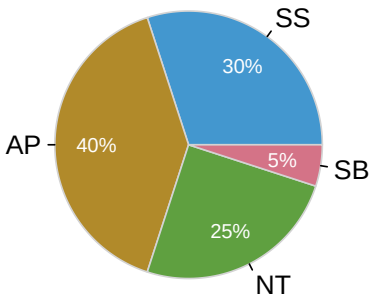
EXAM OF STATISTICS
(DESCRIPTIVE STATISTICS AND REGRESSION)

2nd Physiotherapy — Version A — **June, 24 2022**

Duration: 1 hour.

(5 pts.) 1. The chart below shows the percentage of grades in a Statistic course with 60 students.



(a) Plot the ogive of the score, assuming the following correspondence between grades and scores

Grade	Score
SS	$[0, 5)$
AP	$[5, 7)$
NT	$[7, 9)$
SB	$[9, 10]$

(b) Compute the median and interpret it.

(c) How many students got a score greater than 8?

(d) Study the dispersion of the distribution.

(e) Study the skewness of the distribution. Is it normal?

(f) If we apply the transformation $y = 10x + 5$ to the scores, how changes the representativeness of the mean. And the skewness?

Use the following sums for the computations (X = Score):
$\sum x_i n_i = 337.5$, $\sum x_i^2 n_i = 2207.25$, $\sum (x_i - \bar{x})^3 n_i = -172.55$ and $\sum (x_i - \bar{x})^4 n_i = 2870.75$.

Solution

(5 pts.) 2. In a Statistics course with 40 students it has been measured the grade in the final exam and the number of hours dedicated to prepare it. The sums below summarize the results (X=Hours and Y=Grade):

$\sum x_i = 759$ hours, $\sum \log(x_i) = 106.7396$ log(hours), $\sum y_j = 172.6$ points, $\sum \log(y_j) = 49.1831$ log(points),

$\sum x_i^2 = 19439$ hours^2, $\sum \log(x_i)^2 = 315.247$ log(hours)^2, $\sum y_j^2 = 1019.58$ points^2, $\sum \log(y_j)^2 = 83.1748$ log(points)^2,
$\sum x_i y_j = 4377$ hours·points, $\sum x_i \log(y_j) = 1237.3465$ hours·log(points), $\sum \log(x_i) y_j = 535.6777$ log(hours)·points, $\sum \log(x_i)\log(y_j) = 154.5956$ log(hours)·log(points).

(a) Which regression models, linear, exponential or logarithmic, explains better the relation between the grade and the hours of study?

(b) Use the best model to predict the grade of a student who has dedicated 25 hours to prepare the final exam of Statistics. Is this prediction reliable?

(c) According to the linear model, how much increases the grade for each additional hour of study?

(d) It is known that the grade in the final exam of Physiotherapy does not depend on the hours of study in Statistics, and the average grade in the Physiotherapy exam was 7.5. What is the expected grade in the final exam of Physiotherapy of the student who has dedicated 25 hours to prepare the final exam of Statistics.

Solution

(a) $\overline{x} = 18.975$ weeks, $s_x^2 = 125.9244$ weeks^2.
$\bar{y} = 4.315$ Kg, $s_y^2 = 6.8703$ Kg^2.
$s_{xy} = 27.5479$ weeks

.

Kg.
Regression coefficient of Y on X: $b_{yx} = 0.2188$ Kg/week. The grip strength increases 0.2188 Kg per week.

(b) $\overline{\ln(x)} = 2.6685$ ln(weeks), $s^2_{\ln(x)} = 0.7603$ ln(weeks)^2 and $s_{\ln(x)y} = 1.8774$ ln(weeks)Kg.
Logarithmic regression model of Y on X: $y = -2.274 + 2.4692\ln(x)$.
Predictions: $y(5) = 1.7$ Kg and $y(25) = 5.674$ Kg.
Logarithmic coefficient of determination: $r^2 = 0.6747$. The predictions are not reliable because the sample size is small.
Linear coefficient of determination: $r^2 = 0.8772$.
As the linear coefficient of determination is less than the logarithmic one, the predictions with the logarithmic model are more reliable.

(c) Exponential regression model of X on Y: $x = e^{1.4893+0.2733y}$.
Prediction: $x(25) = 4109.2011$ Weeks.

(d) As $r^2 = 0.6747$, the exponential models explains 67.47% of the variabilidty of the weeks.
