EXAM OF STATISTICS (DESCRIPTIVE STATISTICS AND REGRESSION)

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

Duration: 1 hour and 15 minutes.

(5 pts.) 1. In a course there are 150 students, of which 50 are working students and the other 100 non-working students. The table below shows the frequency distribution of the grade in an exam of these two groups:

Grade	Num non-working students	Num working students
$0-2$	8	2
$2-4$	15	9
$4-6$	25	19
$6-8$	38	11
$8-10$	14	9

(a) Compute the percentage of students that passed the exam (a grade 5 or above) in both groups, working and non-working students.

(b) In which group is there a higher relative dispersion of the grade with respect to the mean?

(c) Which grade distribution is more asymmetric, the distribution of working students, or the non-working students one?

(d) To apply for a scholarship to go abroad, the grade must be transformed applying the linear transformation Y = 0.5 + X * 1.45. Compute the mean of Y for the two groups. How changes the asymmetry of the two groups?

(e) Which grade is relatively higher, 6 in the working students group, or 7 in the non-working students group?

Use the following sums for the computations:
Non-working students: $\sum x_i n_i = 570$, $\sum x_i^2 n_i = 3764$, $\sum (x_i - \bar{x})^3 n_i = -547.8$ y $\sum (x_i - \bar{x})^4 n_i = 6475.73$.
Working students: $\sum y_i n_i = 282$, $\sum y_i^2 n_i = 1826$, $\sum (y_i - \bar{y})^3 n_i = -1.31$ y $\sum (y_i - \bar{y})^4 n_i = 2552.14$.

Solution

(a) 66.5% of non-working stsudents passed and 59% of working students passed.

(b) Non-working students: $\bar{x} = 5.7$, $s^2 = 5.15$, $s = 2.2694$ and $cv = 0.3981$.
Working students: $\bar{y} = 5.64$, $s^2 = 4.7104$, $s = 2.1703$ and $cv = 0.3848$.
The sample of non-working students has a slightly higher relative dispersion with respect to the mean as the coefficient of variation is greater.

(c) Non-working students: $g_1 = -0.4687$.
Working students: $g_1 = -0.0026$.
Thus, the sample of non-working students is more assymetric as the coefficient os skewness is further from 0.

(d) Non-working students: $\bar{y} = 8.765$.
Working students: $\bar{x} = 8.678$.
The coefficient of skewness does not change as the slope of the linear transformation is positive.

(e) Non-working students: $z(7) = 0.5728$.
Working students: $z(6) = 0.1659$.
Thus, a 7 in the sample of non-working students is relatively higher than than a 6 in the sample of working students, as its standard score is greater.

(5 pts.) 2. The effect of a doping substance on the response time to a given stimulus was analyzed in a group of patients. The same amount of substance was administered in successive doses, from 10 to 80 mg, to all the patients. The table below shows the average response time to the stimulus, expressed in hundredths of a second:

Dose (mg)	10	20	30	40	50	60	70	80
Response time (10^{-2} s)	28	46	62	81	100	132	195	302

(a) According to the linear regression model, how much will the response time increase or decrease for each mg we increase the dose?

(b) Based on the exponential model, what will be the expected response time for a 75 mg dose?

(c) If a response time greater than one second is considered dangerous for health, from what level should the administration of the doping substance be regulated, or even prohibited, according to the logarithmic model?

Use the following sums for the computations:
$\sum x_i = 360$ mg, $\sum \log(x_i) = 29.0253$ log(mg), $\sum y_j = 946\ 10^{-2}$ s, $\sum \log(y_j) = 36.1538 \log(10^{-2}\text{ s})$, $\sum x_i^2 = 20400$ mg^2, $\sum \log(x_i)^2 = 108.7717$ log(mg)^2, $\sum y_j^2 = 169958\ 10^{-2}$ s^2, $\sum \log(y_j)^2 = 167.5694 \log(10^{-2}\text{ s})^2$,
$\sum x_i y_j = 57030$ mg$\cdot 10^{-2}$ s, $\sum x_i \log(y_j) = 1758.6576$ mg$\cdot \log(10^{-2}\text{ s})$, $\sum \log(x_i) y_j = 3795.4339$ log(mg)10^{-2} s, $\sum \log(x_i)\log(y_j) = 134.823$ log(mg) $\log(10^{-2}\text{ s})$.

Solution

(a) $\bar{x} = 45$ mg, $s_x^2 = 525$ mg^2.
$\bar{y} = 118.25\ 10^{-2}$ s, $s_y^2 = 7261.6875\ 10^{-4}$ s^2.
$s_{xy} = 1807.5$ mg$\cdot 10^{-2}$ s.
$b_{yx} = 3.4429\ 10^{-2}$ s/mg.
Therefore, the response time increases 3.4429 hundredths of a second for each mg the dose is increased.

(b) $\overline{\log(y)} = 4.5192 \log(10^{-2}\text{ s})$, $s^2_{\log(y)} = 0.5227 \log(10^{-2}\text{ s})^2$.
$s_{x\log(y)} = 16.4669$ mg$\cdot \log(10^{-2}\text{ s})$.
Exponential regression model: $y = e^{3.1078+0.0314x}$.
Prediction: $y(75) = 235.1434\ 10^{-2}$ s.
Exponential coefficient of determination: $r^2 = 0.988$
Thus, the exponential model fits almost perfectly to the cloud of points of the scatter plot, but the sample is too small to get reliable predictions.

(c) Logarithmic regression model: $x = -97.3603 + 31.501 \ln(y)$.
Prediction: $x(100) = 47.7072$ mg.
