

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.

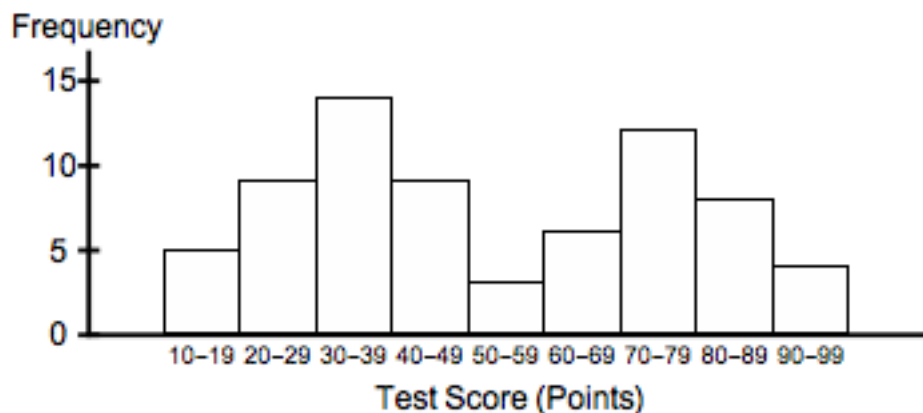
LESSON 4: EXERCISES

- 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 = 200		

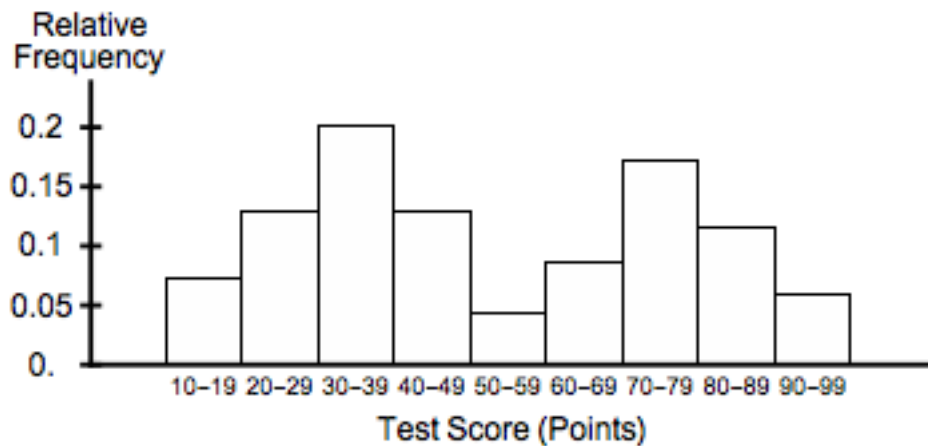
LESSON 5: EXERCISES

- 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?

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

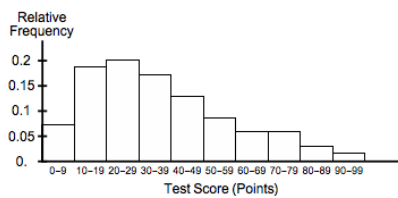


- a) Estimate the relative frequency of scores in the 80s (between 80 and 89 points).
- b) Estimate the relative frequency of scores in the 70s (between 70 and 79 points).
- c) Estimate the relative frequency of scores between 10 and 19 points.
- d) Estimate the relative frequency of scores in the 30s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

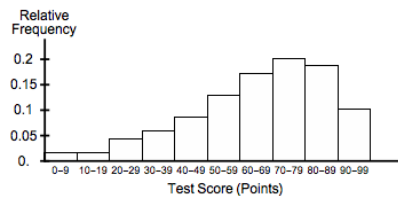
LESSON 6: EXERCISES

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

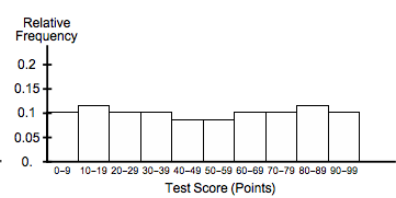
• a)



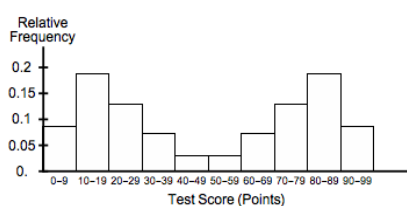
• b)



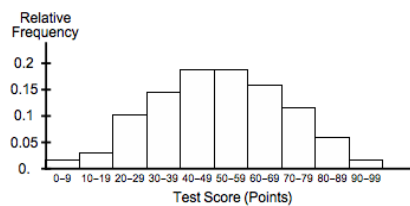
• c)



• d)



• e)



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 98 104 92 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 98 104 92 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?

- 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) $N = 1001$

- 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 6th 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 (and it's possible for students to go over 100%), 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? (Assume that there was a mode or there were modes to begin with.)
 - d) ... the **midrange** of the test scores for the class?

LESSON 8: EXERCISES

1) This frequency table describes the annual incomes of 35 workers in a company:

Income Class	Frequency
\$10,000-\$19,999	8
\$20,000-\$29,999	16
\$30,000-\$39,999	6
\$40,000-\$49,999	5

- a) Estimate the mean annual income of these 35 workers.
- b) Why is the estimated mean closer to \$10,000 than to \$50,000?

2) Dum and Dee are twins who take the same classes at a college.

Here are their grade reports for a term:

Dum's Grade Report			Dee's Grade Report		
Course	Number of Units	Grade	Course	Number of Units	Grade
Psychology	3	C	Psychology	3	D
Economics	4	B+	Economics	4	B
Math	6	A	Math	6	C-

- a) Find Dum's GPA for the term. Round off to two decimal places.
 - b) Find Dee's GPA for the term. Round off to two decimal places.
 - c) Give two reasons why Dee's GPA is so much lower than Dum's.
- 3) All exams in a class are graded out of 100 points. A student gets quiz grades of 65, 83, 80, and 90. The student gets a 92 on the Final. Find the student's weighted class average if the quizzes each count for 15% and the Final counts for 40% of the overall grade.
- 4) So far, your grade record in a class looks like this:

Exam	% of overall grade	Your score (out of 100 points)
Quiz 1	10%	55
Quiz 2	10%	50
Midterm 1	20%	65
Midterm 2	20%	40
Final	40%	c

What must you get on the Final to get at least 70% in the class overall?
(What kind of score do you need c to be?)

LESSON 9: EXERCISES

1) Notation: Population data.

- a) What does N stand for?
- b) What does μ stand for?
- c) What does σ stand for?
- d) What does σ^2 stand for?

2) Notation: Sample data.

- a) What does n stand for?
- b) What does \bar{x} stand for?
- c) What does s stand for?
- d) What does s^2 stand for?

3) What is really the purpose of finding s ?

4) The four students in a class take a test. Their scores in points are as follows:

72 80 91 85

- a) Find the **range** of the population data values.
- b) Use the formula $\mu = \frac{\sum x}{N}$ to find the **population mean**.
- c) Fill out the following table:

Data (x) values	Deviations $(x - \mu)$ values	Squared Deviations $(x - \mu)^2$ values
72		
80		
91		
85		

- d) What do the deviations from the mean add up to?
- e) Use the formula $\sigma^2 = \frac{\sum(x - \mu)^2}{N}$ to find the **population variance**.
The answer will be exact to one decimal place.
- f) Use the formula $\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$ to find the **population standard deviation**. Round it off to one decimal place.

5) 1000 students in a large lecture class take a test, and all the tests are graded. Four of the tests are randomly selected. Their scores in points are as follows:

64 85 71 72

- a) Find the **range** of the sample data values.
- b) Use the formula $\bar{x} = \frac{\sum x}{n}$ to find the **sample mean**.
- c) Fill out the following table:

Data (x) values	Deviations ($x - \bar{x}$) values	Squared Deviations ($x - \bar{x}$) ² values
64		
85		
71		
72		

- d) What do the deviations from the mean add up to?
- e) Use the formula $s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$ to find the **sample variance**.
Round it off to one decimal place, but avoid rounding when doing f) below.
- f) Use the formula $s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$ to find the **sample standard deviation**.
Round it off to one decimal place.

- 6) Assume that women's heights in Fredonia have a population mean of 65 inches and a population standard deviation of 5 inches.
- a) According to **Chebyshev's Theorem**, what proportion of women's heights in Fredonia are between 55 inches and 75 inches?
 - b) According to **Chebyshev's Theorem**, what proportion of women's heights in Fredonia are between 50 inches and 80 inches?
 - c) Let's assume that women's heights in Fredonia are approximately normally distributed. According to the **Empirical Rule**, what proportion of women's heights in Fredonia are between 60 inches and 70 inches?
 - d) Let's assume that women's heights in Fredonia are approximately normally distributed. According to the **Empirical Rule**, what proportion of women's heights in Fredonia are between 55 inches and 75 inches?
 - e) Let's assume that women's heights in Fredonia are approximately normally distributed. According to the **Empirical Rule**, what proportion of women's heights in Fredonia are between 50 inches and 80 inches?
 - f) Use the "Two SD" (2σ) Rule for Usual Values to give an appropriate interval of usual heights for women in Fredonia.
 - g) Based on f), would a 57-inch-tall Fredonian woman have an unusual height?
 - h) Based on f), would a 77-inch-tall Fredonian woman have an unusual height?
- 7) Refer to Lesson 1, Example 1 on presidential ages. The population mean of presidential ages is 55.0 years, and the population standard deviation is about 6.5 years.
- a) Use the "Two SD" (2σ) Rule for Usual Values to give an appropriate interval of usual presidential ages.
 - b) Based on a), would 40 years old be an unusual age for a president?
 - c) Based on a), would 45 years old be an unusual age for a president?
 - d) Based on a), would 65 years old be an unusual age for a president?
 - e) Based on a), would 70 years old be an unusual age for a president?

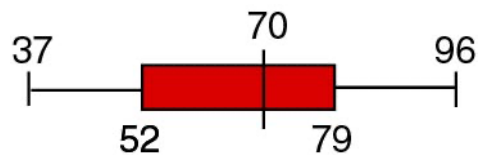
- 8) A nice and generous professor now thinks that the recent test was too hard. If the professor gives every student in the class 10 extra points on the test (and it's possible for students to go over 100%), what happens to ...
- a) ... the **range** of the test scores for the class?
 - b) ... the **standard deviation (SD)** of the test scores for the class?

LESSON 10: EXERCISES

1) Refer to Lesson 9, Exercise 7 on presidential ages. The population mean of presidential ages is 55.0 years, and the population standard deviation is about 6.5 years. Round off the z scores below to two decimal places.

- a) What would be the z score for a 40-year-old president? Interpret it. Would that be unusual, based on the “Two SD” Rule?
- b) What would be the z score for a 45-year-old president? Interpret it. Would that be unusual, based on the “Two SD” Rule?
- c) What would be the z score for a 65-year-old president? Interpret it. Would that be unusual, based on the “Two SD” Rule?
- d) What would be the z score for a 70-year-old president? Interpret it. Would that be unusual, based on the “Two SD” Rule?

2) The scores on a test (in points) in a large class are summarized by the boxplot (also known as a “box-and-whisker” plot) below. The minimum score is 37 points. The maximum score is 96 points. There are no extreme outliers.



- a) A score of 52 points is at which quartile?
- b) A score of 52 points is at which percentile?
- c) A score of 70 points is at which quartile?
- d) A score of 70 points is at which percentile?
- e) A score of 79 points is at which quartile?
- f) A score of 79 points is at which percentile?
- g) What is the median?
- h) What is the range?
- i) What is the IQR (Interquartile Range)?