Part I. (20 points) According to Webster's New World Dictionary, a tongue twister (= dil büken söz) is "a phrase that is hard to speak rapidly". Do tongue twisters have an effect on the length of time it takes to read silently? To answer this question, a randomized controlled experiment is designed where each subjects is asked to read a list of phrases. Two lists of phrases are formed; the one with tongue twisters is given to the treatment group and the list with no tongue twisters is given to the control group.

The following is a MINITAB output for the time (in minutes) it takes to read the whole list for the two groups:

Variable	N	Mean	StDev
Tongue Twister	11	7.59	1.94
Control	9	6.34	1.92

1. (2 points) What is the variable under study?

The time it takes to read silently.

2. (12 points) Is the reading time significantly large in the treatment group? First, write your assumption(s) for solving this problem with the methods you learned in Math 202.

Assumption: The reading time is normally distributed. We can apply t-test under this assumption, as the sample size is small.

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 > 0$

$$SE = \sqrt{\frac{(1.94)^2}{11} + \frac{(1.92)^2}{9}} \cong 0.87$$

$$t = \frac{7.59 - 6.34 - 0}{0.87} = 1.44$$

$$d.f. = (11-1) + (9-1) = 18$$

t-table $\Rightarrow$ $5\% < P\text{-value} < 10\%$

Since P-value $> 5\%$, do not reject H_0.

There is no difference in the mean reading time of "tongue twisters" and "control" lists.

3. (6 points) Form a 95% confidence interval for the difference in the mean reading times of the two groups.

$(\bar{x}_1 - \bar{x}_2) \mp t\,SE$ $\quad$ $d.f. = 18$

$1.25 \mp 2.10(0.87) \Rightarrow [-0.577, 3.077]$

Part II. (20 points) Kleenex, the top-selling brand of tissues (= kağıt mendil), has announced that the median for the number of tissues people use at the time of a cold is 58, in view of the company's extensive (=geniş) market analysis in the past.

1. (2 points) What is the percentage of people who use more than 58 tissues at the time of a cold?

50% since 58 is the median.

2. (3 points) The percentage mentioned in Question 1 is : (a) parameter b) statistics (circle one)
Explain why, in 2 sentences at most:

This result comes from an extensive market analysis which can be considered to cover all population.

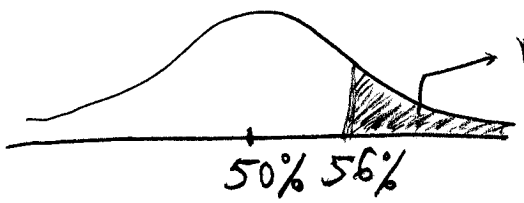
3. (12 points) A recently taken random sample is composed of 100 people with cold. It is found that 56% of them use more than 58 tissues. Is the median number of tissues used by people with cold increased recently? Perform an appropriate hypothesis test to answer this question.
(Hint: Consider the percentage of people who use more than 58 tissues.)

$H_0: p = 50\%$

$H_a: p > 50\%$

$SE = \sqrt{\frac{(0.50)(0.50)}{100}} = 0.05 = 5\%$

$z = \frac{56\% - 50\%}{5\%} = \frac{6}{5} = 1.2$



$P\text{-value} = \frac{100 - 76.99\%}{2} \cong 11.5\%$

Since P-value > 5%, do not reject H_0.

There is no significant evidence that the median has increased.

4. (3 points) Complete the following sentence about the interpretation of the *P*-value:

The P-value (of the test performed in Question 3) corresponds to the probability of observing 56 % or more of the sample using 58 or more tissues, when actually the true percentage in the population who use more than 58 tissues is 50 %.

Part III. (20 points) To determine the level of cell phone use by drivers while they are driving a vehicle, data are collected by observers at randomly selected intersections (= kavşak). A total of 450 commercial vehicles and 675 personal-use vehicles are recorded. The numbers of drivers who speak on their cell phone are 68 and 135 in the commercial and personal-use vehicles, respectively.

1. (14 points) Is the difference between the commercial and personal-use vehicle drivers about the usage of cell phone significant, or just due to chance?

$H_0: p_1 - p_2 = 0$

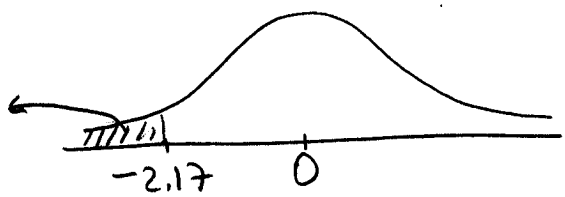
$H_a: p_1 - p_2 < 0$

$\hat{p}_1 = \frac{68}{450} = 0.15 \quad \hat{p}_2 = \frac{135}{675} = 0.20$

$$SE = \sqrt{\frac{(0.15)(0.85)}{450} + \frac{(0.20)(0.80)}{675}}$$

$$\cong 0.023$$

$$z = \frac{0.15 - 0.20 - 0}{0.023} = -2.17$$



P-value

$= \frac{100 - 96.84\%}{2}$

$= 1.58\%$

Since P-value $< 5\%$, reject H_0.

There is a significant difference between the two types of drivers: Personal-use vehicle drivers speak on the cell phone more often than commercial drivers while driving.

2. (6 points) Find a 99% confidence interval for the percentage of drivers who speak on the phone while driving (without paying attention to what type of vehicle they are driving!).

$$\hat{p} = \frac{68 + 135}{450 + 675} = 0.18$$

$$0.18 \mp 2.6 \sqrt{\frac{(0.18)(0.82)}{1125}}$$

$$\Rightarrow \quad [0.15, 0.21]$$

Part IV. (20 points) In the current education system, very few children repeat a grade (= sınıf) in elementary school. Suppose the average height of all elementary school children is 145cm. A study on randomly selected 55 children who repeat a grade reports an average height of 140cm and a standard deviation of 12cm.

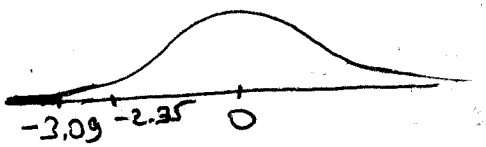
1. (10 points) Does the evidence support the claim that the children who repeat a grade in elementary school are shorter, on average, than their peers (= sınıf arkadaşı) at $\alpha = 1\%$?

$H_0 : \mu = 145$

$H_a : \mu < 145$

$$z = \frac{140 - 145}{12/\sqrt{55}} = -3.09$$

$\alpha = 1\% \Rightarrow z_{critical} = -2.35$ (z-table → area 98%)

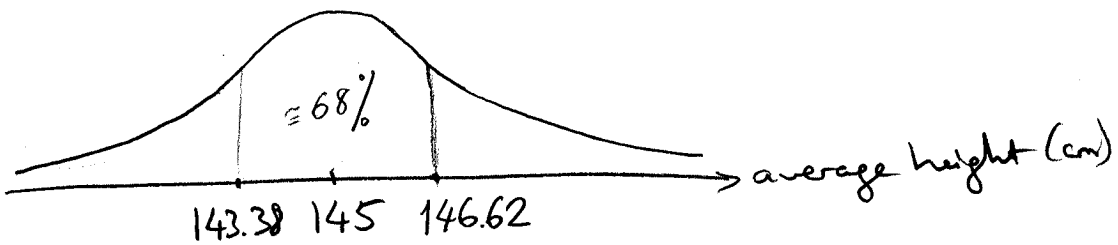


$-3.09 < -2.35 \Rightarrow$ Reject H_0

(OR, p-value $= \frac{100 - 99.806}{2} \approx 0.1\%$ since $0.1\% < 1\%$, reject H_0)

Children who repeat a grade are significantly shorter than their peers.

2. (5 points) Sketch an approximate histogram of the average height of all possible random samples of 55 elementary school grade children, and show <u>three</u> height values on the x-axis.



3. (5 points) Consider another scenario different from the one described in the question. Suppose two samples are taken, one of them from children who repeat a grade and another one from the general population of children in elementary school. Can you apply a two-sample test you learned in Math 202? Explain WHY in at most in 1 sentence in each of the following cases.

a) If the two samples are chosen independently from the respective populations, namely the population of children who repeat a grade, and the general population.

Yes, we can apply as the samples are independent. (Overlap between the populations is not a problem)

b) If some of the children who repeat a grade participate in both samples?

No, we cannot apply because the samples are not independent. We need a different test in this case.

Part V. (20 points) A question of current interest is "How is history written by historians?". A large survey of 2000 historians working academically in many universities all over the world concentrated in three subfields of history: the subjects are chosen from diplomatic, social, and cultural history fields (the other fields are ignored). They are asked which method they mostly use for collecting information about the past, namely, interviews or archival documents. The following table summarizes the outcome of this survey, which approximates the relative frequencies in the population.

	Diplomatic History	Social History	Cultural History
Interviews	0.04	0.16	0.22
Archival documents	0.25	0.23	0.10

1. (6 points) What are the variables in this study?
What are their respective values?

Variables: Subfields of history and Method used in collecting information.

Values: Diplomatic history, Social history, Cultural history and Interviews, Archival Documents

2. (4 points) What is the probability that a subject selected from the participants of the survey works in Social History?

$$0.16 + 0.23 = 0.39$$

3. (4 points) What percent of diplomatic historians use archival documents mostly?

$$0.04 + 0.25 = 0.29$$

$$\Rightarrow \quad \frac{0.25}{0.29} \cong 0.86 = 86\%$$

4. (6 points) Write down the conditional distribution of the fields of subjects who mostly use interviews as their method of research.

% who use interviews: $0.04 + 0.16 + 0.22 = 0.42$

Diplomatic History	Social History	Cultural History
$\frac{0.04}{0.42} = \frac{2}{21}$	$\frac{0.16}{0.42} = \frac{8}{21}$	$\frac{0.22}{0.42} = \frac{11}{21}$