Mine Çağlar November 23, 2018

Math 202: Statistics for Social Sciences

Fall 2018 EXAM 2

Calculator OK, 90 min.

Instructions: There are ~~six~~ five parts to this exam I-V~~I~~. Please inspect the exam and make sure you have all five pages of questions. Do all your work on these pages. If you use the back of a page, make sure to indicate that.

Remember: ***You must show your work to get proper credit.*** ***In hypothesis testing questions, show all steps of the test and state your conclusion in plain English.***

Academic Honesty Code: Koç University Academic Honesty Code stipulates that "copying from others or providing answers or information, written or oral, to others is cheating." By taking this exam, you are assuming full responsibility for observing the Academic Honesty Code.

NAME: ______________________________

Part I:	/20
Part II:	/20
Part III:	/25
Part IV:	/20
Part V:	/25
Total:	/110

10 points bonus

Formulas:

Confidence interval (CI) : $\overline{X} \mp z\, SE$ or $\overline{X} \mp t\, SE$ or $\hat{p} \mp z\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$$(\hat{p}_1 - \hat{p}_2) \mp z\, SE_{2-sample} \quad \text{or} \quad (\overline{X}_1 - \overline{X}_2) \mp z\, SE_{2-sample} \quad \text{or} \quad (\overline{X}_1 - \overline{X}_2) \mp t\, SE_{2-sample}$$

Standard Error (SE) : $SE_{2-sample} = \sqrt{(SE_1)^2 + (SE_2)^2}$, $SE = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ for percentages,

and $SE = \dfrac{SD}{\sqrt{n}}$ *for averages.*

Part I. **(20 points)** A survey organization takes a simple random sample of 1000 people from the residents of a large city. In the sample, 337 people are found to have moved from elsewhere to the city over the past 5 years.

1. (8 points) Find a 99% confidence interval for the percentage of people who have moved to the city over the past 5 years.

$\hat{p} = \frac{337}{1000} = 0,337 = 33,7\%$

$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,337 \cdot 0,663}{1000}} = 0,0149 \simeq 0,015$

z-table: $99\% \rightarrow z = 2,60$

CI: $\hat{p} \pm zSE = 0,337 \pm (2,60) 0,015 \Rightarrow [0,2980, 0,3760]$

or $[29,8\%, 37,6\%]$

2. (2 points) Fill in the blanks with i) sample or ii) city :

The 99% confidence interval found in Question 1 is for the percentage of residents in the

___city___ who have moved to the city over the past 5 years.

3. (10 points) Test the claim that the percentage of people who moved to the city over the past 5 years is more than 30%, at the level of significance $\alpha = 1\%$.

$H_0: p = 0,30$

$H_a: p > 0,30$

$SE = \sqrt{\frac{0.30 \cdot 0.70}{1000}} = 0,01449$

$z = \frac{0.337 - 0.30}{0,01449} \simeq 2.55$

$z_\alpha = 2,35$

$2.55 > 2,35 \Rightarrow$ Reject H_0

$z = 2,55 \rightarrow 99,20\%$

$\frac{100 - 98.92}{2} = 0.54\% = P\text{-value}$

$0.54\% < 1\%$

$\Rightarrow$ Reject H_0.

Yes, the percentage of people who moved to city over the past 5 years is more than 30%..

Part II. (20 points) According to TÜİK, Türkiye İstatistik Kurumu, the mean age of men at the time of their first marriage is 27.7 in 2017 in Turkey. In Istanbul, the mean age for first marriage is 29.2 as found from a sample of 35 men, selected randomly from those who have married at least once. The standard deviation of the sample is 3.4.

1. (11 points) Is the mean age for first marriage of men higher in Istanbul? Report the P-value.

$H_0: \mu = 27.7$

$H_a: \mu > 27.7$

$SE = \frac{SD}{\sqrt{n}} = \frac{3.4}{\sqrt{35}} = 0.5747$

$z = \frac{29.2 - 27.7}{0.5747} = 2.61$

$\downarrow$

99.07%

$\boxed{\frac{100\% - 99.07\%}{2} = 0.465\% = \text{p-value} < 5\% \Rightarrow \text{Reject } H_0}$

$\alpha = 5\%$

$100\% - (2)5\% = 90\%$

$\downarrow$

$z_\alpha = 1.65$

$2.61 > 1.65 \Rightarrow$ Reject Ho

Yes, the mean age in İstanbul is significantly higher.

2. (9 points) Construct a 95% confidence interval for the mean age of first marriage of men in Istanbul.

$SE = \frac{SD}{\sqrt{n}} = \frac{3.4}{\sqrt{35}} = 0.5747$

$95\% \rightarrow z = 1.95$

$CI = \mu \pm z\,SE$

$= 29.2 \pm (1.95)\,0.5747$

$= [28.1, 30.3]$

Part III. (25 points) A randomized, controlled and double-blind study investigates the relationship between the primary school childrens' scores on intelligence tests and their family backgrounds. 12 students from the rural area has scored 25.4 on the average on the intelligence test with a standard deviation of 3.2, while an average of 28.7 is obtained by 10 students from the urban area with a standard deviation of 4.9.

1. (14 points) Is there a significant difference between the intelligence scores of the students from rural and urban areas? Report the P-value.

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$SE_1 = \frac{3,2}{\sqrt{12}} \simeq 0,92$

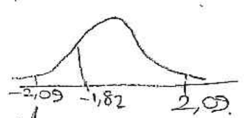
$SE_2 = \frac{4,9}{\sqrt{10}} \cong 1,55$

$SE_{diff} = \sqrt{0,92^2 + 1.55^2} \cong 1,80$

$t = \frac{25.4 - 28.7}{1,80} = \frac{-3,3}{1,80} = -1,83$

$d.f = 12 - 1 + 10 - 1 = 20$

$\Rightarrow t = 2,09$ with $2,5\%$



- Since $-2,09 < -1,83 < 2,09$, do not reject H_0!

There is no significant evidence that students from rural and urban areas differ.

Also $1,72 < 1,83 < 2,09 \Rightarrow \%2.5 < \frac{P\text{-value}}{2} < \%5 \Rightarrow \%5 < P\text{-value} < \%10$

2. (8 points) Construct a 90% confidence interval for the difference in the intelligence scores.

$d.f = 20$

$t_{0.05} = 1,72$

$CI = (\bar{X}_1 - \bar{X}_2) \mp t\,SE$

$= (25,4 - 28,7) \mp (1,72)\,1,80$

$= [-6,3960, -0,2040] \cong [-6.40, -0.20]$

3. (3 points) If 50 samples are drawn from rural area, and 50 samples are drawn from urban area and they are used to form 50 confidence intervals as in Question 2, how many of them is expected to contain the true difference in the intelligence scores?

$50 \cdot (0.9) = 45 //$

Part IV. **(20 points)** In a survey of a random sample of adults, the respondents are classified according to their current employment and marital status as given in the following table:

	Married at least once	Never married
Employed	490	209
Unemployed	77	40

1. (14 points) Is the employment ratio lower in the population of those who never married than those who married at least once?

$H_0: p_1 = p_2$

$H_1: p_1 > p_2$

$\hat{p}_1 = \frac{490}{490+77} \approx 0.86$

$\hat{p}_2 = \frac{209}{209+40} \approx 0.84$

$SE_1 = \sqrt{\frac{(0.86)(0.14)}{567}} = 0.015$

$SE_2 = \sqrt{\frac{(0.84)(0.16)}{249}} = 0.023$

$SE_{diff} = \sqrt{0.015^2 + 0.023^2} \approx 0.027$

$z = \frac{0.86 - 0.84}{0.027} = 0.74$

$100 - 2.5\% = 90\% \rightarrow 1.65$

$0.74 < 1.65$

Do not reject H_0.

So, we cannot say the employment ratio is lower in the population of those who never married.

2. (6 points) Construct a 90% confidence interval for the difference in the percentage of employment in the two populations.

$90\% \rightarrow z = 1.65$

$SE_{diff} = 0.027$

$CI = (\bar{X}_1 - \bar{X}_2) \mp z\,SE$

$= 0.02 \mp (1.65)\,0.027$

$= [-0.0245, 0.0645] \approx [-0.025, 0.065]$

$= [-2.5\%, 6.5\%]$

Part V. (25 points) Medical equipment are used in measuring the brain activity which result from psychological or physical factors. Several thousands of measurements taken on a given equipment show that the mean activity of a certain part of the human brain is 167.6 units and the standard deviation of the measurements is 52.3 units. The equipment is calibrated (=adjusted) for precision (=correctness) if necessary.

1. (12 points) Recent measurements from 23 randomly selected subjects have a mean of 159.8 units. They are made with the same equipment mentioned above. Is there sufficient evidence to indicate that the equipment underestimates the human brain activity, and therefore needs to be calibrated?

$H_0: \mu = 167.6$

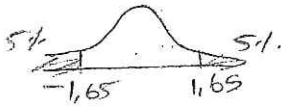
$H_a: \mu < 167.6$

$\Rightarrow$ Not sufficient evidence to indicate underestimation.

$SE = \frac{52,3}{\sqrt{23}} \simeq 10,91$

$z = \frac{159,8 - 167,6}{10.91} \simeq -0,71$

$-1,65 < -0,71 \Rightarrow$ Do not reject H_0!



$0.71 \to 51.61\%$

p-value $= \frac{100 - 51.61}{2}\%$

$\simeq 24.20\%$

$= 0,242 > 0,05$

Do not reject H_0!

2. (10 points) Assume that the standard deviation of the measurements from the equipment has changed and it should be estimated from the data given in Question 1. The data set has a standard deviation of 49.2 units. Construct a 98% confidence interval for the mean of the measurements from the equipment.

$d.f = 22$

$t_{0.01} = 2,51$

$SE = \frac{49,2}{\sqrt{23}} \simeq 10,26$

$CI = \bar{X} \pm t\,SE = 159,8 \mp 2,51 \cdot 10,26 = [134,05,\ 185,55]$

3. (3 points) What would you assume about the distribution of measurements with the equipment, for answering questions 1 and 2 above?

Assume normal distribution.