EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy | Version A | **June, 29 2021**

Duration: 1 hour.

(5 pts.) 1. The table below shows the number of credits obtained by the students of the first year of the physiotherapy grade.

$$48, 52, 60, 60, 24, 48, 48, 36, 39, 54, 54, 60, 12, 46$$

(a) Compute the median and the mode and interpret them.

(b) Draw the box and whiskers plot. Are there outliers in the sample?

(c) Can we assume that the sample comes from a normal population?

(d) If the the second year the mean of credits obtained is 102 and the standard deviation is 12.5, which year has a higher relative dispersion?

(e) Which number of credits is relatively higher, 50 in the first year, or 105 in the second year?

Use the following sums for the computations:
$\sum x_i = 641$, $\sum x_i^2 = 31901$, $\sum (x_i - \bar{x})^3 = -40158.06$ and $\sum (x_i - \bar{x})^4 = 1672652.57$.

Solution

(5 pts.) 2. A researcher is studying the relation between the obesity and the response to pain. The obesity is measured as the percentage over the ideal weight, and the response to pain as the nociceptive flexion pain threshold. The results of the study appears in the table below.

Obesity	89	90	77	30	51	75	62	45	90	20
Pain threshold	10	12	11.5	4.5	5.5	7	9	8	15	3

(a) According to the scatter plot, what model explains better the relation of the response to pain on the obesity, the linear or the logarithmic model?

(b) What percentage of the variability of the pain threshold explains the linear model? And the logarithmic model?

(c) According to the best regression model, what is the response to pain expected for a person with an obesity of 50%? Is this prediction reliable?

(d) According to the best regression model, what is the expected obesity for a person with a pain threshold of 10? Is this prediction more or less reliable than the previous one?

Use the following sums (X=Obesity and Y=Pain threshold):
$\sum x_i = 629$, $\sum \log(x_i) = 40.4121$, $\sum y_j = 85.5$, $\sum \log(y_j) = 20.4679$,
$\sum x_i^2 = 45445$, $\sum \log(x_i)^2 = 165.6795$, $\sum y_j^2 = 854.75$, $\sum \log(y_j)^2 = 44.0891$,
$\sum x_i y_j = 6124$, $\sum x_i \log(y_j) = 1390.1397$, $\sum \log(x_i) y_j = 360.0725$, $\sum \log(x_i)\log(y_j) = 84.8069$.

Solution

(a) $\overline{x} = 62.9\ \mu$g, $s_x^2 = 588.09\ \mu\text{g}^2$.
$\bar{y} = 8.55$ bacteria, $s_y^2 = 12.3725$ bacteria^2.
$s_{xy} = 74.605\ \mu$g·bacteria.
Linear coefficient of determination $r^2 = 0.765$.
$\overline{\log(y)} = 2.0468$ log(bacteria), $s^2_{\log(y)} = 0.2196$ log(bacteria)^2.
$s_{x\log(y)} = 10.2709\ \mu$g· log(bacteria).
Exponential coefficient of determination $r^2 = 0.817$.
Thus, the exponential model explains better the number of residual bacteria as a function of the antibiotic dose because the exponential coef. of determination is greater.

(b) Exponential regression model: $y = e^{0.9483+0.0175x}$.
Prediction: $y(3.5) = 2.7439$ bacteria.
Although the coef. of determination is close to 1, the this prediction is not reliable because the sample size is very small.

(c) $b_{yx} = 0.1269$, therefore the number of bacteria decreases 0.1269 per each μg more of antibiotic.