

(Exercises) E.1

## LESSON 1: EXERCISES

- 1) The Carbo Foods company produced 2000 cupcakes in its first day of operation. We would like to know the average amount of sugar in these cupcakes. Scientists randomly select 50 of the company's cupcakes produced on that first day. The scientists find that the 50 cupcakes have an average of 30 grams of sugar each.
- a) What is the population of interest in this experiment?
  - b) What is the population size,  $N$  ?
  - c) What is the sample in this experiment?
  - d) What is the sample size,  $n$  ?
  - e) Describe the sample results.
  - f) Based on the sample results, what might we infer about the population?

## LESSON 2: EXERCISES

- 1) For each scenario below, determine whether the sampling method used is closest to: simple random sampling, systematic sampling, cluster sampling, or stratified sampling.
- a) A news network interviews 30 registered voters about politics. Ten registered Democrats, ten registered Republicans, and ten "other" registered voters are randomly selected.
  - b) A corporation owns 700 apartment buildings. Ten of those buildings are randomly selected. All of the residents in those ten buildings are surveyed.
  - c) A supermarket gives a prize to every hundredth customer that comes in.
  - d) Each employee in a company randomly receives a free, numbered raffle ticket. Six raffle numbers are randomly selected from a bowl, and the six employees with the six matching numbers win a day off from work.

## LESSON 3: EXERCISES

None, but familiarize yourself with the terminology.

(Exercises) E.2

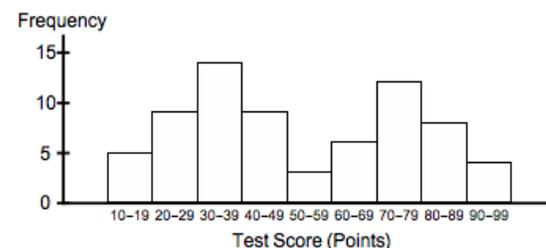
## LESSON 4: EXERCISES

- 1) A six-sided die is tossed 200 times. The frequencies for each die value are listed below. Find the relative frequencies for each die value.

| Die Value | Frequency          | Relative Frequency | Relative Frequency (as a percent) |
|-----------|--------------------|--------------------|-----------------------------------|
| 1         | 22                 |                    |                                   |
| 2         | 32                 |                    |                                   |
| 3         | 44                 |                    |                                   |
| 4         | 34                 |                    |                                   |
| 5         | 38                 |                    |                                   |
| 6         | 30                 |                    |                                   |
|           | Sum = $N$<br>= 200 |                    |                                   |

## LESSON 5: EXERCISES

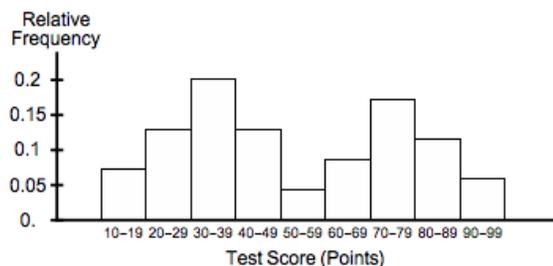
- 1) 70 students take an exam. A frequency histogram for their scores is below.



- a) Estimate the number of students who scored in the 90s (between 90 and 99 points).
- b) Estimate the number of students who scored in the 80s (between 80 and 89 points).
- c) Estimate the number of students who scored in the 70s (between 70 and 79 points).
- d) How many students scored between 10 and 19 points?

(Exercises) E.3

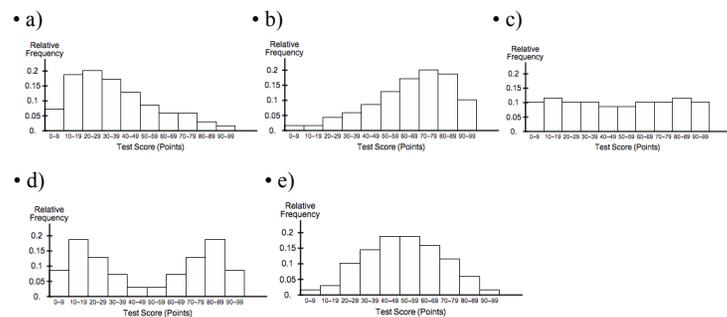
2) Refer to Exercise 1). A relative frequency histogram for the scores is below.



- a) Estimate the relative frequency of scores in the 80s (between 80 and 89 points).
- b) Estimate the relative frequency of scores in the 70s (between 70 and 79 points).
- c) Estimate the relative frequency of scores between 10 and 19 points.
- d) Estimate the relative frequency of scores in the 30s (between 30 and 39 points.)

### LESSON 6: EXERCISES

1) 70 students take an exam. Consider the possible relative frequency histograms for their scores below. Describe each distribution shape.



(Exercises) E.4

### LESSON 7: EXERCISES

1) A drug is designed to lower LDL cholesterol levels in humans. Seven people participate in a study. After eight weeks of using the drug, their LDL cholesterol levels were recorded as follows (in milligrams per deciliter, or mg/dL).

89 137 98 92 104 89 87

Based on this data ...

- a) Find the **mean** of the LDL cholesterol levels among the seven people.
- b) Find the **median position number** of this data set.
- c) Find the **median** of the LDL cholesterol levels among the seven people.
- d) Find the **mode** of the LDL cholesterol levels among the seven people.
- e) Find the **midrange** of the LDL cholesterol levels among the seven people.
- f) Find the **20% trimmed mean** of the LDL cholesterol levels among the seven people. Hint: You would only delete the lowest and highest values in the data set before computing the 20% trimmed mean here.

2) Refer to Exercise 1). Let's say we find that the researchers conducting the study look into the outlier – the “137” recorded in the data set. They find that there was a machine error when that person's LDL cholesterol level was measured. The researchers decide to drop that person from the study. The “137” is deleted from the data set. The modified data set now has these six LDL cholesterol levels (in mg/dL):

89 98 92 104 89 87

Based on this modified data set ...

- a) Find the **mean** of the LDL cholesterol levels among the six people.
- b) Find the **median position number** of this modified data set.
- c) Find the **median** of the LDL cholesterol levels among the six people.
- d) Find the **mode** of the LDL cholesterol levels among the six people.
- e) Find the **midrange** of the LDL cholesterol levels among the six people.
- f) Find the **20% trimmed mean** of the LDL cholesterol levels among the six people. Hint: You would only delete the lowest and highest values in the data set before computing the 20% trimmed mean here.
- g) Compared to Exercise 1), which changes more: the mean or the median?

(Exercises) E.5

- 3) Which measure of center is typically more sensitive to outliers: the mean or the median? (The midrange is very sensitive.)
- 4) Let's say we are analyzing a **left-skewed** distribution. If we consider the mean and the median, which tends to be lower and which tends to be higher?
- 5) Let's say we are analyzing a **right-skewed** distribution. If we consider the mean and the median, which tends to be lower and which tends to be higher?
- 6) In 2004, the George W. Bush administration claimed that the tax cuts they proposed led to an average tax cut of \$1586 for individuals and families. FactCheck.org claimed that the median tax cut was \$470. Could both claims be true? What are possible reasons for the difference?
  - Note 1: The \$1586 figure excludes the 25% of individuals and families who received no tax cut (\$0).
  - Note 2: Many middle-class families did see their income tax drop to \$0.Source: <https://www.factcheck.org/2004/02/here-we-go-again-bush-exaggerates-tax/>
- 7) The **mean** is a very appropriate measure of center for what kinds of distributions?
- 8) What is an advantage of **trimmed means** over the “regular” mean as measures of center?
- 9) The **median** is typically a more appropriate measure of center than the mean for what kinds of distributions?
- 10) Interpret the formula:  $\mu = \frac{\sum x}{N}$
- 11) What does  $\bar{x}$  stand for?
- 12) Find the **median position number** for a sorted population data set if ...
  - a)  $N = 1000$
  - b)  $N = 1001$

(Exercises) E.6

- 13) If the sample size  $n = 11$ , then the **median position number** is 6 for a sorted, sample data set; the formula is  $\frac{n+1}{2}$ , similar to the one for population data. In a survey of annual incomes, 11 American adults are randomly called for interviews. Do we necessarily get the median income for these 11 adults if we take the income of the 6<sup>th</sup> adult that we call? Why or why not?
- 14) A nice and generous professor now thinks that the recent test was too hard. If the professor gives every student in the class 10 extra points on the test, what happens to ...
  - a) ... the **mean** of the test scores for the class?
  - b) ... the **median** of the test scores for the class?
  - c) ... the **mode(s)** of the test scores for the class?
  - d) ... the **midrange** of the test scores for the class?