

Part I. (20 points) Answer the following short questions.

1. (3 points) Define "parameter" in one sentence. (Please note that any other sentences will be ignored).

A numerical fact about the population.

2. (5 points) What is an observational study? Explain by comparing it to a controlled experiment in at most 3 sentences. (Please note that any other sentences will be ignored).

An observational study is a study where the subjects assign themselves to the control and treatment groups. In contrast, the investigator assigns them in a controlled experiment.

3. (7 points) A social scientist wishes to investigate the following research question: "Do seatbelts (=emniyet kemeri) reduce car crash injuries?"

- a) Identify two variables that the researcher should collect data for.

Seatbelt status

Level of injury from a car crash

- b) Another researcher thinks that there may be a confounding variable that should be controlled for, namely, "the condition of being risk averse or not". Explain why this could be a confounding variable. (3 sentences at most!)

(risk averse person = a person who stays away from risky situations).

Being risk averse or not affects the way a person behaves; s/he may buckle up (wear seat belt) and be more careful while driving. The latter may reduce the level of injury. In short, it is a confounding variable because it affects both variables in a).

4. (5 points) The following is a frequency table based on a random sample of the number of rooms in a house around Sariyer district. Estimate the average number of rooms in the houses of Sariyer.

Number of rooms	1	2	3	4	5
Percentage	2%	5%	34%	47%	12%

$$\begin{aligned} \text{average} &= 1(0.02) + 2(0.05) + 3(0.34) + 4(0.47) \\ &\quad + 5(0.12) \\ &= 3.62 \end{aligned}$$

Part II. (20 points) The following are IQ scores from a random sample of 5 individuals.

101, 118, 114, 95, 136

1. (3 points) Fill in the blanks in the following frequency table.

Score	90-100	100-120	120-140
Percentage	20%	60%	20%

($\frac{1}{5}$, $\frac{3}{5}$, $\frac{1}{5}$)

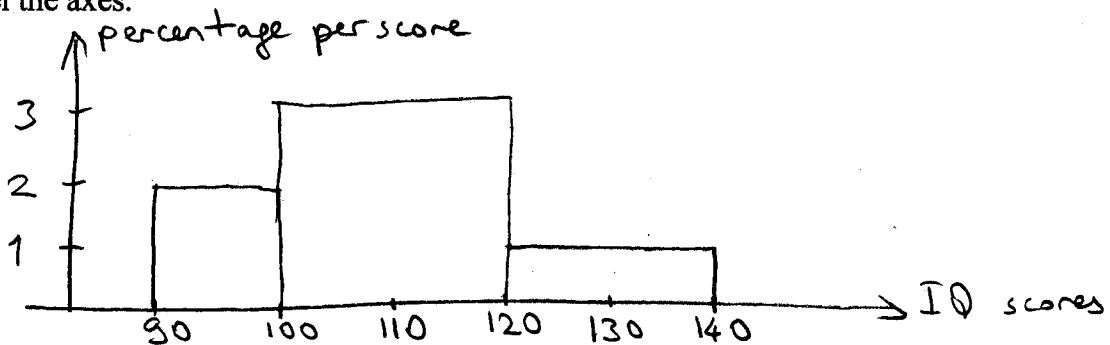
2. (5 points). Find the standard deviation (SD) of the sample.

$$\text{average} = \frac{101 + 118 + 114 + 95 + 136}{5} = 112.8$$

$$\text{SD} = \sqrt{\frac{(101-112.8)^2 + (118-112.8)^2 + (114-112.8)^2 + (95-112.8)^2 + (136-112.8)^2}{5}}$$

$$= \sqrt{204.56} = 14.3$$

3. (5 points) Draw a density scale histogram using the frequency table above. Do not forget to label the axes.

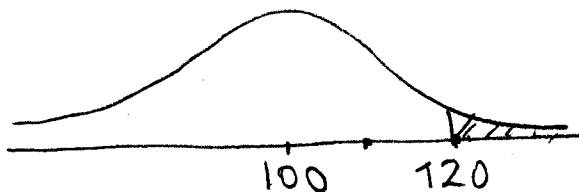


4. (3 points) What is the 40th percentile of the IQ distribution as estimated from the sample?

sorted data: 95 101 114 118 136

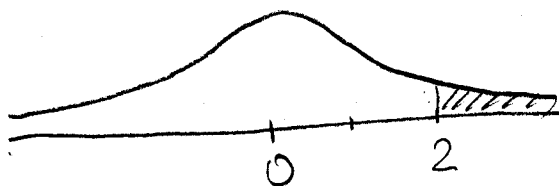
$\frac{40}{100} = \frac{2}{5} \Rightarrow 101$

5. (4 points) If IQ scores in the population can be approximated by a normal curve with mean 100 and standard deviation 10, what percentage of people would have an IQ greater than 120?



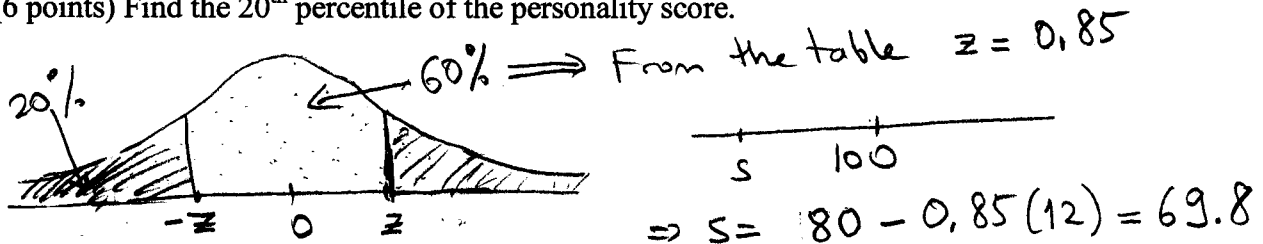
shaded area:

$$\Rightarrow \frac{100 - 95}{2} \% = 2.5\%$$

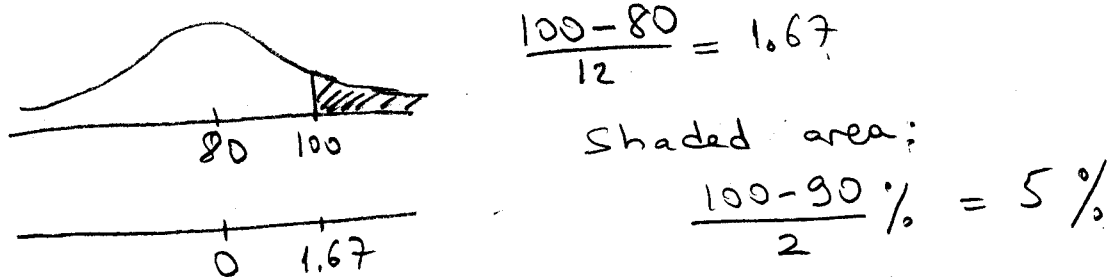


Part III. (25 points) The scores on a personality questionnaire follow a normal curve with population average of 80 and standard deviation of 12.

1. (6 points) Find the 20th percentile of the personality score.



2. (6 points) What is the probability that a randomly selected person from this population will have a score above 100?

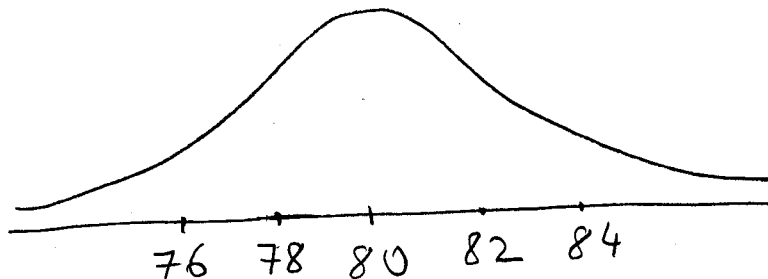


Consider random sampling with a sample size of 36 from this population in the following questions:

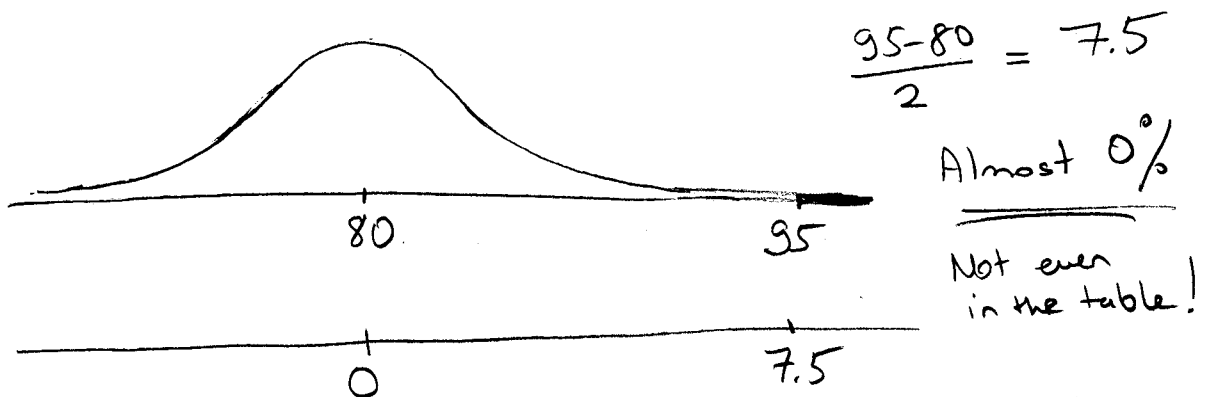
3. (3 points) Find the standard error of the sample average.

$$SE = \frac{SD}{\sqrt{n}} = \frac{12}{6} = 2$$

4. (5 points) Sketch the histogram of all possible sample averages that could be obtained (sketch as a smooth curve). Show the values of at least 3 points on the x-axis.



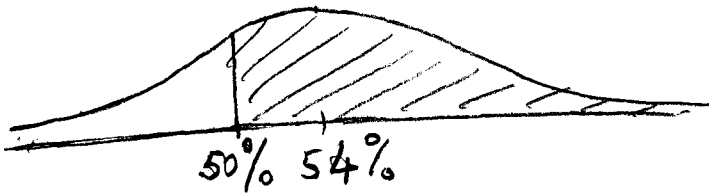
5. (5 points) Of all the possible sample averages, what percent will be greater than 95?



Part IV. (20 points) In a given country, a “referandum” is going to be held to elect the president. In other words, a candidate is elected only if s/he gets at least 50% of all votes. A polling company takes a simple random sample of size 60 to predict the result of the elections. Consider a specific candidate named Mr. Bean who is favored as president by 54% of the sample.

1. (8 points) Estimate the chance that Mr. Bean is going to win the presidential election.

$$SE = \sqrt{\frac{(0.54)(0.46)}{60}} \approx 0.064 \approx 6.4\%$$



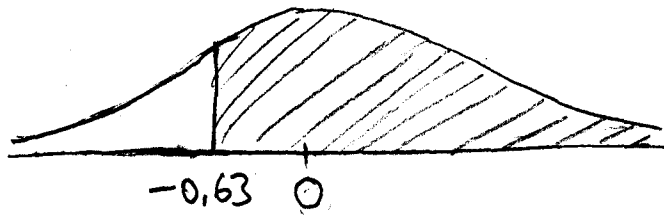
$$\frac{50-54}{6.4} = -0.63$$

$$\text{Table} \rightarrow 0.65 \rightarrow 48.43\%$$

$$\Rightarrow \frac{100 - 48.43}{2} \% = 25.8\%$$

\Rightarrow shaded area:

$$48.43\% + 25.8\% = 74.2\%$$



2. (8 points) Construct a 95% confidence interval for the percentage of voters who will favor Mr. Bean in the election.

$$54\% \pm 2(6.4\%)$$

$$\Rightarrow [41.2\%, 66.8\%]$$

3. (4 points) Is there a “statistic” or a “parameter” in this question? What is its value?

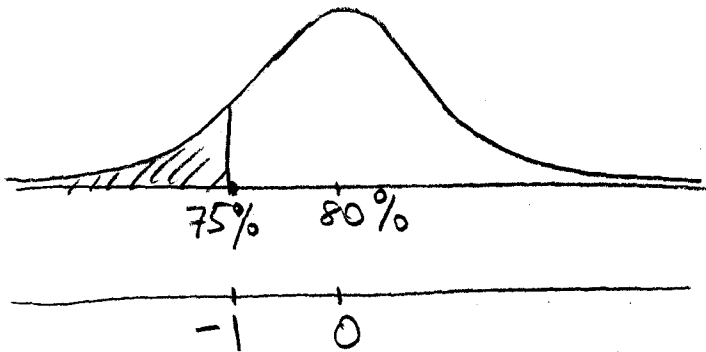
Statistic, which is 54% and is a numerical fact from the sample.

Part V. (15 points) In Turkey, the percentage of all crimes among women that are committed with the husband's gun is believed to be 80%.

1. (10 points) Find the probability that the percentage (of crimes committed with husband's gun) in a random sample of size 64 (from the population of women who committed a crime) will be less than 75%.

$$\sqrt{64} = 8$$

$$\Rightarrow SE = \frac{\sqrt{(0.80)(0.20)}}{8} = \frac{0.4}{8} = 0.05 = 5\%$$



Shaded area:

$$\frac{100 - 68}{2} \% = 16\%$$

2. (5 points) If 10 different samples are withdrawn from Turkey and a 90% confidence interval is constructed using each sample, how many of these intervals do you expect to contain 80%?

Each has a chance of containing the true percentage by 90%.

$$\Rightarrow \frac{90}{100} \cdot 10 = 9 \text{ of them.}$$