Math 211: Statistics

December 30, 2017

Fall 2017, FINAL EXAM
2 hrs 15 minutes

<u>Instructions</u>: There are six questions in this exam. Please inspect the exam and make sure you have all 6 questions. You may only use your calculator and **two pages** of A4-size help sheet. Do <u>all</u> your work on the paper provided.

<u>Remember</u>: ***You <u>must</u> show your work to get proper credit.***

<u>Academic Honesty Code</u>: 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: Solutions

1	/23
2	/24
3	/18
4	/15
5	/15
6	/15
Total:	/110

10 points bonus

Question 1: (17 Points) In the advertisement of Supradyn All Day Vitamin, it is announced that 67% of the population press the snooze button (=erteleme düğmesi) of their alarm clock at least once in the morning.

a) (7 points) In a random sample of 99 people, what is the approximate probability that at most 60 of them press the snooze button? Use the central limit theorem.

$$P(X_i=1) = 0.67 \ , \quad P(X_i=0) = 0.33$$

$$P(X_1+\cdots+X_{99} \leq 60) = P\left(\frac{\sum X_i - 99\cdot(0.67)}{\sqrt{99\cdot(0.67)(0.33)}} \leq \frac{60.5-99(0.67)}{\sqrt{99\cdot(0.67)(0.33)}}\right)$$

$$= P(Z \leq -1.246) = 0.1064$$

b) (10 points) If a random sample of 99 people is drawn from the population and it is found that 55 of them press the snooze button at least once in the morning, is the claim of the advertisement supported by the data or not?

$H_0: p = 67\%$ versus $H_1: p < 67\%$

$n = 99$ $\quad k = 55$ $\quad z_{0.05} = 1.65$

$$z = \frac{k-np}{\sqrt{n\cdot p\cdot(1-p)}} = -2.421$$

Since $z < -z_\alpha$, reject H_0. The claim "67% press the snooze button" is not supported by data.

c) (6 points) Calculate the Type II error when the true percentage of people who press the snooze button is 60%.

Reject H_0 If $k < -1.96\sqrt{99\cdot(0.67)(0.33)} + 99\cdot(0.67)$

$= 53.66$

$$1-\beta = P(\text{Reject } H_0 \mid p=0.6)$$

$$= P\left(\frac{k-99(0.6)}{\sqrt{99\cdot(0.6)(0.4)}} \leq \frac{57.16-99(0.6)}{\sqrt{99\cdot(0.6)(0.4)}}\right)$$

$$= P(Z \leq -0.46) = 0.3228$$

Question 2: (24 Points) Answer the following short questions.

a) (5 points) A scientist injects mice, one at a time, with a disease germ until he finds a mouse that has contracted the disease. If the probability of contracting (=having) the disease in the population of mice is 1/6, what is the probability that 8 mice are required?

7 mice won't get the disease and 8th will!

$$P = \left(\frac{5}{6}\right)^7 \cdot \frac{1}{6} = 0.0465$$

b) (7 points) Let $X_1, \dots, X_n$ be independent and identically distributed random variables, as in a random sample. If their mean μ is known, show that the estimator

$$\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \mu^2$$

for their variance σ^2 is unbiased.

$$E\left[\frac{1}{n}\sum^{n} x_i^2 - \mu^2\right] = \left(\frac{1}{n}\sum^{n} E[X_i^2]\right) - \mu^2 \quad \text{since they are independent}$$

$$= E[x_i^2] - E[x_1]^2 = Var[x_1] = \sigma^2$$

c) (5 points) In a random sample of 9 people from a population with 25% minor blood disorder, what is the probability that at most 2 people have this blood disorder?

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$= \left(\frac{3}{4}\right)^9 + \binom{9}{1}\cdot\frac{1}{4}\cdot\left(\frac{3}{4}\right)^8 + \binom{9}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^7 = 0.60$$

d) (7 points) Find $\mathrm{Cov}(X, Y)$ if the joint distribution of X and Y is given by $p(x, y) = \frac{2y}{9x}$ for $x = 1,2$ and $y = 1,2$.

$$P(1,1) = \frac{2}{9}, \quad P(1,2) = \frac{4}{9}, \quad P(2,1) = \frac{1}{9}, \quad P(2,2) = \frac{2}{9}$$

$$E[XY] = \sum_{x,y=1}^{2} x \cdot y \cdot P(x,y) = 1.1.\frac{2}{9} + 1.2.\frac{4}{9} + 2.1.\frac{1}{9} + 2.2.\frac{2}{9} = \frac{20}{9}$$

$$E[X] = 1.P(x=1) + 2.P(x=2) = 1.(P(1,1)+P(1,2)) + 2.(P(2,1)+P(2,2))$$

$$= \frac{6}{9} + \frac{6}{9} = \frac{4}{3}$$

$$E[Y] = 1.P(Y=1) + 2.P(Y=2) = 1.\frac{3}{9} + 2.\frac{6}{9} = \frac{15}{9} = \frac{5}{3}$$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X].E[Y] = 0$$

Question 3: (18 Points) The concentration of an active ingredient in the output of a chemical reaction is strongly influenced by the catalyst used in the reaction. It is believed that when catalyst A is used, the population mean concentration is 65. A sample of results from 10 independent experiments gives the average concentration 64.5 and a standard deviation of 4.8

a) (7 points) Does the sample indicate that the mean concentration is less than 65?

$H_0: \mu = 65$ v.s $H_1: \mu < 65$

$n = 10$ $\bar{y} = 64.5$, $s = 4.8$

$t_{0.05,9} = 1.831$ $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = -0.33$

If $t < -t_{0.05,9}$, reject H_0. But $-0.33 > -1.831$.

So we fail to reject H_0.

b) (11 points) Suppose a similar experiment is repeated with another catalyst, catalyst B, again 10 times. The mean is 70 and the standard deviation is 5.2
Does the sample information suggest any significant difference between the catalysts? Assume the variances of the measurements with different catalysts are equal.

$H_0: \mu_A = \mu_B$ versus $H_1: \mu_A \neq \mu_B$.

$\bar{y}_A = 64.5$, $\bar{y}_B = 70$, $n = 10$, $m = 10$ $s_A = 4.8$ $s_B = 5.2$

$$S_p = \sqrt{\frac{(n-1)S_A^2 + (m-1)S_B^2}{n+m-2}} = \sqrt{\frac{9 \cdot (4.8)^2 + 9 \cdot (5.2)^2}{10+10-2}} = 5.004$$

$$t = \frac{\bar{y}_A - \bar{y}_B}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{64.5-70}{5.004 \cdot \sqrt{\frac{1}{5}}} = -2.458$$

$t_{\alpha/2, n+m-2} = t_{0.025,18} = 2.1009$

Since $t < -t_{0.025,18}$, reject H_0.

There is no significant difference between the catalysts.

Question 4: (15 Points) The following prices in TL are obtained from five different main dishes (ana yemek) in a couple of competitive restaurants, which serve a large variety of dishes both at the moment and also change their menu over time.

	1	2	3	4	5
Restaurant A	12	17	13	15	18
Restaurant B	19	18	17	15	14

a) (8 points) Construct a 95% confidence interval for the difference of the mean price for a main dish in the two restaurants.

$n = 5$, $d_i = A_i - B_i$, $\bar{d} = \frac{\sum^n d_i}{n} = \frac{-7+-1+-4+4}{5} = \frac{-8}{5}$

$t_{0.025,4} = 2.7764$

$$S_D^2 = \frac{n \sum^n d_i^2 - (\sum d_i)^2}{n \cdot (n-1)} = \frac{5(49+1+16+16) - 64}{20} = 17.3$$

$S_D = 4.16$

95% interval is $\left(\bar{d} \mp t_{\alpha/2,4} \cdot \frac{S_D}{\sqrt{n}}\right) = (-6.765, 3.565)$

b) (5 points) Is there a significant difference between the prices of the restaurants at $\alpha = 0.05$?

The difference is not significant at $\alpha = 0.05$ since 0 is in the interval and $\alpha = 0,05 = 1 - 0,95$.

c) (2 points) What are your assumption(s) to be able to answer parts a) and b) above, if any?

Distribution of prices should be Normal.

Question 5: (15 Points) In a study conducted at the Virginia Polytechnic Institute and State University on the development of symbiotic relationship between the roots of trees and a fungus in which minerals are transferred from the fungus to the trees and sugars from the trees to the fungus, 20 northern red oak seedlings with the fungus Pisolithus tinctortts were grown in a greenhouse. All seedlings were planted in the same type of soil and received the same amount of sunshine and water. Half received no nitrogen at planting time in order to serve as a control and the other half received 368 ppm of nitrogen. The stem weights, in grams, at the end of 140 days were recorded. The mean and variance of the control group are 0.40 gr and 0.045, respectively. The mean and the variance for the nitrogen treatment group are 0.57 gr and 0.039, respectively.

In a hypothesis test for the equality of the variances of the control and treatment groups, the null hypothesis was rejected.

a) (11 points) Is there a significant difference between the two groups?

$\sigma_X^2 = 0.045$, $\mu_X = 0.4$, $n = 10$

$\sigma_Y^2 = 0.039$, $\mu_Y = 0.57$, $m = 10$

$H_0: \mu_X = \mu_y$ v.s $H_0: \mu_X \neq \mu_Y$

$$\theta = \frac{\sigma_X^2}{\sigma_Y^2} = 1.15, \quad V = \frac{(\theta + \frac{n}{m})^2}{\frac{1}{n-1}\cdot\theta^2 + \frac{1}{m-1}\cdot(\frac{n}{m})^2} = \frac{(\theta+1)^2}{\frac{1}{9}(\theta+1)} = 17.9, \quad dof = 18$$

So $t_{0.025,18} = 2.1009$.

$$t^* = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{0.4 - 0.57}{\sqrt{\frac{0.045}{10} + \frac{0.039}{10}}} = -1.85.$$

$|t^*| < t_{0.025,18}$. Fail to reject H_0.

There is no significant difference between the two groups for $\alpha = 0.05$.

b) (4 points) Report the P-value.

$$P\text{-value} = P(t < -1.85) + P(t > 1.85)$$
$$= 2 \cdot P(t > 1.85)$$

From table $0.025 < P(t > 1.85) < 0.5$

So $0.05 < P\text{-value} < 0.1$.

Question 6: (15 Points) A professor in a university polled a dozen colleagues about the number of professional meetings professors attended in the past five years and the number of papers submitted by those professors to refereed journals during the same period. The summary data are given as follows:

$$\sum x_i^2 = 232 \ \sum x_i y_i = 318 \ \ n=12 \ \ \bar{x} = 4 \ \bar{y} = 12$$

Fit a linear regression line.

Comment whether attending professional meetings would result in publishing more papers by conducting an appropriate hypothesis test. (Hint: $S = 7.5$ and $\sum x_i^2 - n\bar{x}^2 = \sum(x_i - \bar{x})^2$)

$$b = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} = \frac{12(318) - (48)(144)}{12(232) - 48^2}$$

$$= -6.45$$

$$a = \frac{\sum y_i - b\sum x_i}{n} = \bar{y} - b\bar{x} = 12 + (6.45)4$$

$$= 37.8$$

$$y = 37.8 - 6.45x$$

$$H_0: \beta_1 = 0 \quad \text{v.s.} \quad H_1: \beta_1 > 0$$

$$t_{\alpha, n-2} = t_{0.05, 10} = 1.8125$$

$$t = \frac{\hat{\beta}_1}{S/\sqrt{\sum(x_i - \bar{x})^2}} = \frac{-6.45}{7.5/\sqrt{140}} = -5.44$$

Since $t < 0$, we cannot reject Ho in favor of H_1, which states $\beta_1 > 0$.

So, attending professional meetings do not result in publishing more papers.