Mine Çağlar

Math 202: Statistics for Social Sciences

Spring 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: _Solutions_

Formula:

To find SD, sum all (entry-average)^2 values, then divide by n, and then take the square root

Part I:	/22
Part II:	/20
Part III:	/25
Part IV:	/20
Part V:	/23
Total:	/110

(10 points bonus)

Part I. (22 points) Consider the studies described below and answer the questions that follow.

Study 1: As a part of an international Human Resources study, 800 companies selected randomly in Turkey were contacted in 2015. Of these, 154 accepted to participate in the survey and answered the question "Is the Human Resources Manager consulted by the upper administration for making strategic decisions about the company?". Here are the results for 2015, comparatively with 2005 when a similar question was asked in a similar study:

	2005	2015
Percentage of "No" answers	12%	21%

From 2005 to 2015, the Human Resources managers responding the survey have been trained about "strategic decisions". It is believed that this has increased awareness.

1. (3 points) Is there a bias in the results obtained in 2015? If yes, what type? Explain in one sentence.

   There is non-response bias as many companies did not accept to participate in the survey.

2. (3 points) Is this a controlled study? If so, are the controls historical or contemperaneous?

   This is a controlled study, where the controls are those in 2005, So, these are historical controls.

3. (5 points) Can you conclude that the Human Resources managers are consulted less often in 2015 compared to 2005 for strategic decision making?

   We cannot conclude this even if 12% < 21%. Non-response bias and historical controls make the comparison unreliable. Besides, the managers have been trained in between; the difference may be due to awareness.

Study 2: In an experiment for improving attention, volunteers have been divided randomly into two groups. The first group practiced with computer exercises devised for increasing their visual and auditory attention, and the second group practiced arbitrary computer games/puzzles. Their attention ability was measured after the exercises by psychologists who do not know which group a subject belongs to.

4. (3 points) Is this a randomized controlled experiment? Explain in one sentence.

   Yes, the treatment and control groups are assigned randomly by the experimenter.

5. (4 points) Is it a double blind study or not? Explain in one sentence.

   It is a double blind study because the psychologists do not know and volunteers do not know the content of the computer work they do.

6. (4 points) As a conclusion of the study, it is found that the first group's attention ability is better than the second group after the exercises. On the other hand, some experts have counter opinions that "computer games might decrease attention ability". Accordingly, can you identify a confounding variable in this study?

   Computer medium is a confounding variable in this study. It can decrease the attention of both groups, which fact makes the conclusion unreliable.

Part II. **(20 points)** Sir Francis Galton (1822-1911) is known for his groundbreaking work in using fingerprints for identification purposes, in addition to all his contributions in genetics. The number of ridge crossings, called the ridge count (simply the number of lines that appear in a finger print) in males can be approximated by a normal distribution with mean 146 and standard deviation 52.

1. (7 points) What is the probability that a randomly selected male has a ridge count between 70 and 170?
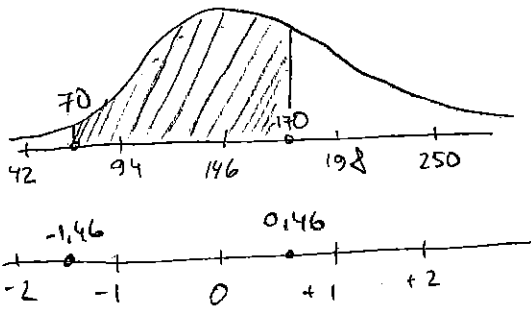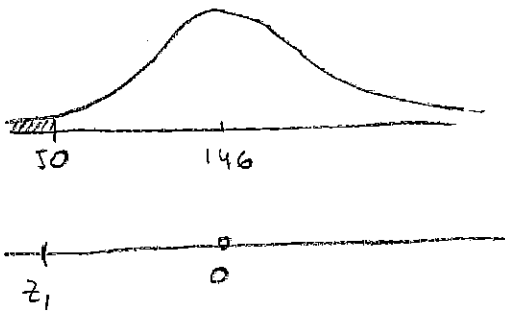
$z_1 = \dfrac{70 - 146}{52} = -1.46$

$z_2 = \dfrac{170 - 146}{52} = 0.46$

z_1	Area
-1.46	85.29

z_2	Area
0.46	34.73

$\dfrac{85.29 + 34.73}{2} = 60.01\%$

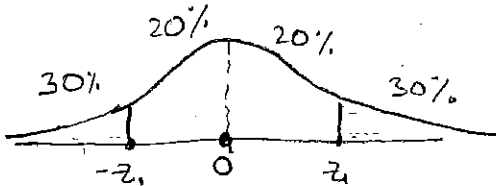2. (4 points) What percent of males in the population have a ridge count less than 50?

$z_1 = \dfrac{50 - 146}{52} = -1.85$

z_1	Area
1.85	93.57

$\dfrac{100 - 93.57}{2} = \dfrac{6.43}{2} = 3.215$

Percentage of males with ridge count below 50 is 3.215 %

3. (6 points) Find the 30th percentile of the ridge count distribution.

$\dfrac{x - 146}{52} = -z_1 = -0.55$

$x = 117.4$

Area	z_1
40%	0.55

The 30th percentile is 117.4

4. (3 points) If you are told that there is a mistake in the standard deviation given above, and the true 30th percentile is 116, what is the true standard deviation?

$\dfrac{116 - 146}{SD} = -0.55$ $SD = 54.55$

Part III. (25 points) The reaction time (in seconds) of the brain to a stimulus in a neurological experiment is recorded in a random sample of subjects and the following data set is obtained given on the left of the page.

11.5
27.2
38.8
47.5
49.6
53.4
54.1
59.3
66.3
68.7
69.3
69.9
71.2
77.5
79.7
84.3

1. (5 points) Form a frequency table by grouping the data in class intervals of size 15, starting with the interval "10-25".

	10-25	25-40	40-55	55-70	70-85
Freq.	1	2	4	5	4
Rel. Freq.	6.25%	10.25%	25%	31.25%	25%

2. (8 points) Draw a relative frequency histogram.
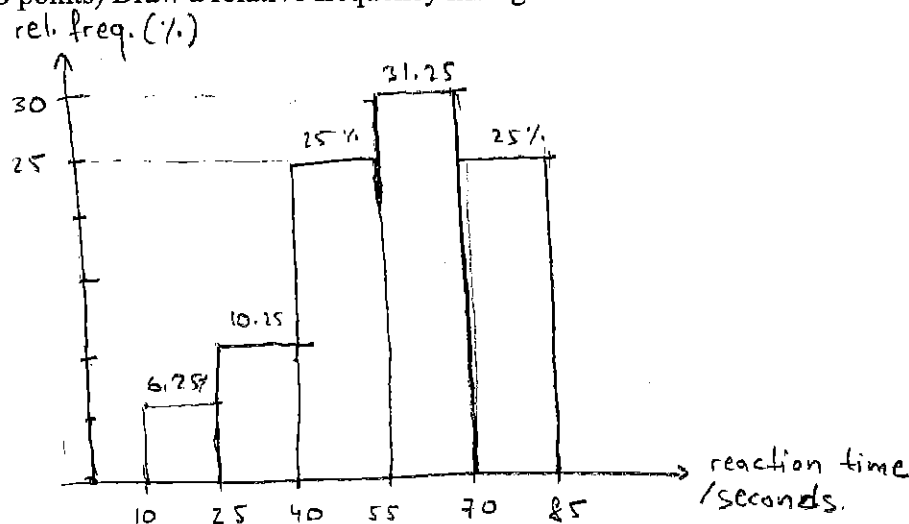
3. (4 points) Find the 90th percentile of the reaction time.

$$\frac{90}{100} \cdot 16 = 14.4$$

..., 71.2, 77.5, 79.7, 84.3

14th

77.5 is the 90th percentile.

4. (8 points) Find the mean and the median. Which one is smaller? Is this what you expect from the shape of the histogram? Explain in 1 sentence.

mean = $\frac{\text{sum of all entries}}{\text{\# of entries}}$ = $\frac{928.3}{16}$ = 58.01

Yes, because the histogram is left skewed

median = $\frac{8th + 9th}{2}$ = $\frac{125.6}{2}$ = 62.8

Part IV. (20 points) Medical conditions are compared on "seriousness of illness rating scale" (SIRS). That is, SIRS quantifies how serious a medical condition is, such as high blood pressure, diabetes, alcoholism. For a large sample, the following table gives the percentages of SIRS scores.

	400	100	100	200	200
SIRS	0- 400	400 – 500	500 - 600	600 - 800	800 - 1000
Frequency	15%	17%	23%	12%	18%
Density	0.0375	0.17	0.23	0.06	0.09

1. (7 points) Draw a density scale histogram for SIRS.
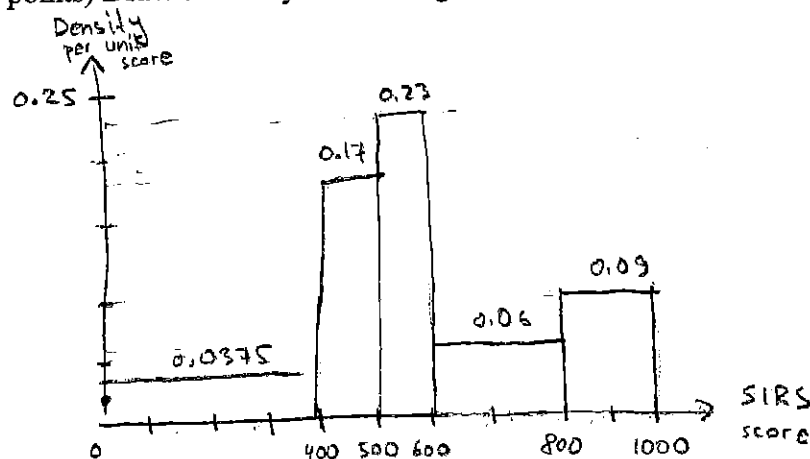
2. (4 points) The 25th percentile is considered to be the cutoff for less serious medical conditions. At which interval is the 25th percentile?

$$15\% < 25\% < (15 + 17)\%$$
$$= 32\%$$

In the 2nd interval.

3. (7 points) Consider the following SIRS scores, which form a smaller data set:

$$173, 621, 636, 454, 312, 733$$

Find its mean and standard deviation.

$$mean = \frac{173+621+636+454+312+733}{6} = \frac{2929}{6} = 488.17 \approx 488$$

$$SD = \sqrt{\frac{(173-488)^2+(621-488)^2+...+(733-488)^2}{6}} = \sqrt{46,194} = 214.9$$

4. (2 points) Are the mean and standard deviation that you found in Question 3 statistics or parameters?

Both are statistics since the smaller data set is a sample.

Part V. (23 points) A newspaper recently reported that the Pizza sector is in the rise (=yükseliş). Six brands aim at opening 600 new branches. Assume that the Pizza companies do a market research by applying a multistage cluster sampling, for finding out the demand for different types of pizza. They ask the respondents which type of pizza they prefer, in addition to others.

1. (3 points) State the variable of interest.

   Types of pizza in demand.

2. (4 points) What are two possible values of the variable that you stated above? Is this a qualitative or a quantitative variable?

   Mixed pizza, Turkish pizza.
   Qualitative variable

3. (6 points) Explain how a multistage cluster sampling can be performed all over Turkey, in four sentences at most.

   - The country is divided into regions of comparable population size.
   - From each region, select certain number of cities randomly, and select certain number of districts randomly from those cities.
   - Select randomly households from the districts
   - Survey those households.

4. (6 points) Which of the following are applicable in the above research: i) non-response bias, ii) interviewer bias, iii) selection bias? Explain in three sentences at most.

   iii) can be neglected because the sampling procedure randomizes the selection as much as possible.
   ii) Interviewer bias is also minimized, but still within a household, the interviewer can choose individuals to question by convenience.
   i) is applicable, the households could refuse to do a survey

5. (4 points) Is there chance error in multistage cluster sampling? Explain in one sentence.

   Yes, because the cities, districts and households are all selected at random and there is chance error in random selection.