



**FELLOWSHIP**  
**CHRISTIAN SCHOOL**  
COLOSSIANS 1: 9-12

Entering

# DE Honors Statistics

---

# Summer Math Packet

2019-2020

Students,

This packet is to be completed by the first day of school and will be used as a study guide for the first assessment in the course. Please show all steps when working through the packet.

It is a mistake to do this packet at the beginning of the summer. We want these techniques to be relatively fresh in your mind in the fall. If you work a couple of problems a day, the whole packet will be completed in no time.

As math department, we hope you take this seriously, as we sincerely wish for you to be successful throughout this next year. Your preparation over the summer will be rewarded in unexpected ways during the year.

Here are some helpful websites to use, if needed:

- [www.khanacademy.org](http://www.khanacademy.org)
- [www.patrickjmt.com](http://www.patrickjmt.com)
- [www.youtube.com](http://www.youtube.com) to find specific math related topics with accompanying videos

Sincerely,

Fellowship Math Department

## DE Honors Statistics Summer Packet 2019

Name: \_\_\_\_\_

VOCABULARY- Study & MEMORIZE the following definitions:

- 1) **Statistics** is the science (and art) of learning from data.

### Population vs. Sample

The main difference between a population and sample has to do with how observations are assigned to the data set.

A **population** includes all of the elements or measurements from a set of data.

A **sample** consists one or more observations drawn from the population. A sample is a subset of the population.

Depending on the sampling method, a sample can have fewer observations than the population or the same number of observations. More than one sample can be derived from the same population.

### Parameter vs. Statistic

A measurable characteristic of a population, such as a population mean ( $\mu$ ) or a population standard deviation ( $\sigma$ ), is called a **parameter**; a measurable characteristic of a sample, such as the sample mean ( $\bar{x}$ ) and the sample standard deviation ( $s$ ), is called a **statistic**.

In summary, a measure that describes a population is a parameter, and a measure describing a sample is a statistic. (HINT: Think...  $p \rightarrow p$  and  $s \rightarrow s$ .)

### **EXAMPLE:**

100,000 randomly selected U.S. adults were asked whether they drink at least 48 oz of water each day and only 45% said yes. Identify the sample and population.

### **SOLUTION:**

**Sample:** the 100,000 selected U.S. adults; **population:** all U.S. adults

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

- 2) A \_\_\_\_\_ is the complete collection of all measurements or data collected, whereas, a \_\_\_\_\_ is a subcollection of members selected from the complete collection. 2) \_\_\_\_\_

A) sample; population  
C) population; sample

B) population; parameter  
D) sample; census

Answer the question.

3) A magazine publisher mails a survey to every subscriber asking about the quality of its subscription service. The total number of subscribers represents what? 3) \_\_\_\_\_

A) The population

B) The sample

4) A magazine publisher always mails out a questionnaire six months before a subscription ends. This questionnaire asks its subscribers if they are going to renew their subscriptions. On average, only 3% of the subscribers respond to the questionnaire. Of the 3% who do respond, an average of 40% say that they will renew their subscription. This 3% who respond to the questionnaire are known as what? 4) \_\_\_\_\_

A) The population

B) The sample

5) In a poll of 50,000 randomly selected college students, 74% answered "yes" when asked "Do you have a television in your dorm room?" Identify the sample and population. 5) \_\_\_\_\_

A) Sample: the 74% who answered "yes"; population: all college students

B) Sample: the 50,000 selected college students; population: all college students

C) Sample: the 50,000 selected college students; population: the 74% who answered "yes"

D) Sample: all college students; population: the 50,000 selected college students

6) A computer network manager wants to test the reliability of some new and expensive fiber-optic Ethernet cables that the computer department just received. The computer department received 7 boxes containing 40 cables each. The manager does not have the time to test every cable in each box. The manager will choose one box at random and test 8 cables chosen randomly within that box. What is the population? 6) \_\_\_\_\_

A) 280 cables

B) The 8 cables chosen randomly for testing

C) The one box that was chosen at random from the 7 boxes

D) The 7 boxes

7) A computer network manager wants to test the reliability of some new and expensive fiber-optic Ethernet cables that computer department just received. The computer department received 4 boxes containing 50 cables each. The manager does not have the time to test every cable in each box. The manager will choose one box at random and test 10 cables chosen randomly within that box. What is the sample? 7) \_\_\_\_\_

A) 200 cables

B) The 4 boxes

C) The one box that was chosen at random from the 4 boxes

D) The 10 cables chosen for testing

- 8) The spell-checker in a desktop publishing application may not catch all misspellings (e.g. their, there) or correctly interpret the spellings of proper names. Jackie is an expert editor and can proofread extremely quickly. Jackie is hired by a book publisher to check the spelling of every word in the latest proof of a history book. With regard to Jackie's assignment, what is the population? 8) \_\_\_\_\_
- A) Finding misspellings in the latest proof of the history book  
B) The total number of misspellings that Jackie finds in the latest proof of the history book  
C) Every word in the latest proof of the history book  
D) The latest proof of the history book

SHORT ANSWER. Complete the following short answer questions. If you need more space to write your answer, use a separate sheet of paper.

Provide an appropriate response.

- 9) A group of medical researchers is interested in knowing the mean cholesterol level for all men in the U.S. aged between 70 and 80. They pick a sample of 5,000 men and measure their cholesterol levels. They then calculate the mean and standard deviation of these cholesterol levels. Do the mean and standard deviation obtained in this way represent parameters or statistics? Why? 9) \_\_\_\_\_

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

- 10) Determine whether the given value is a statistic or a parameter. Thirty percent of all dog owners poop scoop after their dog. 10) \_\_\_\_\_
- A) Parameter B) Statistic

Fill in the blank to complete the statement.

- 11) Researchers are interested in learning more about the age of men when they marry for the first time so they survey 500 married or divorced men and ask them how old they were when they first married. The mean of age of the 500 men when they married for the first time would be a \_\_\_\_\_. 11) \_\_\_\_\_
- A) Parameter B) Statistic C) Population D) Sample
- 12) The mean age of all the U. S. vice presidents when they took office would be a \_\_\_\_\_. 12) \_\_\_\_\_
- A) Sample B) Statistic C) Parameter D) Population

VOCABULARY- Study & MEMORIZE the following definitions:

- 13) **Individuals** are the objects described by a set of data. Individuals may be people, but they may also be animals or things.

A **variable** is any characteristic of an individual. A variable can take different values for different individuals.

**EXAMPLE:** The AP Statistics class list that I received from guidance contains information about the students enrolled. The students enrolled are the individuals, and for each individual, there are variables such as student number, gender and home phone number.

There are 2 types of variables:

A **categorical** (or **qualitative**) variable places an individual into one of several groups or categories.

A **quantitative** variable takes numerical values for which arithmetic operations such as adding and averaging make sense.

**EXAMPLE :** The following is a small section of a data set describing education in the US.

State	Region	Population(1000s)	SAT Verbal	SAT Math	% taking	% No HS
CA	PAC	35,894	499	519	54	18.9
CO	MTN	4,601	551	553	27	11.3
CT	NE	3,504	512	514	84	12.5

Using the distribution above, identify the individuals, and then identify the variables. Determine if each variable is categorical or quantitative.

**SOLUTION:** Individuals = States

Variables = Region, Population (1000s), SAT Verbal, SAT Math, % taking, % No HS

Categorical = Region

Quantitative = Population (1000s), SAT Verbal, SAT Math, % taking, % No HS

Sometimes those values are clustered close together, and other times they are spread far apart. When we look at how those values vary, we are looking at the distribution of the variable.

The **distribution** of a variable tells us what values the variable takes and how often it takes these values.

### **Quantitative Variables - Discrete vs. Continuous**

If a quantitative variable can take on any value between two specified values, it is called a **continuous** variable. Continuous variables can be rounded to a certain decimal value, for example, the nearest tenth. Some common continuous variables are: height, weight, temperature, time running a race, & GPA.

**Discrete** variables are ones that CANNOT take on all values within an interval and are often whole number values (i.e. finite or countable values). Some common discrete variables are: number of students in a room, number of days in a certain month, & number of boxes in a warehouse.

**Some examples will clarify the difference between discrete and continuous variables:**

Suppose the fire department mandates that all fire fighters must weigh between 150 and 250 pounds. The weight of a fire fighter would be an example of a continuous variable; since a fire fighter's weight could take on any value between 150 and 250 pounds.

Suppose we flip a coin and count the number of heads. The number of heads could be any integer value between 0 and plus infinity. However, it could not be any number between 0 and plus infinity. We could not, for example, get 2.5 heads. Therefore, the number of heads must be a discrete variable.

**MULTIPLE CHOICE.** Choose the one alternative that best completes the statement or answers the question.

14) The United States Census collects data on many variables about individuals and households. Which variable is categorical? 14) \_\_\_\_\_

- A) monthly mortgage
- B) annual electricity cost
- C) family size
- D) type of residence
- E) hours worked per week

15) What type of data values are quantitative and the number of values is finite or countable? 15) \_\_\_\_\_

- A) Continuous
- B) Categorical
- C) Interval
- D) Discrete

16) Determine whether the given value is from a discrete or continuous data set. The time it takes a computer to complete a task. 16) \_\_\_\_\_

- A) Continuous
- B) Discrete

Classify the data as either discrete or continuous.

17) The average speed of cars passing a busy intersection between 4:30 P.M. and 6:30 P.M. on a Friday is 32.3 mi/h. 17) \_\_\_\_\_

- A) Discrete
- B) Continuous

18) The total number of phone calls a sales representative makes in a month is 425. 18) \_\_\_\_\_

- A) Discrete
- B) Continuous

19) What type of data is provided by the statement "Helen finished in 7th place in the ice dancing competition"? 19) \_\_\_\_\_

A) Discrete

B) Continuous

20) The average height of all freshmen entering college in a certain year is 68.4 inches. 20) \_\_\_\_\_

A) Discrete

B) Continuous

SHORT ANSWER. Write your answers in the space provided. If you need more space to write your answer, use a separate sheet of paper.

21) Distinguish between categorical and quantitative data. Give an example of each.

22) The following table gives the top five movies at the box office this week.

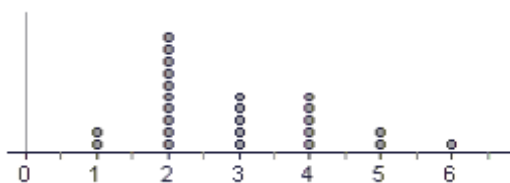
Rank	Last week	Movie title	Studio	Box office sales (\$ millions)
1	N/A	Pirate Adventure	Movie Giant	35.2
2	2	Secret Agent Files	G.M.G.	19.5
3	1	Epic Super Hero Team	22nd Century	14.3
4	5	Reptile Ride	Movie Giant	10.1
5	4	Must Love Cats	Dreamboat	9.9

For each column above, identify whether the variable is categorical or quantitative. For each quantitative variable, determine whether it is discrete or continuous.



MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

23) The distribution below is the number of family members reported by 25 people in the 2010 Census. 23) \_\_\_\_\_

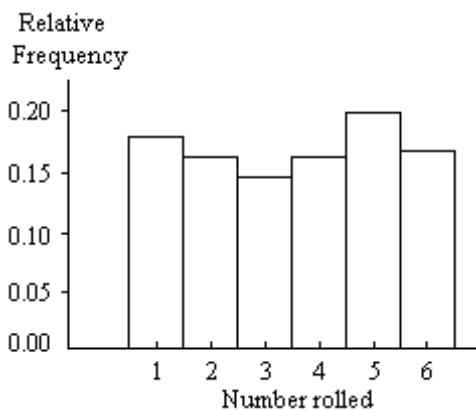


The best description for the shape of this distribution is

- A) uniform
- B) bell-shaped
- C) skewed left
- D) skewed right
- E) bimodal

A graphical display of a data set is given. Identify the overall shape of the distribution as (roughly) bell-shaped, uniform, right skewed, left skewed, or bimodal.

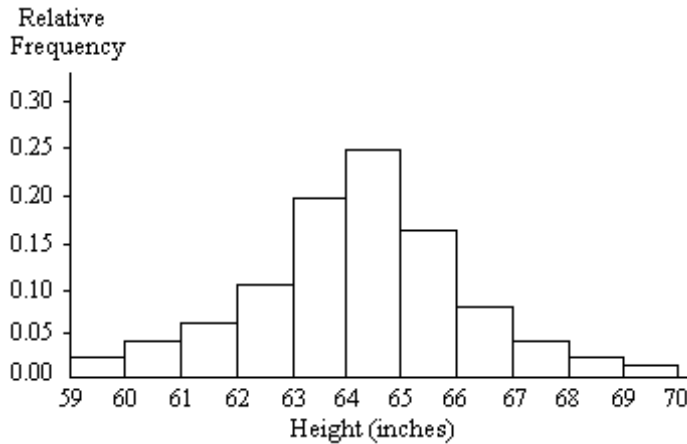
24) A die was rolled 200 times and a record was kept of the numbers obtained. The results are shown 24) \_\_\_\_\_  
in the relative frequency histogram below.



- A) Right skewed
- B) Left skewed
- C) Bell-Shaped
- D) Uniform

25) A relative frequency histogram for the heights of a sample of adult women is shown below.

25) \_\_\_\_\_



A) Left skewed

B) Bell-shaped

C) Right Skewed

D) Uniform

Find the average (mean) for the given sample data. Unless otherwise specified, round your answer to one more decimal place than that used for the observations.

26) 13, 15, 11, 13, 10

26) \_\_\_\_\_

A) 11

B) 15.5

C) 12.4

D) 13

Find the median for the given sample data.

27) 3, 7, 18, 21, 30, 30, 49

27) \_\_\_\_\_

A) 25.5

B) 18

C) 21

D) 30

Find the mode(s) for the given sample data.

28) 86, 62, 32, 62, 29, 86

28) \_\_\_\_\_

A) 62

B) 86, 62

C) 59.5

D) 86

29) The blood types for 30 people who agreed to participate in a medical study were as follows.

29) \_\_\_\_\_

O A A O A AB O B A O  
A O A B O O O AB A A  
A B O A A O O B O O

Find the mode of the blood types.

A) O, A

B) 13

C) A

D) O

Find the range for the given data set.

30) 27, 37, 16, 43, 58

30) \_\_\_\_\_

A) 58

B) 10

C) 42

D) 16

VOCABULARY- Study & MEMORIZE the following definitions:

31) **Descriptive statistics** uses the data to provide descriptions of the population, either through numerical calculations or graphs or tables.

**Inferential statistics** makes inferences and predictions about a population based on a sample of data taken from the population in question.

**EXAMPLE:** A news article appearing in a national paper stated that "The fatality rate from use of firearms sank to a record low last year, the government estimated Friday. But the overall number of violent fatalities increased slightly, leading the government to urge an increase in police forces in major urban areas. Overall, 15,600 people died from violent crimes in 2005, up from 15,562 in 2004, according to projections from a government source. Is the figure 15,600 a descriptive statistic or an inferential statistic? Is the figure 15,562 a descriptive statistic or an inferential statistic?

**SOLUTION:** The figure 15,600 is an **inferential statistic** since it is indicated in the statement that it is a projection (probably based on incomplete data for the year 2005). The figure 15,562 is a **descriptive statistic** since it reflects the actual number of deaths from violent crimes for the year 2004.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Provide an appropriate response.

32) A researcher randomly selects a sample of 100 students from the students enrolled at a particular college. She asks each student his age and calculates the mean age of the 100 students. It is 21.3 years. Based on this sample, she then estimates the mean age of all students enrolled at the college to be 21.3 years. In what way are descriptive statistics involved in this example? In what way are inferential statistics involved?

32) \_\_\_\_\_

A) When calculating the mean age of the students in the sample, the researcher is using descriptive statistics. When estimating the mean age of all students at the college, the researcher is using inferential statistics.

B) When calculating the mean age of the students in the sample, the researcher is using inferential statistics. When estimating the mean age of all students at the college, the researcher is using descriptive statistics.

33) The table below shows the number of homicides in the U.S. in each of the years 1989-1993.

33) \_\_\_\_\_

Murder and non-negligent manslaughter	
Year	Number of offenses
1989	21,500
1990	23,440
1991	24,700
1992	23,760
1993	24,530

Classify the study as either descriptive or inferential.

A) Descriptive

B) Inferential

34) Based on a random sample of 1000 people, a researcher obtained the following estimates of the percentage of people lacking health insurance in one U.S. city.

34) \_\_\_\_\_

Age	Percentage not covered
18-24	28.2
25-39	24.9
40-54	19.1
55-65	16.5

Classify the study as either descriptive or inferential.

A) Descriptive

B) Inferential

VOCABULARY- Study & MEMORIZE the following definitions:

### 35) The Interquartile Range (IQR)

The interquartile range rule is useful in detecting the presence of outliers. Outliers are individual values that fall outside of the overall pattern of the rest of the data. The interquartile range shows how the data is spread about the median. It is less susceptible than the range to outliers.

To find the first quartile, Q1, representing a quarter of the way through the list of all the data, find the median of the data, and then find the median of the bottom half of the data.

To find the third quartile, Q3, representing three quarters of the way through the list of all the data, find the median of the data, and then find the median of the top half of the data.

These five numbers can be used to tell us quite a bit about our data. For example, the range, which is just the minimum subtracted from the maximum, is one indicator of how the data is spread out. The data set is.

Similar to the range, but less sensitive to outliers, is the interquartile range. The interquartile range is calculated in much the same way as the range. All that we do is subtract the first quartile from the third quartile:  **$IQR = Q3 - Q1$** .

### Interquartile Rule for Outliers (1.5 IQR Rule)

Calculate the interquartile range for our data

Multiply the interquartile range (IQR) by the number 1.5

Add  $1.5 \times (\text{IQR})$  to the third quartile. Any number greater than this is a suspected outlier.

Subtract  $1.5 \times (\text{IQR})$  from the first quartile. Any number less than this is a suspected outlier.

It is important to remember that this is a rule of thumb and generally holds. In general, we should follow up in our analysis. Any potential outlier obtained by this method should be examined in the context of the entire set of data.

### EXAMPLE:

We will see this interquartile range rule at work with a numerical example. Suppose we have the following set of data: 1, 3, 4, 6, 7, 7, 8, 8, 10, 12, 17. Use the 1.5 IQR rule to determine whether or not 17 is an outlier

**SOLUTION:** The five number summary for this data set is minimum = 1, first quartile = 4, median = 7, third quartile = 10 and maximum = 17. The interquartile range is:  $Q3 - Q1 = 10 - 4 = 6$

We now multiply by 1.5 and have  $1.5 \times 6 = 9$ . Nine less than the first quartile is  $4 - 9 = -5$ . No data is less than this. Nine more than the third quartile is  $10 + 9 = 19$ . No data is greater than this. Despite the maximum value being five more than the nearest data point, the interquartile range rule shows that it should NOT be considered an outlier for this data set.

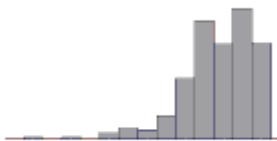
### **\*\*IMPORTANT: Which measures of center and spread should I report?**

The IQR is a useful measure of spread because quartiles are much less affected by outliers for a skewed data set than the equivalent measures of mean and standard deviation. For this reason, the IQR is often reported along with the median as the best choice of measure of spread and central tendency (center), respectively, when dealing with skewed and/or data with outliers. If the shape of the distribution of data is roughly symmetric or bell-shaped, then report the mean as the measure of center and the standard deviation as the measure of spread.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

36) Which is true of the data whose distribution is shown?

36) \_\_\_\_\_



- I. The distribution is skewed to the right.
- II. The mean is probably smaller than the median.
- III. We should summarize with median and IQR.

- A) II and III
- B) I, II, and III
- C) I only
- D) I and II
- E) II only

SHORT ANSWER. Complete the following short answer questions. If you need more space to write your answer, use a separate sheet of paper.

- 37) Testgen questions still do not copy to other applications. Auto insurance The Insurance Institute for Highway Safety publishes ratings for all models of vehicles to compare the relative risk of payouts. 100 is the mean rating for all vehicles. A rating of 122 means the vehicle is 22% worse than average. The table shows the summary statistics for the collision ratings of 27 midsize cars. 37) \_\_\_\_\_

Min	57
Q1	99
Median	109
Q3	122
Max	173
Mean	110.9
SD	23.99

a. Were any of the ratings outliers? Show how you made your decision.

b. A histogram of the data is shown below. Is it more appropriate to use the mean and standard deviation, or the median and IQR to describe these data? Explain.

