EXAM OF STATISTICS
(DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy — Version A — **June, 18 2019**

Duration: 1 hour and 15 minutes.

(5 pts.) 1. To see if the confinement due to COVID-19 has influenced the performance of a course, the number of failed subjects of each student in the current course and in the previous year course has been counted, obtaining the table below.

Failed subjects	Previous year course	Current course
0	7	8
1	15	12
2	11	8
3	5	7
4	4	3
5	2	2
6	1	2
8	0	1

Se pide:

(a) Draw the box plots of the failed subjects in the current and the previous year courses and compare them.

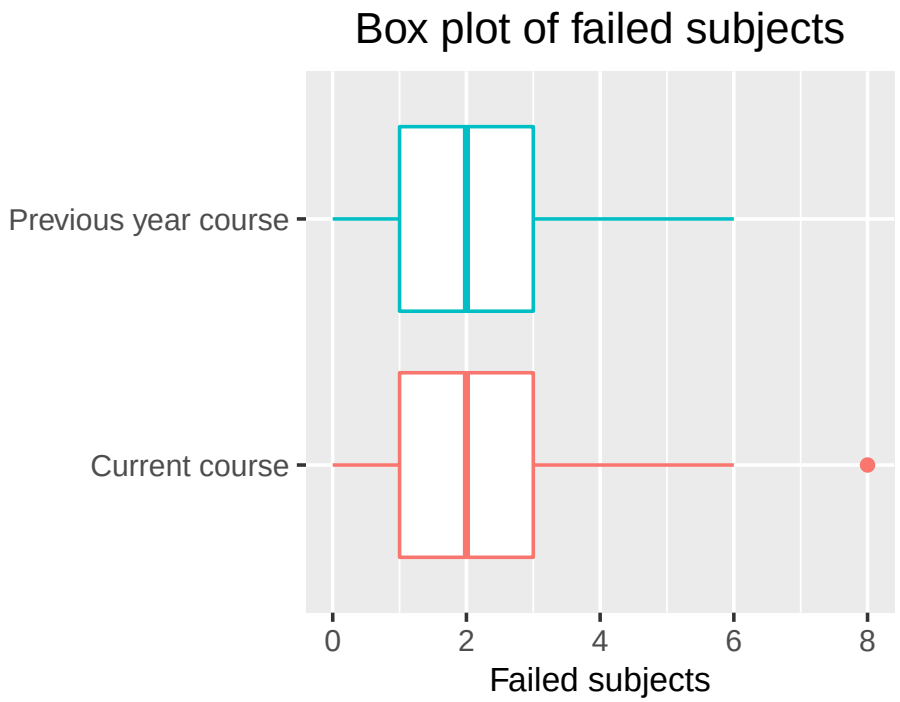
(b) Can we assume that both samples come from a normal population?

(c) In which sample the mean is more representative?

(d) Which number of failed subjects is greater, 7 in the current course or 6 in the previous year course?

Use the following sums for the computations:
Previous year course: $\sum x_i n_i = 84$, $\sum x_i^2 n_i = 254$, $\sum (x_i - \bar{x})^3 n_i = 122.99$ y $\sum (x_i - \bar{x})^4 n_i = 669.21$.
Current course: $\sum y_i n_i = 91$, $\sum y_i^2 n_i = 341$, $\sum (y_i - \bar{y})^3 n_i = 301.16$ y $\sum (y_i - \bar{y})^4 n_i = 2012.88$.

Solution



(a) Both distributions are pretty similar. The central dispersion is the same and both are right skewed. The only difference is that there is an outlier in the current year distribution.

(b) Previous year course: $\bar{x} = 1.8667$, $s^2 = 2.16$, $s = 1.4697$, $g_1 = 0.8609$ and $g_2 = 0.1874$.
Current course: $\bar{y} = 2.1163$, $s^2 = 3.4516$, $s = 1.8578$, $g_1 = 1.0922$ and $g_2 = 0.9292$.
As the coefficients of skewness and kurtosis are between -2 and 2, we can assume that both distributions come from a normal distribution.

(c) Previous year course: $cv = 0.7873$.
Current year: $cv = 0.8779$.
Thus, the mean is more representative in the previous year course, since the coefficient of variation is smaller.

(d) Previous year course: $z(6) = 2.8124$.
Current course: $z(7) = 2.6287$.
Thus, 7 failed subjects in the current course is relatively less than 6 in the previous year course, since the standard score is smaller.

(5 pts.) 2. A study tries to develop a new technique for detecting a certain antibody. For this, a piezoelectric immunosensor is used, which allows to measure the change in the signal in Hz by varying the concentration of the antibody (μg/ml). The table below presents the data collected.

Concentration (μg/ml)	5	8	20	35	50	80	110
Signal (Hz)	50	70	100	150	170	190	200

Se pide:

(a) Compute the logarithmic model of the change in the signal on the concentration of the antibodies.

(b) It was observed that at a concentration of 100 μg/ml the change in signal tends to stabilize. Predict the value of the signal corresponding to such concentration using the logarithmic model.

(c) Predict the antibody concentration that corresponds to a change in the signal of 120 using the exponential model.

Use the following sums for the computations (X=Concentration and Y=Signal):
$\sum x_i = 308$ Hz, $\sum \log(x_i) = 23.2345$ log(Hz), $\sum y_j = 930$ μg/ml, $\sum \log(y_j) = 33.4575$ log(μg/ml),
$\sum x_i^2 = 22714$ Hz^2, $\sum \log(x_i)^2 = 85.1299$ log(Hz)^2, $\sum y_j^2 = 144900$ μg/ml^2, $\sum \log(y_j)^2 = 161.6475$ log(μg/ml)^2,
$\sum x_i y_j = 53760$ Hz·μg/ml, $\sum x_i \log(y_j) = 1580.3905$ Hz· log(μg/ml), $\sum \log(x_i) y_j = 3496.6333$ log(Hz)μg/ml,
$\sum \log(x_i) \log(y_j) = 114.7297$ log(Hz) log(μg/ml).

Solution

(a) $\overline{\log(x)} = 3.3192$ log(μg/ml), $s^2_{\log(x)} = 1.1442$ log(μg/ml)^2. $\bar{y} = 132.8571$ Hz, $s_y^2 = 3048.9796$ Hz^2.
$s_{\log(x)y} = 58.5379$ log(μg/ml)Hz.
Logarithmic regression model: $y = -36.9501 + 51.1589 \log(x)$.

(b) Prediction: $y(100) = 5078.9396$ Hz.

(c) Exponential regression model: $y = e^{0.7685 + 0.0192y}$.
Prediction: $y(120) = 21.5929$ μg/ml.
