EXAM OF STATISTICS
(DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy — Version A — **March, 3 2020**

Duration: 1 hour.

(5 pts.) 1. In a study on the reconstruction of the anterior cruciate ligament (ACL), the postoperative recovery time was evaluated depending on whether the patients underwent a meniscal suture or not. The table below contains the results:

Postoperative months	Patients without suture	Patients with suture
$0-1$	11	5
$1-2$	18	13
$2-3$	10	6
$3-4$	8	10
$4-5$	6	9
$5-6$	4	1
$6-7$	1	0
$7-8$	2	8

(a) Plot the ogive for the sample of patients without meniscal suture.

(b) Are there outliers in the number of postoperative moths of patients without meniscal suture?

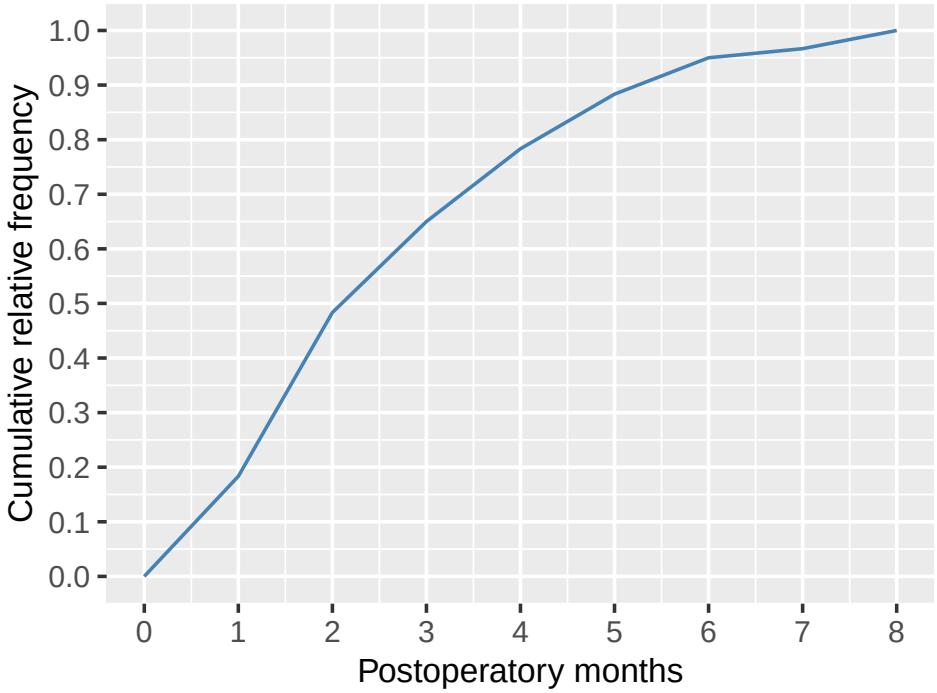
(c) In which of the two groups is more representative the mean of the postoperative months?

(d) Can we assume that the sample of patients with meniscal suture comes from a normal population?

(e) What value is relatively greater, 5 months for a patient without meniscal suture, or 6 for a patient with meniscal suture?

Use the following sums for the computations:
Without meniscal suture: $\sum x_i n_i = 156$ months, $\sum x_i^2 n_i = 601$ months^2, $\sum (x_i - \bar{x})^3 n_i = 313.32$ months^3 y $\sum (x_i - \bar{x})^4 n_i = 1990.94$ months^4.
With meniscal suture: $\sum y_i n_i = 178$ months, $\sum y_i^2 n_i = 853$ months^2, $\sum (y_i - \bar{y})^3 n_i = 340.26$ months^3 y $\sum (y_i - \bar{y})^4 n_i = 2788.04$ months^4.

Solution



(a)

(b) $Q_1 = 1.222$ months, $Q_3 = 3.7502$ months, $IQR = 2.5282$ month, $f_1 = -2.5703$ and $f_2 = 7.5425$. Since the upper limit of the last interval is greater than the upper fence, there could be outliers in the sample.

(c) Without suture: $\bar{x} = 2.6$ months, $s^2 = 3.2567$ months^2, $s = 1.8046$ months and $cv = 0.6941$.
With suture: $\bar{y} = 3.4231$ months, $s^2 = 4.6864$ months^2, $s = 2.1648$ months and $cv = 0.6324$.
Thus, the mean of the sample with suture is more representative since its coefficient of variation is smaller.

(d) $g_1 = 0.645$ and $g_2 = -0.5587$. Since both coefficients are between -2 and 2, we can assume that the sample comes from a normal population.

(e) Without suture: $z(5) = 1.3299$.
With suture: $z(6) = 1.1904$.
Thus, 5 months in the sample without suture is relatively greater, since its standard score is greater.

(5 pts.) 2. The table below shows the evolution of the number of coronavirus infections since the virus was detected.

Days	25	29	32	35	38	40	43	45	47
Infections	282	846	2798	7818	14557	20630	31481	37558	43103

(a) Which regression model, the linear or the exponential, is better to predict the number of coronavirus infections with time?

(b) According to the best of the two previous regression models, how many infections will there be after 100 days? Is this prediction reliable?

(c) If he number of deaths from coranvirus is linearly related to the number of infections, with a linear coefficient of determination 0.99, how much will increase or decrease the number of infections for each death more if it is know that the number of deaths increases 0.022 for each infection more.

Use the following sums for the computations:
$\sum x_i = 334$ days, $\sum \log(x_i) = 32.3517$ log(days), $\sum y_j = 159073$ infections, $\sum \log(y_j) = 80.3657$

log(infections),
$\sum x_i^2 = 12842$ days^2, $\sum \log(x_i)^2 = 116.6525$ log(days)^2, $\sum y_j^2 = 4966773651$ infections^2, $\sum \log(y_j)^2 = 743.3009$ log(infections)^2,
$\sum x_i y_j = 6842750$ days·infections, $\sum x_i \log(y_j) = 3086.808$ days· log(infections), $\sum \log(x_i) y_j = 597633.103$ log(days)infections, $\sum \log(x_i) \log(y_j) = 291.8911$ log(days) log(infections).

Solution

(a) $\bar{x} = 37.1111$ days, $s_x^2 = 49.6543$ days^2.
$\bar{y} = 17674.7778$ infections, $s_y^2 = 239465969.5062$ infections^2.
$\overline{\log(y)} = 8.9295$ log(infections), $s_{\log(y)}^2 = 2.8526$ log(infections)^2.
$s_{xy} = 104374.9136$ days·infections, $s_{x\log(y)} = 11.5941$ days· log($infections$).
Linear coefficient of determination: $r^2 = 0.9162$
Exponential coefficient of determination: $r^2 = 0.949$
Thus, the exponential model is better to predict the number of infections since its coefficient of determination is greater.

(b) Exponential regression model: $y = e^{0.2642+0.2335x}$.
Prediction: $y(100) = 18002893169.2954$ infections. The prediction is not reliable because 100 days is far away from the range of days in the sample.

(c) According to the regression coefficient, ther will be 44.689 infections more for each death more.
