

**Part I. (20 points)** For studying the relationship between terrorist attacks and the cultural events (theatre, concert, etc.) in a society, several countries in the world are considered. It is agreed that the terrorist attacks picked up after 2015 globally. The variables are

- The number of cultural events during 2013-2014
- The number of cultural events during 2015-2016

The two variables are compared in a random sample of countries over the world.

1. (2 points) Is this an observational study or an experiment?

Observational study, since the cultural events are categorized according to date, not by the experimenter.

2. (4 points) What are the treatment and control groups above?

Treatment: Countries during 2015-2016  
Control: Countries 2013-2014

3. (2 points) Are the controls historical or contemporary?

Historical as 2013-2014 are before 2015-2016.

4. (2 points) Is this a longitudinal study or not? Explain in at most one sentence (any extra sentences will be ignored).

This is a longitudinal study because the same random sample of countries are followed up over time.

5. (5 points) It is claimed that different societies could behave differently at times of crisis. Accordingly, what is a confounding variable in the study above? Explain why in two sentences at most.

"Cultural or social behavior" is a confounding variable. It affects the number of cultural events both before and after 2015.

6. (5 points) How would you control for the confounding variable? Explain in two sentences at most.

By grouping the data according to different countries, or cultural regions of the world. I can analyze each such group separately for events before and after 2015.



**Part II. (20 points)** The following are scores from a psychology experiment, on a scale 0-5.

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

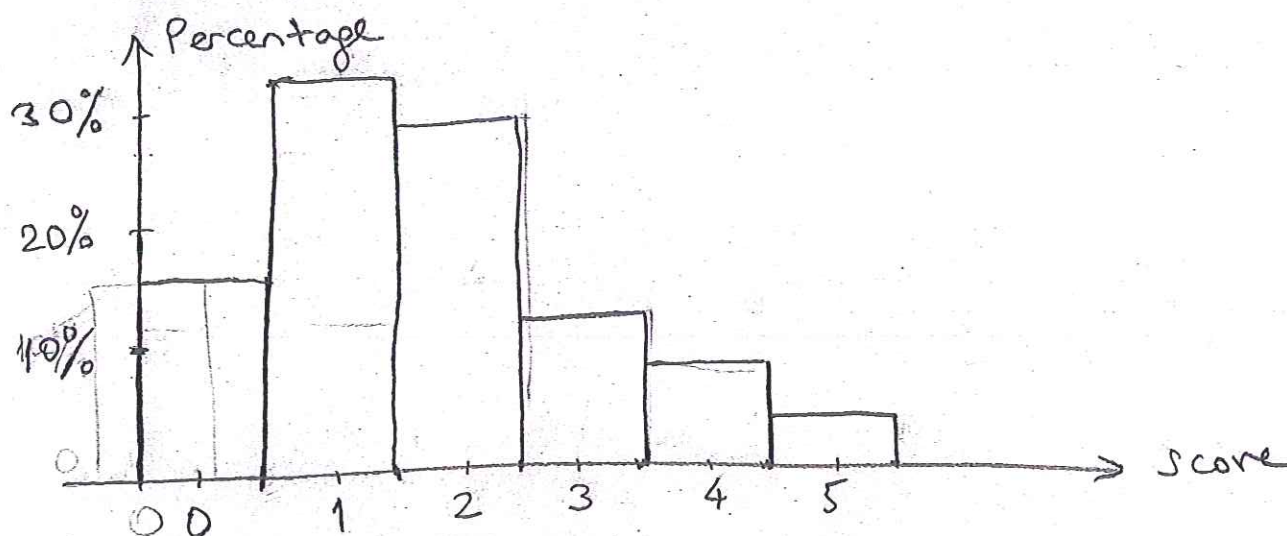
1. (4 points) Fill in the following frequency table.

Observation	0	1	2	3	4	5
Frequency	4	8	7	3	2	1
Rel. freq.	$4/25$ 16%	$8/25$ 32%	$7/25$ 28%	$3/25$ 12%	$2/25$ 8%	$1/25$ 4%

2. (4 points) Find the sample mean using the frequency of each observation.

$$\text{sample mean} = \frac{0 \times 4 + 1 \times 8 + 2 \times 7 + 3 \times 3 + 4 \times 2 + 5 \times 1}{25} = 1.76$$

3. (7 points). Draw a histogram and label both of the axes (either density scale or relative frequency histogram).



4. (2 points) The standard deviation of the data set is 1.3. What percent of the observations is within one standard deviation of the mean?

$$1.76 \pm 1.3 \Rightarrow [0.46, 3.06]$$

Scores 1, 2, 3 are within one SD.

$$8 + 7 + 3 = 18 \text{ out of } 25 : \frac{18}{25} = 72\%$$

5. (3 points) Find the interquartile range.  $\rightarrow$  Order the data first

$$(2)(0.25)(25) = 6.25 \rightarrow 6^{\text{th}} \text{ or } 7^{\text{th}} \text{ observation} : 1$$

$$(0.75)(25) = 18.75 \rightarrow 18^{\text{th}} \text{ or } 19^{\text{th}} : 2$$

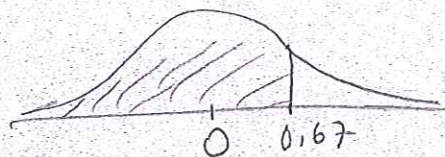
$$Q_1 = 1 ; Q_3 = 2 \Rightarrow \text{Interquartile range} = 2 - 1 = 1$$



**Part III. (20 points)** The time it takes to complete a statistics exam has a normal distribution with mean 70 minutes and a standard deviation 15 minutes at a certain university.

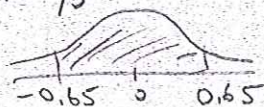
1. (7 points) What is the chance that a randomly student from this university will complete the exam in at most 80 minutes?

$$z = \frac{80 - 70}{15} = 0.67$$



z-table  $\rightarrow$  Area = 48.43 %  
 $0.67 \rightarrow 0.65$

$$50\% + \frac{48.43\%}{2} \approx 74.2\%$$



2. (6 points) If the instructor wishes 99% of the students to complete the exam on time, how many minutes should she allow for the exam?

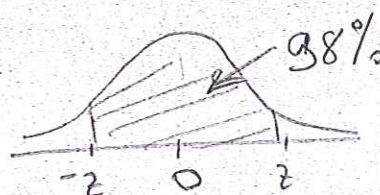
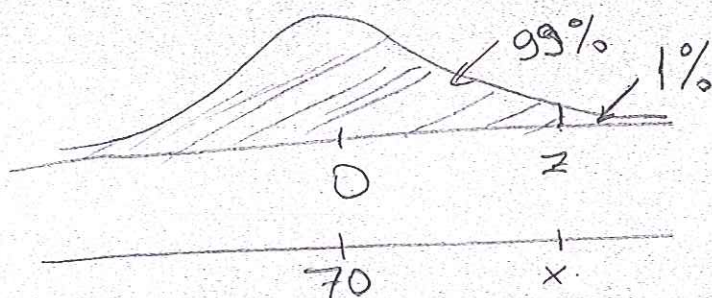


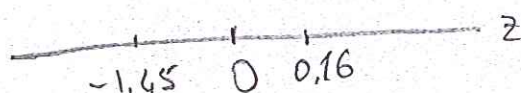
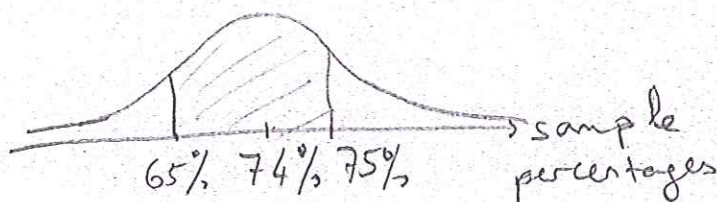
table  $\Rightarrow z = 2.35$

$$\frac{x - 70}{15} = 2.35 \Rightarrow x = 105.25 \text{ min's}$$

3. (7 points) Consider the percentage of students who finish the exam in 80 minutes or less, as you found in Part 1. In a class of 50 students, what is the probability that this percentage will be between 65% and 75%?

$$SE = \sqrt{\frac{(0.74)(0.26)}{50}} = 0.062 = 6.2\%$$

table



$$\frac{65 - 74}{6.2} = -1.45 \rightarrow 85.29\%$$

$$\frac{75 - 74}{6.2} = 0.16 \rightarrow 11.92\%$$

$$\frac{85.29\%}{2} + \frac{11.92\%}{2} = 48.6\%$$



**Part IV. (20 points)** In a newly-established residential complex, the management aims to investigate the usage of the gym, which is available only for the residents. In particular, the management would like to know what proportion uses the gym at least twice a week. For this purpose, a survey is conducted in the café next to the gym by interviewing the customers there that evening.

1. (4 points) What is the population and what is the sample?

Population: All residents of the complex.

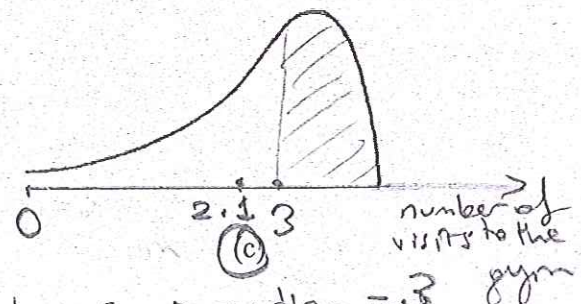
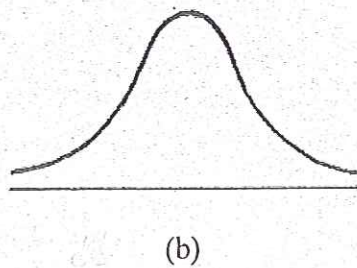
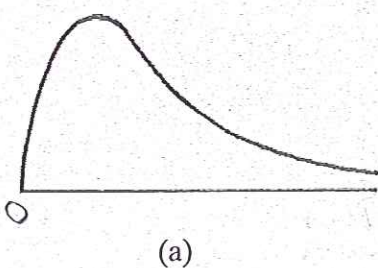
Sample: Customers at the café that evening.

2. (6 points) Define one possible variable in this investigation. Is it quantitative or qualitative? What are its possible values?

1) Number of times a week a resident uses the gym.  
- Quantitative.

2) The gym being used at least twice a week by a resident or not.  
- Qualitative

3. (5 points) It is found that half of the people go to the gym 3 or more times per week in the sample. The average number of times they go to the gym is 2.1. Which one of the following histograms is more plausible do you think? Why? Explain in two sentences at most, also mentioning possible values of the variable in the x-axis.



50% of people go 3 or more times  $\Rightarrow$  median = 3  
mean = 2.1 So mean < median, it is likely that there is a long tail to the left.

4. (5 points) Explain two possible sampling biases that can happen while surveying people, in four sentences at most.

- Selection bias: The sample is not random, but is selected to favor certain groups more than others. (Here, only those that went to the café)

- Noresponse bias: There may be people who refuse to answer the survey.



**Part V. (20 points)** In US elections of November 8, the Republicans won the 241 of the 435 available seats in the House of Representatives, and the remaining 194 seats went to the Democrats.

1. (5 points) What percent of the representatives is Republican in the House of Representatives?

$$\frac{241}{435} = 55.4\%$$

2. (2 points) Is the result you found in Question 1 a parameter or a statistic? Why?

Parameter, because it is from the population - real election result.

3. (2 points) Before the elections, the results of a poll indicated 58% for the percentage of Republicans to be elected for the House of Representatives in a random sample of size 600. Is 58% a parameter or a statistic?

Statistic, because it is a numerical fact from a sample before the election.

4. (9 points) Find the 95% confidence interval for the percentage of Republican seats in the House of Representatives (to be known for sure only after the election), as estimated by the poll mentioned in Question 3.

$$\text{Take } \hat{p} = 0.58$$

$$SE = \sqrt{\frac{(0.58)(0.42)}{600}} = 0.02 = 2\%$$

$$95\% \rightarrow z = 1.95 \quad (\text{or } z \approx 2)$$

$$0.58 \pm (1.95)(0.02)$$

$$[0.54, 0.62] \quad \text{or} \quad [54\%, 62\%]$$

5. (2 points) Does the interval you constructed in Question 4 contain the true percentage or not? Explain in one sentence.

Yes, because 55.4% falls between 54% and 62%.