Second midterm exam
November 30th, 2022

Time:
75 minutes

- **You are not allowed to use any documentation apart from the formula sheet you have received, and the Z(0,1) table.**
- **Use 4 decimal digits in all calculations and results.**

1. A survey has been carried out in a population using a samle of 200 people, resulting in 72 smokers.

   a) (1 point) Calculate the proportion of smokers in the sample and the standard deviation for the sample.

   $$\hat{p} = \frac{72}{200} = 0{,}36$$

   Since we have a sample with 200 items we can approximate the binomial distribution B(200;0,36) by a normal distribution with standard deviation $\hat{s} = \sqrt{\hat{p}\hat{q}/n}$

   $$\hat{s} = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{\frac{0{,}36\,.\,0{,}64}{200}} = 0{,}0339$$

   b) (1 point) Find a 90% Confidence Interval for the population proportion of smokers.

   $$p \,\epsilon \left\{ \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \right\} = \left\{ 0{,}36 \pm 1{,}645\sqrt{\frac{0{,}36.0{,}64}{200}} \right\}$$

   $\{0{,}3042; 0{,}4158\}$ *is the interval requested*

   c) (1 point) In case we want to increase the acuracy of the estimation by reducing the lenght (amplitud*2)of the confidence interval by half, what size sample should we use assuming that the rest of the sample values do not change?

   $$z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = \frac{1}{2}\, z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n^*}}$$

   $$n^* = 4n = 4.200 = 800\, people$$

   .

   d) (1 point) Is there enough evidence to affirm that at least 1/3 of the population is smoker with a significance level α=0.10?

   We perform a hypothesis test with the following assumptions:

   H_0 –- p = 1/3 = 0,3333
   H_1 –- $p \neq 0{,}3333$
   Which is a bilateral test, so we can check if 0,3333 is contained in the confidence interval calculated previously.

$0{,}3333 \in \{0{,}3042; 0{,}4158\}$*so we can say, with a level of significance of 10%, that at least 1/3 of the population smokes.*

e) (1 point) Calculate the p-value for the hypothesis test in the previous question.
The test statistic is:

$$z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0{,}36-0{,}3333}{\sqrt{\frac{0{,}3333\,.\,0{,}6667}{200}}} = 0{,}801$$

$$P - value = P(Z > z_0) = P(Z > 0{,}80) = 1 - P(Z < 0{,}80) = 1 - 0{,}7881 = 0{,}2119$$

f) (1 point) In another sample with 400 people, 150 were smokers. Is this sample giving the same information about the proportion of smokers as the previous one?. Use the same level of significance.

To check it we can perform the following hypothesis test:

H₀ –- $p_1 = p_2$
H₁ –- $p_1 \neq p_2$
Which is a bilateral test whose statistic is:

$$z_0 = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\widehat{p_0}\widehat{q_0}(1/n_1 + 1/n_2)}} = \frac{\frac{72}{200} - \frac{150}{400}}{\sqrt{0{,}37\,.\,0{,}63\,(1/200 + 1/400)}} = \frac{-0{,}015}{0{,}01870} = -0.8022$$

$$\widehat{p_0} = \frac{n_1\widehat{p_1} + n_2\widehat{p_2}}{n_1 + n_2} = \frac{200\,.\,0{,}36 + 400\,.\,0{,}375}{200+400} = 0{,}37 \quad \widehat{q_0} = 0{,}63$$

α = 0,1, therefore $|z_{\alpha/2}| = 1{,}645 \;\; so\; z_0 > |z_{\alpha/2}|\; and\; we\; cannot\; reject\; H_0$
The information given by both samples is the sample, the proportion of smokers is at least 1/3.

2. Random variable X denotes the monthly expenses of an enterprise, and it has a density function defined as follows:

$$f(\Theta, x) = \Theta x^{\Theta - 1} \; with\; \Theta > 0 \; and \; 0 < x < 1$$

Simple random samples of size three are obtained, and the following are proposed as estimators for parameter Θ:

$$\hat{\Theta}_1 = \bar{X}$$

$$\hat{\Theta}_2 = \frac{X_3 - 2X_1 + 4X_2}{6}$$

a) (3 points) Calculate the bias for each estimator

$$Bias(\hat{\Theta}) = E(\hat{\Theta}) - \Theta$$

$$E(\hat{\Theta}_1) = E\left[\frac{X_1 + X_2 + 3X_3}{3}\right] = \frac{1}{3}\,(3\mu) = \mu$$

$$\mu = \int_0^1 x.f(x,\Theta)dx = \int_0^1 x\Theta x^{\Theta-1}dx = \int_0^1 \Theta x^{\Theta}dx = \frac{\Theta}{\Theta+1}$$

$$Bias(\hat{\Theta}_1) = \frac{\Theta}{\Theta+1} - \Theta = -\frac{\Theta^2}{\Theta+1}$$

$$E(\hat{\Theta}_2) = E\left[\frac{X_3 - 2X_1 + 4X_2}{6}\right] = \frac{1}{6}\,(3\mu) = \frac{1}{2}\mu$$

$$Bias(\hat{\Theta}_2) = \frac{1}{2}\,\frac{\Theta}{\Theta+1} - \Theta = -\frac{2\Theta^2+\Theta}{2(\Theta+1)}$$

b) (0,5 points) If a sample obtained is (0,7; 0,1; 0,3) calculate the point estimations?

$$\hat{\Theta}_1 = \left[\frac{0{,}7 + 0{,}1 + 0{,}3}{3}\right] = 0{,}367$$

$$\hat{\Theta}_2 = \left[\frac{0{,}3 - 2.0{,}7 + 4.0{,}1}{6}\right] = -0{,}117 \ \textit{which is not possible because } \Theta > 0$$

c) (0,5 points)Which are the estimated functions for the previous estimations?

Only for $\hat{\Theta}_1$since $\hat{\Theta}_2$ is not valid --- f(0,367,x) = 0,367$x^{0{,}367-1} = 0{,}367x^{-0{,}633}$