Mine Çağlar April 19, 2018

Math 202: Statistics for Social Sciences

Spring 2018 EXAM 2

Calculator OK, 90 min.

Instructions: There are five parts to this exam I-V. 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: Solution

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

10 points bonus

Formulas:

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

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

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

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

Binomial Distribution :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0,1,2,\ldots,n$$

Part I. (20 points) Consider a student who is totally unprepared for a multiple-choice exam with 30 questions, and randomly chooses an answer. Each question has five choices, namely, a), b), …, e).

1. (5 points) Fill in the blanks and show your calculations:
   The expected number of questions s/he can answer correctly is 6.

a)…e) → 5 choices

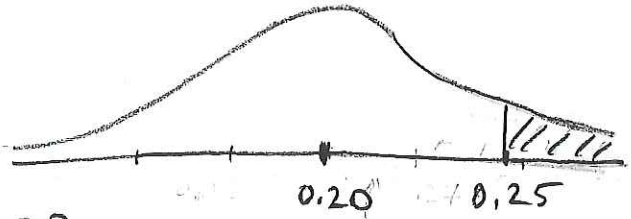
$p = \frac{1}{5} = 20\%$ $\quad 30 \times \frac{1}{5} = 6$

2. (12 points) In a classroom of similar students (all unprepared for the exam, like Hababam sınıfı), there are 24 students. They all take the same exam with 30 questions. What are the chances that the percentage of correctly answered questions in the classroom is greater than 25%?

$n = 24 \times 30 = 720$

$p = 0.20$

$SE = 0.015$



$z = \frac{0.25 - 0.20}{0.015} = 3.33$

→ 99.919% (from normal table)

$\Rightarrow \frac{100\% - 99.919\%}{2} = \boxed{0.0405\%}$

$\Rightarrow SE = \sqrt{\frac{(0.20)(0.80)}{720}} \cong 0.015$

3. (3 points) The percentage of correctly answered questions in the classroom (mentioned in Question 2 above) is a statistics or a parameter? Explain in one sentence.

statistic, because the classroom is in fact a sample.

Part II. **(25 points)** Most people who quit smoking complain of subsequent weight gain. Recently, an innovative program to quit smoking involving weight gain prevention is introduced. It is compared with the standard program to stop smoking as a control. Investigators hypothesize that the smoking abstinence (=quitting) rates in the innovative program would be greater than those in the standard program.

Of 55 subjects assigned to the innovative condition, 29 were not smoking at the end of 52 weeks. On the other hand, 19 of the 54 subjects assigned to the control condition were abstinent at the end of the same period.

1. (12 points) Is the hypothesis of the investigators supported by the data?

$N_1 = 55$; $P_1 = 29/55 = 0.53$
$N_2 = 54$; $P_2 = 19/54 = 0.35$

$H_0: P_1 = P_2$
$H_a: P_1 > P_2$

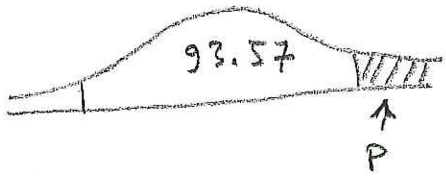
$SE = \sqrt{SE_1^2 + SE_2^2} =$

$= \sqrt{\frac{(0.53)(0.47)}{55} + \frac{(0.35)(0.65)}{54}} =$

0.0936

$z = \frac{(P_1 - P_2) - 0}{SE} = \frac{0.53 - 0.35}{0.0936} = 1.87$

z	Area
1.87	≈ 93.57



$P = \frac{100 - 93.57}{2} = 3.215\% < 5\%$

$P < 5\% \rightarrow$ We reject H_0!
The hypothesis that the new program is more effective is supported by the data

2. (5 points) If the hypothesis of the investigators were "the smoking abstinence rates in the innovative program would be different from those in the standard program", what would your alternative hypothesis be? How much would be the P-value? What would be your conclusion?

$H_a: p_1 \neq P_2$; $P = 6.43\%$ $(2 \times 3.215\%)$ Don't reject H_0! because $P > 5\%$.

3. (8 points) Construct a 95% confidence interval for the difference in the smoking rates with the innovative program and the standard program.

$z_{95} = 1.96$ (or 2)

$SE = 0.093$

$(P_1 - P_2) \pm (SE)(z) =$

$0.18 \pm (0.093)(1.96) =$

$(0.002, 0.362)$

Part III. (20 points) In a study, the effect of psychological counselling on the recovery time from a social trauma such as a major earthquake is investigated. After an earthquake, 23 random subjects were followed until they recovered from trauma. Among them, randomly chosen 12 subjects received counselling; their mean time to recovery was 16.3 months and the standard deviation was 5.2 months. The remaining subjects did not receive any counselling and recovered in 21.8 months on the average with a standard deviation of 6.7 months.

1. (13 points) Do the results support the idea of forming trauma centers for psychological counselling as proposed by some public health officials?

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 < \mu_2$

$N_1 = 12$

$N_2 = 11$ → t-test deg. of freedom $\boxed{21}$

$$SE = \sqrt{\left(\frac{SD_1}{\sqrt{N_1}}\right)^2 + \left(\frac{SD_2}{\sqrt{N_2}}\right)^2} = \sqrt{\frac{5.2^2}{12} + \frac{6.7^2}{11}} = 2.51$$

$$t = \frac{(M_1 - M_2) - 0}{SE} = \frac{16.3 - 21.8}{2.51} = -2.18$$

↓

$1\% < p < 2.5\%$ → we reject H_0.

The results support the idea of forming trauma centers.

2. (7 points) Construct a 95% confidence interval for the mean time to recovery of the people who receive counselling. d.f. $= 12 - 1 = 11$

$t_{2.5\%} = 2.20$

$$SE = \frac{SD}{\sqrt{N}} = \frac{5.2}{\sqrt{12}} = 1.5$$

$16.3 \pm (2.20)(1.5) =$

$(13, 19.6)$ months.

Part IV. (25 points) In a market research study, the number of colors on the package of a bar of candy is investigated. In a sample of 46 bars of candy randomly selected from all types, it is found that 4.3 colors are used on the average, and the standard deviation of the number of colors is 2.5.

1. (3 points) Fill in the blanks: The mean number of colors on the packages of bars of candy in the market can be estimated as 4.3.

   Also, calculate the chance error on your estimation: $\frac{2.5}{\sqrt{46}} = 0.37$

2. (7 points) Construct a 99% confidence interval for the mean number of colors in the package of all candies.

$\bar{X} \pm z \cdot SE =$

$= 4.3 \pm (2.60)(0.37) =$

$= 4.3 \pm 0.95$

$(3.35, 5.25)$

3. (5 points) What is the conclusion of a test of hypotheses in testing whether the mean number of colors in the package of all candies is 5 or not, at $\alpha = 0.01$? Answer by using the result of Question 2 above.

In such a test, we wouldn't reject Ho since 5 is inside the confidence interval.
99% is equivalent to $\alpha = 0.01$ in a two-tailed test.

4. (10 points) Test if the mean number of colors in the package of all candies is 4 or larger, at the level of significance $\alpha = 0.01$? Report the P-value.

$H_0: \mu = 4$

$H_a: \mu > 4$

$z = \frac{4.3 - 4}{0.37} = \frac{0.3}{0.37} = 0.81 \longrightarrow$ Area $= 57.63$ (from normal table)

$P = \frac{100 - 57.63}{2} = \boxed{21.185\%} > 5\%$

So, do not reject Ho.
The mean number of colors is 4.

Part V. **(20 points)** The following computer output is for a test of hypothesis on the commuting time for people who commute to a city from suburbs, given in minutes.

```
Test of mu = 19 vs mu > 19

Variable        N    Mean     StDev    SE Mean    T      P
CommuteTime     26   19.534   ?        0.804      0.66   0.256
```

1. (2 points) Find the standard deviation of the sample, which is left blank in the output above.

$\frac{SD}{\sqrt{26}} = 0.804 \longrightarrow SD = 4.1$

2. (4 points) Is this a one-sample or two-sample test? Is it one-sided or two-sided?

1) one - sample
2) one - sided

3. (4 points) What is the result of the test? State it in plain English as well.

P = 25.6 therefore Ho is not rejected
The average commute time is not significantly greater than 19 minutes.

4. (2 points) Find the P-value using your t-table (give a lower and/or upper bound for it).

deg. of freedom = 26 - 1 = 25
P-value > 25%

5. (5 points) Construct a 99% confidence interval for the mean commuting time of people.

$t_{25, 0.005} = 2.79$

$\alpha = 1\%$

$\alpha/2 = 0.5\%$

$19.534 \pm (0.804)(2.79) =$
$19.534 \pm 2.24 =$
$(17.29 , 21.78)$

6. (3 points) If 99% level confidence intervals are constructed from 200 different samples, how many of those are expected to cover μ?

(1% of 200 is 2, not to cover)
$99\% \times 200 = 198$ are expected to cover true μ.