciente.

(c) $r_{xy}^2 = 0.1705$, de manera que la recta de regresión de X sobre Y explica el 17.05% de la variabilidad de la distancia mínima al borde de la abertura de la membrana de Bruch, y $r_{xz}^2 = 0.0733$, de manera que la recta de regresión de X sobre Z explica el 7.33% de la variabilidad de la distancia mínima al borde de la abertura de la membrana de Bruch.

(d) $x(60) = 329.2939$ μm.
