

Math 202: Statistics for Social Sciences**Fall 2018 EXAM 1****Calculator OK, 90 min.**

Instructions: There are five parts to this exam, I-V. Please inspect the exam and make sure you have all 5 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.*

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

Part I. (20 points) A study investigates the relationship between the primary and secondary school childrens' scores on intelligence tests and their family backgrounds. A group of students from a rural area has scored 25 on the intelligence test on the average while an average of 28 is obtained by the students from an urban area.

1. (10 points) Describe how a randomized controlled double-blind experiment could have been set up (=kurgulamak) to obtain the results above (5 sentences at most).

In a randomized controlled experiment, the investigators choose the students randomly from all the students of primary and secondary schools in urban and rural areas. Then, they split them according to their region, urban or rural, and call one group as treatment and the other as control group.

The students will be given IQ tests, but should not know the comparison to be done at the end, or any other information about the investigation. Similarly, the investigators who perform the IQ measurements should not know whether a student comes from a rural or urban area.

2. (5 points) Describe how the samples could have been obtained to say that the results come from an observational study? (3 sentences at most)

If the investigators do not choose the students "randomly" from rural and urban areas, it means this is an observational study. Instead, they can choose the students in any arbitrary way. Then, using either available IQ result or taking new IQ measurements, they compare "some" rural and urban area students.

3. (5 points) There are no secondary schools in some rural areas and families move to urban areas to have their children receive further education. Explain why "age" could be a confounding variable if this was an observational study. (3 sentences at most).

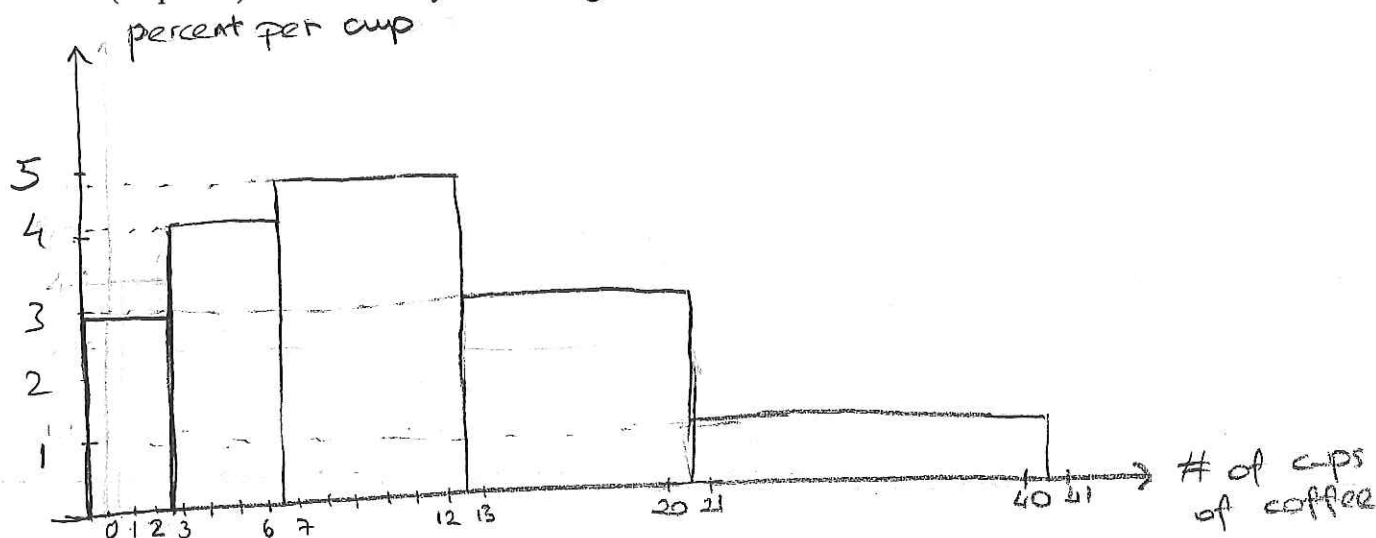
Age is a confounding variable in this case because both it affects the IQ scores, and it is associated with family backgrounds.

- As age increases, IQ also increases.
- Older - age students tend to live in urban areas.

Part II. (25 points) Consider the following frequency table for data collected on the number of cups of coffee students have during the week of final exams. Each subject reports the total number of cups of coffee over the period of finals.

Class interval (number of cups of coffee)	0-2	3-6	7-12	13-20	21-40	Total
Frequency	9	17	30	25	22	103
Rel. Freq.	$9/103 = 8.7\%$	16.5%	29.1%	24.3%	21.4%	
Density	2.9%	14.1%	4.9%	3%	1%	

1. (12 points) Draw a density scale histogram in view of the above table and label both axes.



2. (4 points) In which interval is the 30th percentile? Explain how you found it.

$$\underbrace{8.74\% + 16.5\%}_{\text{First two intervals}} = 25.24\% \quad 25.24\% < 30\% < (25.24 + 29.1)\%$$

↓
In the interval 7-12

3. (2 points) Is the percentile you found in Question 2 above a statistic or a parameter? Explain.

This is a statistic because the percentile comes from a sample, rather than a population.

4. (7 points) A smaller sample was also drawn at random for $\frac{5}{8}$ students: 9, 6, 34, 25, 12. Find the mean and the standard deviation.

$$\text{Mean} = \frac{9 + 6 + 34 + 25 + 12}{5} = 17.2$$

$$\text{S.D.} = \sqrt{\frac{(9-17.2)^2 + (6-17.2)^2 + (34-17.2)^2 + (25-17.2)^2 + (12-17.2)^2}{5}} = \sqrt{112.56} \approx 10.61$$

Part III. (20 points) The weight of Lüfer (a delicious fish) caught in Istanbul Boğazi has a mean of 450 gr (grams) and a standard deviation of 70 g. The following are also known:

The distribution of the weights is left-tailed.

The first quartile (Q_1) is 300 gr and the interquartile range (IQR) is 250 gr

1. (4 points) What percent of the weights lie above 550 g?

$$\Theta_3 = \Theta_1 + IQR = 300 + 250 = 550 \text{ gr.}$$

Since 550 is the 75th percentile, 25% of the weights lie above 550.

2. (6 points) Is the median of the weight distribution likely to be larger or smaller than 450 gr? Explain in one sentence.

Since we know the distribution is left-tailed, median is likely to be larger than the mean.
(mean is smaller than the median)



The information given above is an estimation from a data set of 80 fish. To collect the data, 5 fishermen, each from one of the 5 randomly selected parts of Istanbul Boğazi, were asked to bring a random sample of 20 Lüfer fish from their boat. One of the fishermen did not bring any fish, whereas the other 4 obeyed the instruction. A total of 80 fish obtained in this fashion were used to form the data set of weights. Answer the following questions accordingly.

3. (5 points) Is this a simple random sample? Explain. Also explain the similarity of the data collection to cluster sampling. (3 sentences at most)

This is not a simple random sample because fish are not chosen from all over Istanbul Boğazi at random. Instead, Istanbul Boğazi is split into different parts, as in cluster sampling and 5 of them are first selected at random. Then, in each cluster, fish are chosen randomly.

4. (5 points) Is there any bias in the estimations to be done from this sample? Is there chance error? Explain in 2 sentences at most.

There is non-response bias as one of the fishermen did not bring a sample.

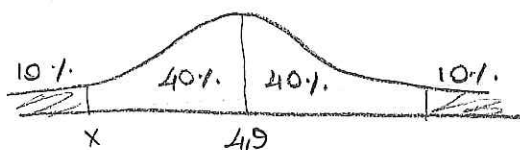
There is chance error because the parts and then the fish are chosen randomly, even if they do not form a "simple" random sample.

Part IV. (25 points) As a result of a Gallup poll, the well-being of Americans is measured on a scale 0 to 10, where “0” represents the worst possible life and “10” represents the best possible life. Respondents are classified by Gallup as “thriving” (=doing well) if they rate their current life a 7 or higher. Assume that well-being scores are approximately normally distributed. The mean and the standard deviation are 4.9 and 1.2, respectively

1. (7 points) What is the 10th percentile of the well-being distribution?

$$z = \frac{x - 4.9}{1.2} = -1.3$$

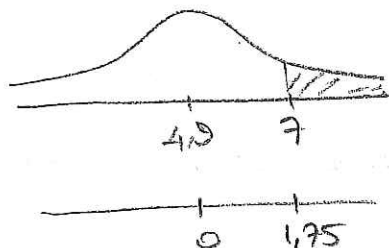
$$x = 3.34 \rightarrow 10^{\text{th}} \text{ percentile}$$



2. (6 points) What percent of the respondents are “thriving”?

$$z = \frac{7 - 4.9}{1.2} = 1.75$$

z	Area
1.75	91.99



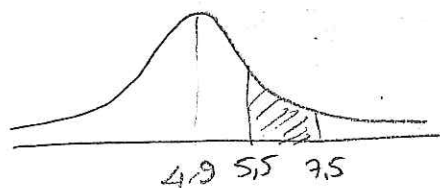
$$\frac{100 - 91.99}{2} \approx 4.1\% \text{ of the respondents are thriving.}$$

3. (8 points) Find the percentage of people who would score between 5.5 and 7.5 on the Gallup poll?

$$z_1 = \frac{5.5 - 4.9}{1.2} = 0.5$$

$$z_2 = \frac{7.5 - 4.9}{1.2} = 2.17$$

z	Area
0.5	38.29
2.17	97.22



$$\frac{97.22}{2} - \frac{38.29}{2} \approx 29.47\% \text{ of people.}$$

4. (4 points) What is the variable in this question? What are the values of this variable?

The variable is the well-being of Americans.

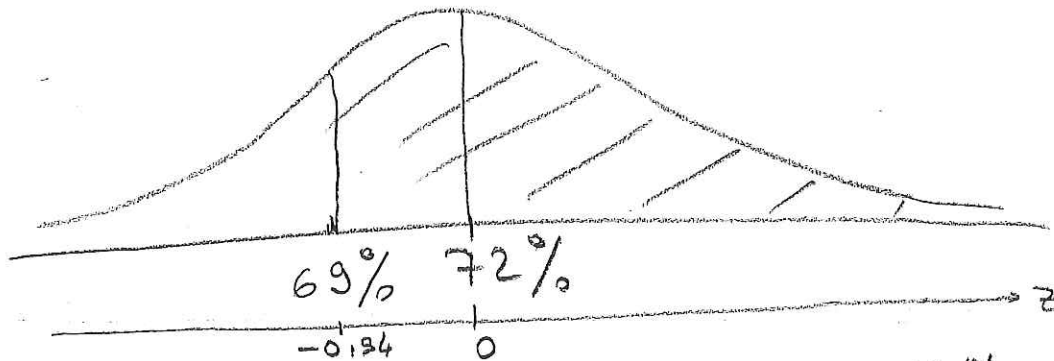
The values of this variable are the integers in the range [9,10]

Part V. (20 points) Eating disorders are found among both genders and among people of all ages, religions, nationalities, and levels of intelligence. It is believed that 72% of all teenagers are concerned (=kaygılı) about their weight.

1. (4 points) What is the expected percentage of teenagers concerned about their weight in a random sample of 200 teenagers?

72% of the sample is expected to be concerned about their weight, because population percentage is 72%.

2. (8 points) In a random sample of 200 teenagers, what are the chances that the percentage of teenagers who are concerned about their weight will be larger than 69%?



$$SE = \sqrt{\frac{(0.72)(0.28)}{200}} \approx 0.032 = 3.2\%$$

$$z = \frac{69\% - 72\%}{3.2\%} = -0.94$$

$$\text{Table} \Rightarrow 50\% + \frac{65.79\%}{2} \approx 82.9\%$$

3. (4 points) Suppose that a random sample of 200 teenagers is taken and it is found that 146 of them are concerned about their weight. What is the percentage of concerned teenagers in the sample?

$$\text{The percentage of concerned teenagers in the sample} = \frac{146}{200} = 73\%$$

4. (4 points) In questions 1, 2, and 3, is (are) there any parameter(s) or statistic(s)? Identify if any.

Parameter: Question 1 $\rightarrow 72\%$, the expected percentage.

Statistic: Question 3 $\rightarrow 146$ or 73% , is a statistic.