

Solutions

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

1. (3 points) Define "statistics" as a discipline in at most 2 sentences (Please note that any other sentences will be ignored).

Statistics is concerned with data collection, summarizing and tabulating data, and drawing conclusions and making generalizations from a sample for the population.

2. (3 points) The results of what type of a study is generalizable to the whole population?

Observational Study or Randomized Experiment

Consider the following information to answer questions 3 to 6.

A statistics student hypothesizes that eating more fruits and vegetables may prevent cancer. She conducts a survey in Sanyer by visiting the households she finds at home and asking each individual two questions:

- i) How many portions of fruits and vegetables do you eat per day?
- ii) Did you have any form of cancer in your life up to this point?

The student found out that there is an association between the variables implied by those two questions.

3. (4 points) What are ~~the names~~ of the variables?

- 1) The number of portions of fruits and vegetables consumed.
- 2) Condition of having cancer or not.

4. (5 points) Is this an observational study or a controlled experiment? Explain briefly.

It is an observational study because the student has not assigned the subjects to different levels of fruits & vegetables consumption. She has just observed what the subjects did themselves (eat less or much...)

5. (5 points) Actually, the student also asked some other questions to determine if the subject follows a healthy life style (score=1) or not (score=0). Shortly called "lifestyle", this is a confounding variable. Explain briefly.

Lifestyle affects both variable 1) and variable 2). It may be the main cause behind their association: healthier lifestyle involves eating more vegetables & fruits, besides it may prevent cancer as a whole rather than just the eating style.

Part II. (20 points) The following are scores from an experiment:

3, -2, 0, -1, -4, -4, 2, 2, 5, -2, 1, -2, 0, 0, 4, 2, -4, -3, 5, -2

1. (7 points) Fill in the blanks in the following distribution table.

Score	-4 to -3	-2 to -1	0 to 1	2 to 3	4 to 5
Percentage	20	25	20	20	15
	4/20	5/20	4/20	4/20	3/20

2. (7 points) Is inter-quartile range a measure of center or spread? Find it for the above data set.

Spread.

-4, -4, -4, -3, -2, -2, -2, -1, 0, 0, 0, 1, 2, 2, 2, 3, 4, 5, 5

25th percentile is -2, 75th percentile is 2

$$IR = 2 - (-2) = 4$$

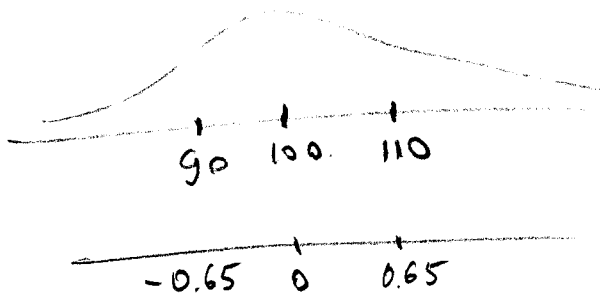
3. (6 points) Here is a table of scores for another data set. Write below the Excel formula to find the average (=mean) for those scores:

$$= B2 * C2 + B3 * C3 + B4 * C4 + \dots + B12 * C12$$

	A	B	C	D	E
1		Score	Percentage (Relative Frequency)		
2		-5	2%		
3		-4	3%		
4		-3	5%		
5		-2	14%		
6		-1	8%		
7		0	10%		
8		1	11%		
9		2	15%		
10		3	19%		
11		4	12%		
12		5	1%		
13		Total	100%		
14					
15					
16					

Part III. (20 points) It is assumed that IQ scores follow a normal distribution with mean 100 and standard deviation 15.4.

1. (6 points) IQ scores between 90 and 110 are designated “ordinary”. What percentage of the people is “ordinary”?



$$\frac{10}{15.4} \approx 0.65$$

Table $\Rightarrow 48.43\%$

2. (7 points) Find the 90th percentile of IQ scores.

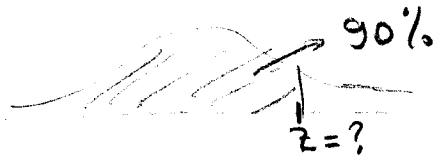
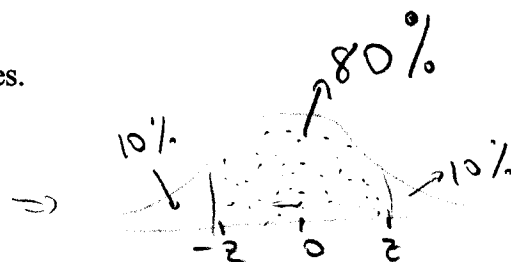


Table $\Rightarrow z = 1.30$

\Rightarrow IQ is



$$100 + 1.30 \times (15.4) \approx 120$$

3. A sample of IQ values are

	A	B	C
1		IQ	
2		95	
3		120	
4		89	
5		134	
6		96	
7		105	
8		114	
9		101	
10		95	
11		108	
12		92	
13		80	
14		88	
15	Average:		
16	Standard Deviation:		
17			

- a) (4 points) Write down a formula to find the average and the standard deviation, any way you like, as long as your procedure gives the answer (either an Excel formula or just hand written operations). DO NOT calculate the final answer.

$$\text{mean} = \text{SUM}(B2:B14)/13$$

$$\text{std. Dev.} = \text{SQRT}((B2-B15)^2 + (B3-B15)^2 + \dots + (B14-B15)^2)/12$$

- b) (3 points) What is the percentage of ordinary people in this sample?

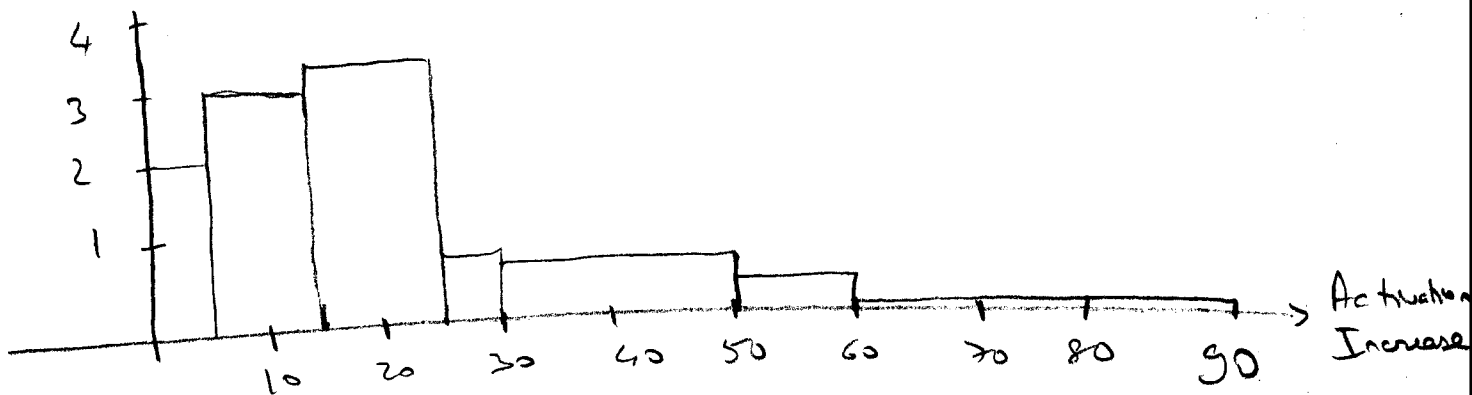
$$\frac{7}{13} = 0.5385 = 53.85\%$$

Part IV. (20 points) A researcher is studying the amygdala (a part of the brain involved in emotion). The subjects are measured for the increase in activation of their amygdala while they are viewing pictures of violent scenes. The distribution of activation increases obtained from a sample is as follows:

Activation Increase	Percentage
0-5	10
5-15	30
15-25	33
25-30	8
30-50	12
50-60	4
60-90	3

$$\begin{aligned} &\rightarrow 10/5 = 2 \\ &\rightarrow 30/10 = 3 \\ &33/20 = 1.65 \\ &8/8 = 1 \\ &12/20 = 0.6 \\ &4/10 = 0.4 \\ &3/30 = 0.1 \end{aligned}$$

1. (10 points) Draw a histogram using the density scale.



2. (5 points) Is the mean (=average) of activation increase smaller or larger than its median? Explain.

Since the histogram has a long tail to the right, the mean is larger than the median. The larger values on the right affect the magnitude of the mean.

3. (5 points) In which interval is the median of activation increase, 15-25, 25-30 or 30-50? Why?

In the interval 15-25, because in previous intervals 0 to 15 there is $10 + 30 = 40\%$ of all data. 50th percentile must be in 15-25 as also the total percentage in 0 to 25 is 73%.

$$(40 < 50 < 73)$$

Part V. (20 points) A survey of young people, namely 18-22 year olds, was published by Sabah last week. Assume that the survey has been conducted by random sampling. One of the questions in the survey is "Do you support death sentence (idam cezası)?" Before taking the survey, the percentage of 18-22 year olds in Turkey who support death sentence was believed to be 28%.

1. (3 points) What is the expected value of the percentage in the sample that support death sentence?

28% , just the population percentage.

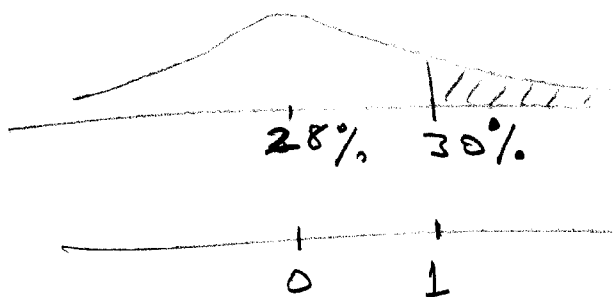
2. (6 points) What should the sample size be so that the standard error of the sample percentage is 2%?

$$S.E. = \sqrt{\frac{(0.28)(0.72)}{n}} = 0.02 \Rightarrow n = \frac{(0.28)(0.72)}{(0.02)^2}$$

$$\Rightarrow n = 504$$

3. (7 points) If 500 young people are surveyed, estimate the chance that the percentage of those who support death sentence is larger than 30%.

$$SE = \sqrt{\frac{(0.28)(0.72)}{500}} \approx 0.02 = 2\%$$



→ Table \Rightarrow in -1 to 1
68%
 $100 - 68 = 32\%$
shaded area $\frac{32}{2} = 16\%$

4. (4 points) Among 500 subjects interviewed, 155 of them responded "YES" to the question.

- a) What is the observed value for the percentage in the sample that support death sentence?

$$\frac{155}{500} = 0.31 = 31\%$$

- b) What is this value called, a statistic or a parameter?

A statistic .

(Remark: This is indeed the observed percentage by Sabah!)

3

(