

**Part I. (20 points)** In a study on the relationship of mental performance and adequate sleep, published in the prestigious magazine *Nature* in 2004, a sample of 100 volunteers was divided into two equal-sized groups. Each volunteer took a test that involved finding a hidden rule in the given problem. Prior to taking the test, one group received eight hours of sleep while the other group stayed awake all night. The scientists monitored the volunteers to determine whether they found the hidden rule, without knowing which group the subject belongs to. Of the volunteers who slept, 39 discovered the hidden rule, of the volunteers who stayed awake all night, 15 discovered the hidden rule.

1. (4 points) Is this a randomized controlled experiment or an observational study? Explain in two sentences at most.

Randomized controlled experiment because the experimenter assigns the subjects to the treatment and control groups.

Treatment group → who sleep adequately (or vice versa)  
Control group → who stay awake

2. (4 points) Is this a double-blind study? Explain in two sentences at most.

It's not double-blind because the volunteers know in which group they are in. However, the scientists who monitor do not know their group, they are "blind" in this experiment. (Note that the treatment group, those who sleep, may also not know what group they are in. But, those who do not sleep definitely know they are left sleepless on purpose!)

3. (4 points) What would you conclude about the relationship of mental performance and adequate sleep based on this study? Explain how you reached this conclusion in at most two sentences.

Adequate sleep improves mental performance

I compared  $\frac{15}{50} = 30\%$  with  $\frac{39}{50} = 78\%$ .

Consider the two different scenarios given in Questions 4 and 5:

4. (4 points) Suppose the scientists obtained the sample of 100 volunteers from the students of a leading Turkish university. They used the sample percentage of people who find the hidden rule after adequate sleep to estimate the corresponding percentage in the population. Is this estimate likely to be too low, too high or just about right? Explain in at most two sentences. (Hint: Note that these students are experts on finding rules in a problem, due to their preparation for the university entrance exam)

The estimate is too high, because the students in the sample are trained better than the rest of the population for finding the rules in a given problem.

5. (4 points) If the scientists want to generalize their findings to the whole city they live in, they should sample 100 people from all around the city. However, the city's population is two million.

Describe how they should pick 100 subjects from this city in at most four sentences. Since the city

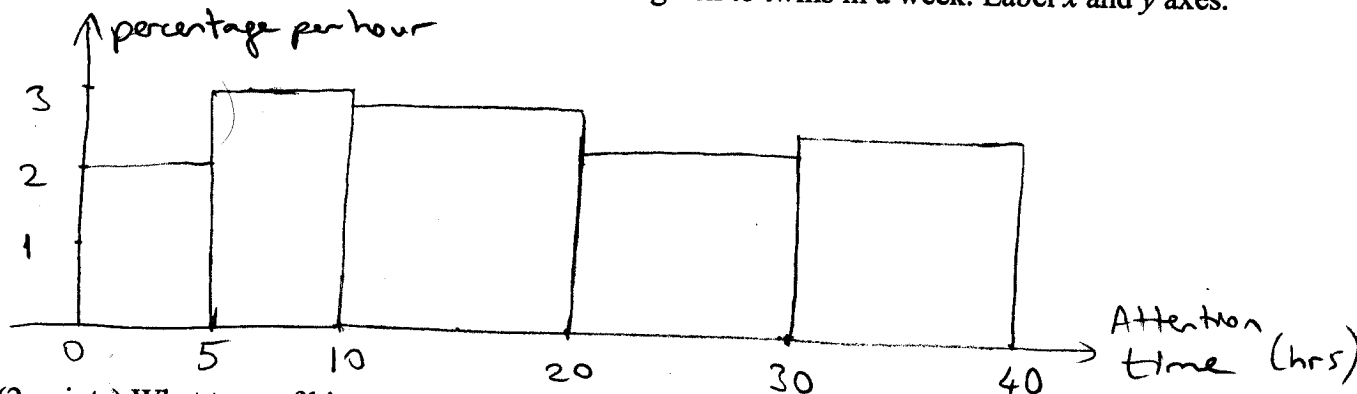
is huge, they should use (multistage) cluster sampling.

The city should be split into smaller geographical regions. Some of the regions should be selected randomly, then districts are to be selected randomly from those and finally households and subjects would be selected randomly.

**Part II. (20 points)** Psychologists have found that twins learn language more slowly in their early years than nontwins. Suspecting that this may be caused by parental neglect, they study the attention time given to each pair of twins by their parents in a week. The data collected from a simple random sample of 50 twin boys (at the same age) are summarized in the following relative frequency table.

Attention Time (hours)	0-5	5-10	10-20	20-30	30-40
Relative frequency	%10	%15	28%	%22	%25

1. (8 points) Draw a histogram of the attention time given to twins in a week. Label  $x$  and  $y$  axes.



2. (2 points) What type of histogram did you draw: density scale or relative frequency? Why?

Density scale histogram because the intervals of attention time are not equal length.

3. (5 points) Can you find the interquartile range for the attention time given to twins in a week using the given information? If yes, find it. If not, explain in one sentence.

25<sup>th</sup> percentile : 10 hrs.  
75<sup>th</sup> percentile : 30 hrs

$$\Rightarrow IQR = 30 - 10 = 20 \text{ hrs.}$$

4. (2 points) Does your histogram have a long right-hand tail, a long left-hand tail, or is it approximately symmetric?

It is approximately symmetric.

5. (3 points) Choose one:

- a) The average is likely to be larger than the median,
- b) The average is likely to be smaller than the median
- ☒ c) The average is about the same as the median

In one sentence, explain why you chose this option.

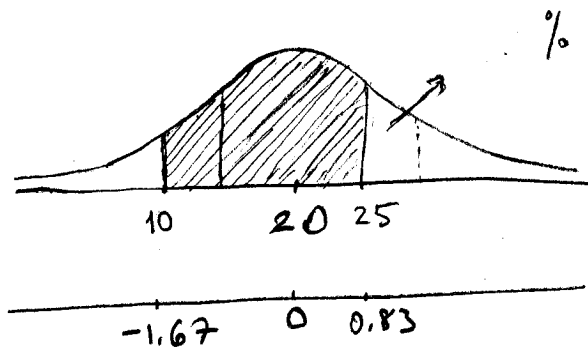
Because the histogram is approximately symmetric.

**Part III. (20 points)** An experienced editor of a magazine reports that the distribution of the number of mistakes found in an article of this magazine can be approximated by a normal curve with mean 20 and standard deviation 6.

1. (2 points) Are the numbers given in this question parameters or statistics? Explain in one sentence.

They are parameters as they are from the population as suggested by the "experienced" editor.

2. (7 points) What percent of all articles have between 10 and 25 mistakes?



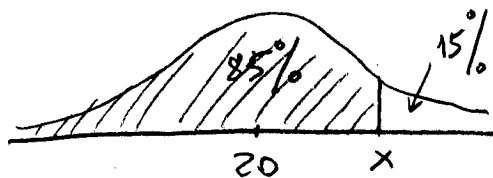
$$\frac{10-20}{6} = -1.67 \rightarrow 90.11\%$$

$$\frac{25-20}{6} = 0.83 \rightarrow 60.47\%$$

$$\frac{90.11 - 60.47}{2} = 14.82\%$$

$$\text{shaded area} = 60.47\% + 14.82\% = 75.29\%$$

3. (6 points) What is the 85th percentile of the mistakes distribution?

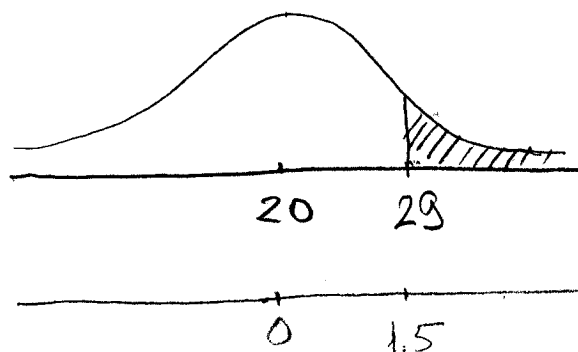


$$\text{Table} \Rightarrow 70\% \rightarrow z = 1.05$$

$$20 + (1.05)6 = 26.3 \text{ mins}$$



4. (5 points) Find the chance that the editor will correct more than 29 mistakes in the next article she edits?



$$\frac{29-20}{6} = 1.5$$

↓ table  
86.64%

$$\frac{100 - 86.64}{2} = 6.68\%$$

**Part IV. (20 points)** At the times of a flu pandemic (=salgin), it is believed that 70% of the people with flu go home and have a rest.

1. (10 points) In a random sample of 84 people with flu, what is the probability that the percentage of people who go home and rest is larger than 60%?

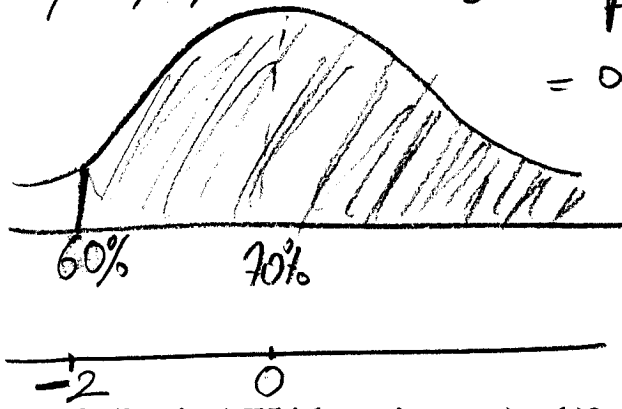
$$p = 70\% \Rightarrow SE = \sqrt{\frac{(0.7)(0.3)}{84}} = 0.05 = 5\%$$

$$z = \frac{60\% - 70\%}{5\%} = -2$$

$$-2 \xrightarrow{\text{Table}} \sim 95\%$$

Shaded area:

$$\frac{100\% - 95\%}{2} + 95\% = 97.5\%$$



2. (3 points) Which one is true, a) or b)?

- ☒ a) The expected value of the percentage of people in the sample who go home and rest when they have flu is exactly 70%.
- b) The expected value of the percentage of people in the sample who go home and rest when they have flu is about 70%, give or take 5% or so.

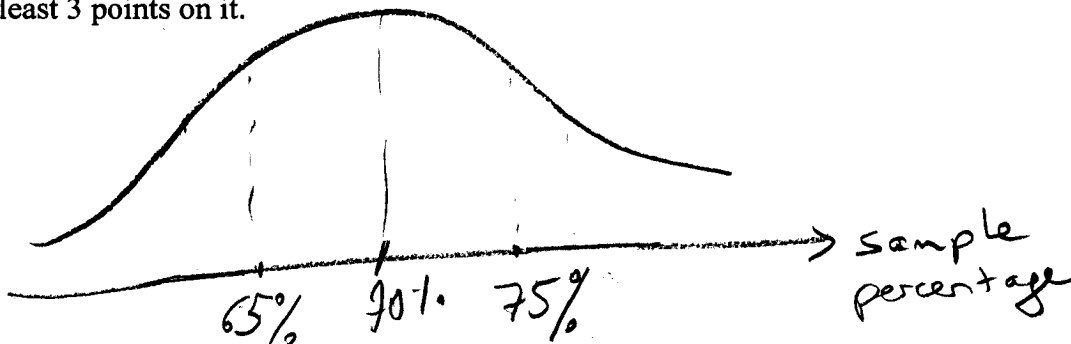
Explain in one sentence at most.

The expected value of the sample percentage is the population percentage, which is 70% exactly, as there is no error associated with population percentage.

Suppose 100 different samples of size 84 are taken and the percentage of people in the sample who go home and rest are recorded for each one as illustrated below.

Sample No.	1	2	3	4	5	6	7	8	---	100
Sample percentage	71%	74%	54%	62%	75%	69%	67%	82%	---	55%

3. (4 points) Sketch an approximate histogram for these sample percentages. Label the x-axis and show at least 3 points on it.



4. (3 points) In Sample No.5, what is the observed number of people who go home and rest?

$$(75\%) 84 = 63$$

**Part V. (20 points)** Kleenex is recognized as the top-selling brand of tissue (=kağıt mendil) in the world. One of its products is a cold-care box of tissues (with lotion, hence soft!) designed especially for people who have cold. Marketing experts of the company conduct a survey to determine how many tissues will be packed in a cold-care box. They want to pack as many as the average number of tissues used by people with cold. In this study, 250 randomly selected consumers report the number of tissues they used when they had a cold, as summarized in the following MINITAB output.

Variable	N	Mean	StDev	Median
Tissues	250	56.7	25.3	58

1. (10 points) Find a 90% confidence interval for the average number of tissues used by Kleenex customers at the time of a cold.

$$\bar{x} \pm z \frac{SD}{\sqrt{n}}$$

$$56.7 \pm 1.65 \frac{25.3}{\sqrt{250}} \Rightarrow [54.1, 59.3]$$

90%, table      1.6

2. (3 points) Interpret the interval you found in Question 1, by completing the following sentence:

I am 90% confident that the mean number of tissues used  
by all Kleenex consumers at the time of a cold is between  
54.1 and 59.3 tissues

3. (3 points) What percent of the customers in the sample used more than 58 tissues? Explain.  
 (Hint: Look at the MINITAB output.)

Since 58 is the median, 50% used  
 more than 58 tissues.

4. (4 points) The percentage of all customers who use more than 58 tissues when they have a cold is estimated as 50%. The standard error attached to this estimate is 3.2%.

(by 3.)

$$SE = \sqrt{\frac{(0.5)(0.5)}{250}} \cong 0.032 = 3.2\%$$