EXAM OF STATISTICS

2nd Physiotherapy Version A **June, 28 2018**

Duration: 2 hours.

(2.5 pts.) 1. The time required by a drug A to be effective has been measured in a sample of 150 patients. The table below summarize the results.

Response time	Patients
$(0, 5]$	5
$(5, 10]$	15
$(10, 15]$	32
$(15, 20]$	36
$(20, 30]$	42
$(30, 60]$	20

(a) Are there outliers in the sample? Justify the answer.

(b) What is the minimum time for the 20% of patients with highest response time?

(c) What is the average response time? Is the mean representative? Justify the answer.

(d) Study the kurtosis of the sample.

(e) If we take another sample of patients with mean 18 min and standard deviation 15 min, in which group is greater a response time of 25 min?

Use the following sums for the computations: $\sum x_i = 3105$ min, $\sum x_i^2 = 83650$ min^2, $\sum(x_i - \bar{x})^3 = 2277448.69$ min^3 y $\sum(x_i - \bar{x})^4 = 82723757.55$ min^4.

Solution

(a) $Q_1 = 12.7344$ min, $Q_2 = 59$ min, $IQR = 13.099$ min, $f_1 = -6.9141$ min and $f_2 = 45.4818$ min. Therefore there are outliers in the sample since the upper limit of the last interval is 60.

(b) $P_{80} = 27.619$ min.

(c) $\bar{x} = 20.7$ min, $s^2 = 129.1767$ min^2, $s = 11.3656$ min and $cv = 0.5491$. The mean is not very representative since the $cv > 0.5$.

(d) $g_2 = 0.2523$, thus the distribution is a little bit leptokurtic.

(e) For the first sample $z_i = 0.3783$ and for the second one $z_i = 0.4667$, thus a time of 25 min is relatively greater in the second sample.

(1.5 pts.) 2. A 40% of a population of athletes have a very athletic mother and a 30% a very athletic father. If 50% of athletes in this population have some very athletic progenitor:

(a) Compute the probability that a radom athlete of this population have a very athletic mother if he or she has a very athletic father.

(b) Compute the probability that a radom athlete of this population have a very athletic father if he or she has a non very athletic mother.

(c) ¿Are the events corresponding to having a very athletic father and having a very athletic mother independents? Justify the answer.

Solution

Let M the event of having a very athletic mother and F the event of having a very athletic father.

(a) $P(M|F) = 0.6667$

(b) $P(F|\overline{M}) = 0.1667$

(c) No, the events are dependent since $P(M) = 0.4 \neq 0.6667 = P(M|F)$.

(2.5 pts.) 3. We have measured the average number of weekly hours of study X and the score in a clinic entrance test Y of 8 candidates, getting the following results:

$\sum x_i = 15.9$, $\sum \log(x_i) = 3.852$, $\sum y_j = 41.5$, $\sum \log(y_j) = 11.511$,
$\sum x_i^2 = 42.23$, $\sum \log(x_i)^2 = 5.559$, $\sum y_j^2 = 274.25$, $\sum \log(y_j)^2 = 20.742$,
$\sum x_i y_j = 103.3$, $\sum x_i \log(y_j) = 28.047$, $\sum \log(x_i) y_j = 32.616$.

(a) Compute the equations of the logarithmic and exponential regression models of the score as a function of the average number of hours of study.

(b) Use the best of the previous models to predict the score for somebody that study 3.2 hours a week.

Solution

(a) $\bar{x} = 1.9875$ hours, $s_x^2 = 1.3286$ hours^2.
$\overline{\log(x)} = 0.4815$ log(hours), $s_{\log(x)}^2 = 0.463$ log(hours)^2.
$\bar{y} = 5.1875$ points, $s_y^2 = 7.3711$ points^2.
$\overline{\log(y)} = 1.4389$ log(points), $s_{\log(y)}^2 = 0.5224$ log(points)^2.
$s_{x\log(y)} = 0.6461$, $s_{\log(x)y} = 1.5792$
Logarithmic regression model: $y = 3.5453 + 3.4106\log(x)$
Exponential regression model: $y = e^{0.4723+0.4863x}$

(b) Logarithmic coefficient of determination: $r^2 = 0.7307$
Exponential coefficient of determination: $r^2 = 0.6015$
Therefore, the best regression model to predict is the exponential.
Prediction: $y(3.2) = 14.4592$.

(1.5 pts.) 4. The average number of injuries in an international tennis tournament is 2.

(a) Compute the probability that in an international tennis tournament there are more than 2 injuries.

(b) If a tennis circuit has 6 international tournaments, what is the probability that there are no injuries in some of them?

Solution

(a) Let X be the number of injuries in a tournament, then $X \sim P(2)$ and $P(X > 2) = 0.3233$.

(b) Let Y be the number of tournaments in the tennis circuit with no injuries, then $Y \sim B(6, 0.1353)$ and $P(Y > 0) = 0.5821$.

(2 pts.) 5. A diagnostic test to determine doping of athletes returns a positive outcome when the concentration of a substance in blood is greater than 4 μg/ml. If the distribution of the substance concentration in doped athletes follows a normal distribution model with mean 4.5 μg/ml and standard deviation 0.2 μg/ml, and in non-doped athletes is normally distributed with mean 3 μg/ml and standard deviation 0.3 μg/ml,

(a) what is the sensitivity and specificity of the test?

(b) If there is a 0.1% of doped athletes in a competition, what are the positive and the negative predicted values? Interpret them.

Solution

Let $X \sim N(4.5, 0.2)$ be the substance concentration in doped athletes and $Y \sim N(3, 0.3)$ the substance concentration in non-doped athletes.

(a) Sensitivity $P(+|D) = P(X > 4) = 0.9938$ and specificity $P(-|\overline{D}) = P(Y < 4) = 0.9996$.

(b) PPV $P(D|+) = 0.9961$ and NPV $P(\overline{D}|-) = 0.9993$.
