EXAM OF STATISTICS (PROBABILITY AND RANDOM VARIABLES)

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

Duration: 1 hour and 15 minutes.

(2 pts.) 1. A study tries to determine the effectiveness of an occupational risk prevention program in jobs that require to be sit a lot of hours. A sample of 500 individuals between 40 and 50 years that spent more than 5 hours sitting was drawn. Half of the individuals followed the prevention program (treatment group) and the other half not (control group). After 5 years it was observed that 12 individuals suffered spinal injuries in the group following the prevention program while 32 individuals suffered spinal injuries in the other group. In the following 5 years it was observed that 21 individuals suffered spinal injuries in the group following the prevention program while 48 individuals suffered spinal injuries in the other group.

(a) Compute the cumulative incidence of spinal injuries in the total sample after 5 years and after 10 years.

(b) Compute the absolute risk of suffering spinal injuries in 10 years in the treatment and control groups.

(c) Compute the relative risk of suffering spinal injuries in 10 years in the treatment group compared to the control group. Interpret it.

(d) Compute the odds ratio of suffering spinal injuries in 10 years in the treatment group compared to the control group. Interpret it.

(e) Which statistics, the relative risk or the odds ratio, is more suitable in this study? Justify the answer.

Solution

Let D be the event of suffering spinal injuries.

(a) Cumulative incidence after 5 years: $R(D) = 0.088$.
Cumulative incidence after 10 years: $R(D) = 0.226$.

(b) Risk in the treatment group: $R_T(D) = 0.132$.
Risk in the control group: $R_C(D) = 0.32$.

(c) $RR(D) = 0.4125$. Thus, the risk of suffering spinal injuries is less than half following the prevention program.

(d) $OR(D) = 0.3232$. Thus, the odd of suffering spinal injuries is less than one third following the prevention program.

(e) Since the study is prospective and we can estimate the prevalence of D, both statistics are suitable, but relative risk is easier to interpret.

(3 pts.) 2. The table below shows the results of a study to evaluate the usefulness of a reactive strip to diagnose an urinary infection.

Outcome	Infection	No infection
Positive	60	80
Negative	10	200

(a) Compute the sensitivity and the specificity of the test.

(b) Compute the positive and the negative predictive values. Is this test better to confirm or to rule out the infection?

(c) If another study has determined that the true prevalence of the infection is 2%, how does this affect to the predictive values?

Solution

Let D be the event corresponding to suffering the urinary infection and $+$ and $-$ the events corresponding to get a positive and negative outcome in the test respectively.

(a) Sensitivity $= 0.8571$ and Specificity $= 0.7143$.

(b) $PPV = 0.4286$ and $NPV = 0.9524$. Since the $PPV < NPV$ the test is better to rule out the infection.

(c) $PPV = 0.0577$ and $NPV = 0.9959$. The positive predictive value descreases a lot while the negative predictive value increases al little bit.

(5 pts.) 3. The time required to recover from an injury follows a normal distribution with variance 64 days. It is also known that 10% of people with this injury require more than 80 days to recover.

(a) What is the expected time required to recover from the injury?
Remark: Use $\mu = 70$ for the next part if you do not know how to compute it.

(b) What percentage of individuals will require between 60 and 75 days to recover?

(c) If we draw a random sample of 12 individuals with this injury, what is the probability of having between 9 and 11 individuals, both included, requiring less than 80 days to recover?

(d) If we draw a random sample of 500 individuals with this injury, what is the probability of having less than 4 requiring a time above the 99th percentile to recover?

Solution

Let X be the time required to recover from the injury. Then $X \sim N(\mu, 8)$.

(a) $\mu = 69.7476$ days.

(b) $P(60 < X < 75) = 0.6327$.

(c) Let Y be the number of individuals with the injury requiring less than 80 days to recover in a sample of 12. Then $Y \sim B(12, 0.9)$ and $P(9 \leq Y \leq 11) = 0.6919$.

(d) Let Z be the number of individuals with the injury requiring a time above the 99th percentile to recover in a sample of 500. Then $Z \sim B(500, 0.01) \approx P(5)$ and $P(Z \leq 4) = 0.265$.
