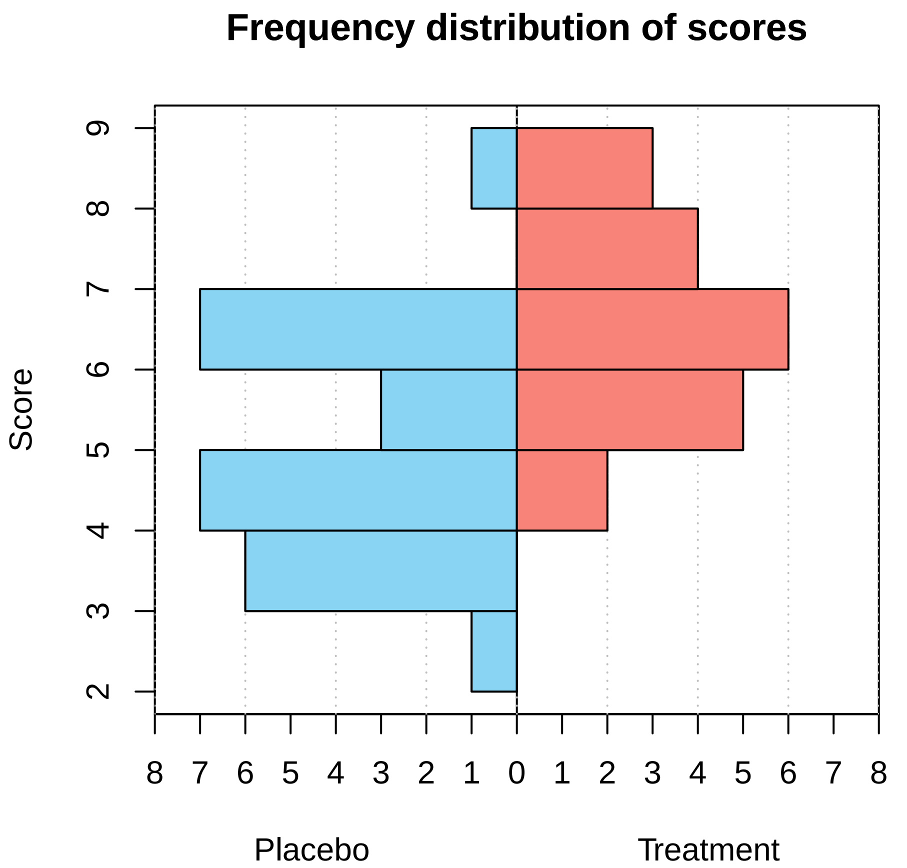
EXAM OF STATISTICS
(DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy Version A **June, 07 2021**

Duration: 1 hour.

(5 pts.) 1. To study the effectiveness of a new treatment for the polymyalgia rheumatica a sample of patients with polymyalgia was drawn and they were divided into two groups. The first group received the new treatment while the second one received a placebo. After a year following the treatment they filled out a survey. The chart below shows the distribution of the survey score of the two groups of patients (the greater the score the better the treatment).



(a) Construct the frequency table of the scores for the placebo group and plot the ogive.

(b) Compute the interquartile range of the scores for the placebo group.

(c) Are there outliers in the placebo group?

(d) In which group the score mean represents better?

(e) Which distribution is more normal regarding the kurtosis?

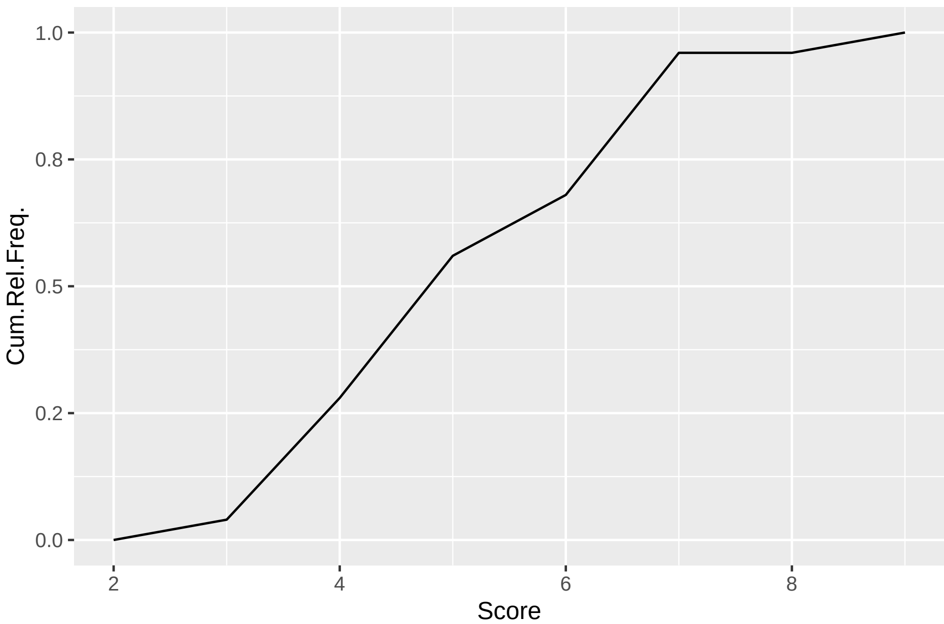
(f) Which score is relatively better, a score of 5 in the placebo group or a score of 6 in the treatment group?

Use the following sums for the computations:
Placebo: $\sum x_i = 125.5$, $\sum x_i^2 = 680.25$, $\sum(x_i - \bar{x})^3 = 27.11$ and $\sum(x_i - \bar{x})^4 = 253.27$.
Treatment: $\sum x_i = 131$, $\sum x_i^2 = 887$, $\sum(x_i - \bar{x})^3 = 2.66$ and $\sum(x_i - \bar{x})^4 = 88.03$.

Solution

(a)

Score	n_i	f_i	N_i	F_i
$[2, 3]$	1	0.04	1	0.0
$(3, 4]$	6	0.24	7	0.3
$(4, 5]$	7	0.28	14	0.6
$(5, 6]$	3	0.12	17	0.7
$(6, 7]$	7	0.28	24	1.0
$(7, 8]$	0	0.00	24	1.0
$(8, 9]$	1	0.04	25	1.0



(b) $Q_1 = 3.875$, $Q_3 = 6.25$ and $IQR = 2.375$.

(c) $f_1 = 0.3125$ and $f_2 = 9.8125$. Thus, there are no outliers in the placebo sample because all the values fall between the fences.

(d) Placebo: $\bar{x} = 5.02$, $s^2 = 2.0096$, $s = 1.4176$ and $cv = 0.2824$.
Treatment: $\bar{x} = 6.55$, $s^2 = 1.4475$, $s = 1.2031$ and $cv = 0.1837$.

(e) Placebo: $g_2 = -0.4914$. Treatment: $g_2 = -0.8992$. Thus, the distribution of the placebo group is more normal as the coef. of kurtosis is closer to 0.

(f) Standard score for the placebo: $z(5) = -0.0141$
Standard score for the treatment: $z(6) = -0.4571$
As the standard score of 5 in the placebo group is greater than the standard score of 6 in the treatment group, a score of 5 in the placebo group is better.

(5 pts.) 2. We have applied different doses of an antibiotic to a culture of bacteria. The table below shows the number of residual bacteria corresponding to the different doses.

Dose (μg)	0.2	0.7	1	1.5	2	2.4	2.8	3
Bacteria	40	32	28	20	18	15	12	11

(a) Which regression model explains better the number of residual bacteria as a function of the antibiotic dose, the linear or the exponential?

(b) Use the best of the two previous regression models to predict the number of residual bacteria for an antibiotic dose of 3.5 μg. Is this prediction reliable?

(c) According to the linear regression model, what is the expected decrease in the number of residual bacteria per each μg more of antibiotic?

Use the following sums for the computations (X=Antibiotic dose and Y=Number of bacteria):
$\sum x_i = 13.6$ μg, $\sum \log(x_i) = 2.1362$ $\log(\mu\text{g})$, $\sum y_j = 176$ bacteria, $\sum \log(y_j) = 23.9638$ log(bacteria),
$\sum x_i^2 = 30.38$ μg^2, $\sum \log(x_i)^2 = 6.3959$ $\log(\mu\text{g})^2$, $\sum y_j^2 = 4622$ bacteria^2, $\sum \log(y_j)^2 = 73.3096$ log(bacteria)^2,
$\sum x_i y_j = 227$ μg·bacteria, $\sum x_i \log(y_j) = 37.4211$ μg· log(bacteria), $\sum \log(x_i) y_j = -17.633$ $\log(\mu\text{g})$bacteria,
$\sum \log(x_i)\log(y_j) = 3.6086$ $\log(\mu\text{g})$ log(bacteria).

Solution

(a) $\overline{x} = 1.7$ μg, $s_x^2 = 0.9075$ μg^2.
$\bar{y} = 22$ bacteria, $s_y^2 = 93.75$ bacteria^2.
$s_{xy} = -9.025$ μg·bacteria.
Linear coefficient of determination $r^2 = 0.9574$.
$\overline{\log(y)} = 2.9955$ log(bacteria), $s^2_{\log(y)} = 0.1908$ log(bacteria)^2.
$s_{x\log(y)} = -0.4147$ μg· log(bacteria).
Exponential coefficient of determination $r^2 = 0.9928$.
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: $y = e^{3.7723+-0.4569x}$.
Prediction: $y(3.5) = 8.7845$ bacteria.
Although the coef. of determination is close to 1, the this prediction is not reliable because the sample size is very small.

(c) $b_{yx} = -9.9449$, therefore the number of bacteria decreases -9.9449 per each μg more of antibiotic.
