Statistics 1 for Economics

Question 1 (2 + 5 + 5 = 12)

The owner of an ice cream stand wants to investigate the relationship between the price p of an ice cream (in €) and the number n of ice creams sold per week. She varies the price during a period of five weeks and finds the following numbers:

week	1	2	3	4	5	mean	st.dev.
price of an ice cream	2.10	2.70	2.80	4.20	2.20	2.80	0.840
numbers of ice creams sold	749	749	885	620	742	749	93.790

a. Is n a nominal, ordinal, discrete or continuous variable?

discrete

b. Calculate the covariance between p and n. (2 decimals)

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(p_i - \bar{p})(n_i - \bar{n})$$

$$= \frac{1}{5-1}\big(0 + 0 + 0 + (4.20 - 2.80)(620 - 749) + (2.20 - 2.80)(742 - 749)\big)$$

$$= -44.10$$

c. After the five weeks of experimental pricing, the owner assumes a linear relation between p and n, and finds the regression line $\hat{n} = 924.00 + b_1 \cdot p$. The owner decides to set the price at € 2.90. Determine the expected number of ice creams sold at this price. (2 decimals)

$$\text{substitute } (\bar{p} = 2.80, \bar{n} = 749) \text{ into } \hat{n} = 924.00 - b_1 \cdot p$$

$$\Longrightarrow \quad b_1 = \frac{924.00 - \bar{n}}{\bar{p}} = \frac{924.00 - 749}{2.80} = 62.50$$

$$\hat{n} = 924.00 - b_1 \cdot p_{new} = 924.00 - 62.50 \cdot 2.90 = 742.87$$

Question 2 (5 + 5 + 7 + 6 + 7 = 30)

49 countries participate in the African Friendship Games. Each country is represented by 2 players. For the games on the first day, 20 players are selected at random, where it is possible that both players from the same country are chosen to play on the same day.

a. Calculate the probability that Tanzania is selected to play on the first day? (3 decimals)

$$X: \text{ number of players from Tanzania} \sim Hyp(n = 20, N = 2 \cdot 49 = 98, k = 2)$$

$$P(X \geq 1) = P(X = 1) + P(X = 2) = \frac{\binom{2}{1}\binom{96}{19}}{\binom{98}{20}} + \frac{\binom{2}{2}\binom{96}{18}}{\binom{98}{20}} \approx 0.040 + 0.328 = 0.368$$

or

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{2}{0}\binom{96}{20}}{\binom{98}{20}} \approx 1 - 0.632 = 0.368$$

Each player will play three matches. Players from the same country will never face each other in a match.

b. Tanzanian players have a probability of 0.6 to win from other players in a match. Calculate the probability that at least one (of the two Tanzanian players) wins three matches? (3 decimals)

For one Tanzanian player to win three matches: $P(\text{win, win, win}) = P(\text{win})^3 = 0.6^3 = 0.216$

For at least one Tanzanian player to win three matches:

$$Y: \text{ number of Tanzanian players who win 3 matches} \sim Bin(n = 2, p = 0.216)$$

$$P(Y \geq 1) = P(Y = 1) + P(X = 2)$$
$$= P(\text{only first wins 3}) + P(\text{only second wins 3}) + P(\text{both win 3})$$
$$= 0.216 \cdot (1 - 0.216) + (1 - 0.216) \cdot 0.216 + 0.216 \cdot 0.216 = 0.385$$

or

$$P(Y \geq 1) = 1 - P(Y = 0)$$
$$= 1 - P(\text{both do not wins 3})$$
$$= 1 - (1 - 0.216) \cdot (1 - 0.216) = 0.385$$

c. 12% of the Tanzanian population have been following the popular African Friendship Games on television. A sample of 50 Tanzanians is selected randomly. Determine the probability that at most 9 of them have been following the Friendship Games on television. (4 decimals)

$$Q: \text{number of Tanzanians following the Games} \sim Bin(n = 50, p = 0.12)$$

1. Conditions

$$np = 50 \cdot 0.12 = 6 \geq 5$$
$$n(1 - p) = 50 \cdot 0.88 = 44 \geq 5$$

2. Translation

$$\mu = np = 50 \cdot 0.12 = 6$$
$$\sigma^2 = np(1 - p) = 50 \cdot 0.12 \cdot 0.88 = 5.28$$
$$Q \sim N(\mu = 6, \sigma^2 = 5.28)$$

3. Continuity correction

$$P(Q \leq 9) \approx P(Q < 9.5) = P\left(\frac{Q^{CC} - \mu}{\sigma} < \frac{9.5 - 6}{\sqrt{5.28}}\right) \approx P(Z < 1.52) = 0.9357$$

d. In the neighboring country Kenya, 5 Kenyans in a random sample of 50 have been following the Friendship Games. Based on this sample, determine a 95%-confidence interval for the percentage of Kenyans that are following the Friendship Games. (4 decimals)

$$p = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.1 \pm 1.96 \sqrt{\frac{0.1(1-0.1)}{50}} = 0.1 \pm 0.0832$$

$$\Longrightarrow \; 0.0168 < p < 0.1832$$

Conditions: random sample, $n\hat{p} \geq 5$ and $n(1-\hat{p}) = 50 \cdot (1-0.1) = 45 \geq 5$

e. Although the previous estimation turned out to be quite unprecise, it does convince us that the population proportion of Kenyans following the Friendship Games can never be more than 30%. Calculate the required minimal sample size in order to have a 95%-confidence interval with a maximum width of 4%?

$$\text{margin} = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq B = \frac{1}{2} \cdot 4\% = 0.02$$

$$\Longrightarrow \; n \geq \left(\frac{z_{\frac{\alpha}{2}}}{B}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.02}\right)^2 0.2(1-0.2) = 1536.64 \;\Longrightarrow\; n \geq 1537$$

Question 3 (19 + 5 + 5 = 29)

A marketing manager of a store wants to attract potential clients. She is particularly interested in the amounts of money that potential clients are spending at the stores of her competitors. These amounts may be assumed to be normally distributed.

From a marketing perspective, she will need a different strategy to attract the potential clients, if the amount of money that potential clients spend at the competitors varies more than the amount that her current clients spend at her store. The standard deviation for the amounts that her current clients spend is 9.7 euros.

She took random sample of 31 potential clients, and found a sample standard deviation of 11.80 euros.

a. Perform the appropriate hypothesis test for the standard deviation. Use $\alpha = 0.01$.

Conditions and assumptions:

- random sample
- values of the population are normally distributed

Hypotheses

$$H_0 : \sigma^2 = 9.7^2 \text{ versus } H_1 : \sigma^2 > 9.7^2$$

Test statistic and its distribution

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(31-1)S^2}{9.7^2} \sim \chi^2[\text{df} = n - 1 = 30]$$

Rejection region

$$\alpha = 0.01 \implies \chi^2 \geq \chi^2_{crit} = \chi^2_{0.01;30} = 50.892$$

Sample outcome

$$\chi^2_{obs} = \frac{(31-1)\cdot 11.80^2}{9.7^2} = 44.396$$

Confrontation and decision

$$\chi^2_{obs} < \chi^2_{crit} \text{ so do not reject } H_0$$

Conclusion

Given the significance level of 1%, we have not found sufficient evidence to infer that the standard deviation of the spent amounts of money by potential clients is greater than 9.7.

b. Estimate the *p-value* of the test. (3 decimals)
(If you could not find the value of the test statistic in 3a. then use 43.210)

$$p\text{-}value = P(\chi^2 > \chi^2_{obs}) = P(\chi^2 > 44.396)$$

$$\text{df} = 30 : \; P(\chi^2 > 43.773) = 0.05 \text{ and } P(\chi^2 > 46.979) = 0.025 \text{ so } 0.025 < p\text{-}value < 0.05$$

c. For which sample standard deviations would we conclude, in this test, that the population standard deviation is larger than 9.7 ? (3 decimals)

We should reject H_0 when:

$$\chi^2_{obs} \geq \chi^2_{crit} = 50.892 \implies \frac{(31-1)s^2}{9.7^2} \geq 50.892 \implies s \geq s_{crit} = \sqrt{\frac{50.892 \cdot 9.7^2}{(31-1)}} = 12.634$$

Question 4 (19)

A researcher investigates the monthly turnover of a company. The monthly turnovers of this company do not follow a normal distribution. The researcher randomly selected eleven monthly turnovers (× 1000 €):

12 8 11 20 16 15 21 9 18 5 11

Use the $p\text{-}value$ to test, with a significance level of 10%, whether the median differs from 11 000 €.

- X : turnover (× 1000 €)
- M : median turnover (× 1000 €)
- p^+ : proportion of turnovers above 11 000 €.

Conditions and assumptions

- random sample
- $n_{\text{eff}} < 10$ so we cannot use Z (approximation by normal)

Hypotheses

$$H_0 : M = 11 \text{ versus } H_1 : M \neq 11 \quad \text{or} \quad p^+ = 0.5 \text{ versus } p^+ \neq 0.5$$

Test statistic and its distribution

X^+: number of turnovers above 11 $\sim Bin(n_{\text{eff}} = 9, p = 0.5)$

("neutral" observations of 11 are excluded)

Rejection region

$$\alpha = 0.12$$

Sample outcome

(sample: 12+ 8– ~~11~~ 20+ 16+ 15+ 21+ 9– 18+ 5– ~~11~~)

$$x^+ = 7 \implies p\text{-}value = 2 \cdot P(X^+ \geq x^+) = 2 \cdot P(X^+ \geq 7) = 2 \cdot \big(1 - P(X^+ \leq 6)\big)$$
$$\approx 2 \cdot (1 - 0.910) = 0.180$$

Confrontation and decision

$$p\text{-}value > \alpha \implies \text{do not reject } H_0$$

Conclusion:

Given a significance level of 12%, there is insufficient evidence to infer that median turnover is different from 11 000 €.

Question 5 (max. 10 pt, min. 0 pt, per incorrect answer: −2 pt)

- If we change the research hypothesis in a sign test about the median, from $H_1 : p^+ > 0.5$ to $H_1 : p^- < 0.5$, then we will come to a different conclusion about the population median.

FALSE

(because $p^+ + p^- = 1$)

- If we replace the answers of an ordinal 7-point-scale (from “disagree very strongly” via “neutral” to “agree very strongly”) by the numbers {-3, -2, -1, 0, 1, 2, 3} then we can do a valid Z-test or T-test for the average opinion.

FALSE

(the 7-point-scale is ordinal, not quantitative)

- The sample mean $\bar{X}$ is not a consistent estimator.

FALSE

(because $V[\bar{X}] = \frac{\sigma^2}{n}$)

- The probability of a Type II Error will increase when increasing the significance level α.

FALSE

(because $\beta = P(\text{not reject} \mid \dots)$ and $\alpha = P(reject \mid \dots)$)

- The power of a test can be increased by increasing the sample size n.

TRUE

(if n increases, then the variance of the distribution decreases, and so the overlap between hypothetical and real distribution decreases)

- The power of a test is smaller when the observed sample mean $\bar{x}$ is closer to μ_0 (the value of the population mean under the null-hypothesis)

FALSE

(the power of a test is independent of the observed sample mean $\bar{x}$)