HOUSTON HIGH SCHOOL

2026-2027 Pre-Term Work

Course Name	**AP Statistics**
Teacher Name and Contact Email	**Ashton Collins** **ashton.collins@gmsdk12.org**
Pre-Term Work Assignment Description	**Introduction to AP Statistics**
Purpose of Pre-Term Assignment	**I. To brief you in the information that must be understood from your mathematics classes for you to be successful in this course.** **II. To introduce you to some of the statistic capabilities of the Numworks calculator.** **III. To teach you some of the basic concepts that we will see in the first few units of the course.**
Assignment Due Date	**First Day of Class**
How will this assignment be submitted to the teacher?	**Printed and physical copy turned in.**
How will this assignment work be assessed (include rubric, if applicable)?	**This packet will be graded two ways:** **1. Assignment Grade for Completion of the Packet. (Due the First Day of School)** **2. Quiz Grade. I will give you a quiz of 10 random questions from the packet on the second day of school.**
Additional Resources:	***** For all Numworks Calculator Steps ***** **1. You can purchase a Numworks calculator (Highly Suggested!): https://www.numworks.com/buy/** **2. You can download the Numworks phone APP. 3. You can use the online emulator: https://www.numworks.com/simulator/**

Name:_____

AP Statistics Introduction

Directions:

1. The purpose of this packet is:

   I. To brief you in the information that must be understood from your mathematics classes for you to be successful in this course.

   II. To introduce you to some of the statistic capabilities of the Numworks calculator.

   III. To teach you some of the basic concepts that we will see in the first few units of the course.

2. This packet should take a few days of work to complete. Please give yourself plenty of time and do not wait until the last minute.

3. This packet will be graded two ways:

   1. **Assignment Grade** for Completion of the Packet. (Due the First Day of School)

   2. **Quiz Grade**. I will give you a quiz of 10 random questions from the packet on the second day of school.

4. All Work must be Shown. There will be a lot of writing in this course. Please make sure everything is neat and legible.

5. Place answers in the spaces and boxes provided.

6. **Turn in Packet on the First Day of School!**

*** For all Numworks Calculator Steps ***

1. You can purchase a Numworks calculator (Highly Suggested!): https://www.numworks.com/buy/

2. You can download the Numworks phone APP.

   Apple Store: https://apps.apple.com/us/app/numworks-graphing-calculator/id1456585807

   Google Play: https://play.google.com/store/apps/details?id=com.numworks.calculatorc

3. You can use the online emulator: https://www.numworks.com/simulator/

Activity 1: Data Types

There are two types of data: ***qualitative*** (or categorical) and ***quantitative***.

Qualitative variables or ***categorical variables*** are variables that categorize individuals (place them in groups). These variables may take on values that are labels for categories. Examples are eye color (blue, hazel, etc.), gender (male or female), method of transportation to school (bike, car, bus, etc.), class rank (senior, junior, etc.).

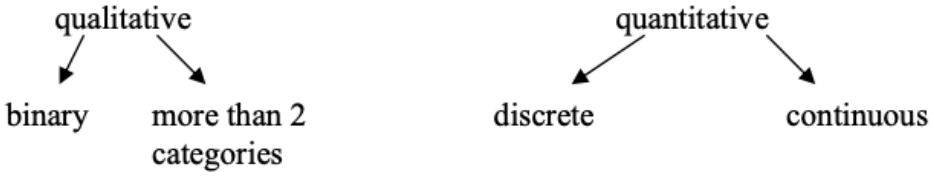
A specific type of qualitative variable is a ***binary variable***. A binary variable is a qualitative variable that has only two outcomes. Examples include gender, approve or disapprove of the president's handling of the war in Iraq, outcome of a coin toss, outcome of a die roll (when restricted to a four or not a four), the response to the question "Do you play basketball?"

Quantitative variables are numerical variables that represent an amount or quantity. There are two kinds of these: ***discrete*** and ***continuous***.

Discrete variables are quantitative variables that assume only a countable number of values. Examples of these include shoe size (..., $6, 6\frac{1}{2}, 7, 7\frac{1}{2}$, ...), score on a test, class size, number of cans collected for MIFA.

Continuous variables are quantitative variables that can assume an infinite number of values. In the case of continuous variables, the values can generally assume any decimal quantity within a small range of values (even though we may round the answer like when we measure our height). *These are, typically, values that result from some kind of measurement*. The units of measurement are pounds/ inches/ Kelvin/ degrees/ feet/ etc. Examples are height, weight, surface area of oranges, ERA in baseball (3.23, 2.78, etc.), GPA.

Just because your variable's values are numbers, don't assume that it's quantitative. For example, Social security number is a numerical output that is not a quantitative variable. SSN doesn't stand for any type of numerical quantity (you are not the 412,327,642 person born in the US!). Phone number is not a quantitative variable either. The 901 area code is a designation for a geographic region; it is not a numerical quantity.



Answer the following questions and then decide if the variable in question is qualitative or quantitative. Then decide if it is also binary, discrete, or continuous.

Question	Answer	Type
1. In which year did you take Algebra I?		
2. What is you zip code?		
3. Choose a random integer from 1 to 20.		

4. How many siblings do you have?		
5. Who is your favorite math teacher?		
6. How tall are you (in inches)?		
7. Where did you eat your last meal? (1=home, 2=restaurant, 3=other)		
8. How many AP classes will you be taking this year?		
9. How far away from school do you live?		
10. How many miles per gallon does you vehicle get while driving in the city?		
11. What time is it?		
12. How fast can you run “the 40”?		

Activity 2: Numerical Descriptions of Quantitative Data

There are two categories of numbers that are used to describe a set of data: measures of center and measures of spread.

Measures of Center:

1. The ***mean*** is the average number. It is the sum of all the data values divided by the number (n) of values.

Ex. {4, 36, 10, 22, 9}

$$Mean = \bar{x} = \frac{\sum x_i}{n} = \frac{4+36+10+22+9}{5} = 16.2$$

2. The ***median*** is the value that separates the bottom 50% of data from the top 50% of data. It is the middle element of an ordered set of data that is odd in number. It is the average of the two middle elements of an ordered set of data that is even in number.

Ex. {4, 36,10,22, 9} = {4, 9, **10**, 22, 36,} ⟶ median = 10

{4, 36,10,22, 9, 43} = {4, 9, 10, 22, 36, 43 } ⟶ median = $\frac{10+22}{2} = 16$

3. The ***mode*** is the value that occurs most often in a set of data. If the data occurs with the same frequency, then there is no mode. If two (or more) values occur the most then they are both the mode.

Measures of Spread:

1. The ***range*** is a measure of the spread of the entire data. It is calculated by subtracting the minimum value from the maximum value.

Ex. {4, 36,10,22, 9} = {4, 9, 10, 22, 36,} ⟶ range = 36-4 =32

2. The ***interquartile range*** (IQR) is a measure of the spread of the middle 50% of the data. It is calculated by subtracting the ***25th percentile*** (Q1) from the ***75th percentile*** (Q3). Q1 is the median of the lower half of the data. It separates the bottom 25% of values from the top 75% of values. Q3 is the median of the upper half of the data. It separates the top 75% of values from the bottom 25% of values. In neither of these cases is the median considered in the top half or the bottom half of the data.

Ex. {4, 36,10,22, 9} = {4, 9, **10**, 22, 36,} IQR = 29-6.5 =22.5

Q1=6.5 Q3=29

Ex. {4, 36,10,22, 9, 43} = {4, 9, 10, 22, 36, 43} IQR = 36-9 =27

Q1=9 Q3=36

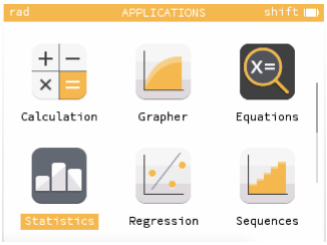
3. The ***standard deviation*** is the measure of the spread around the mean. It is calculated using:

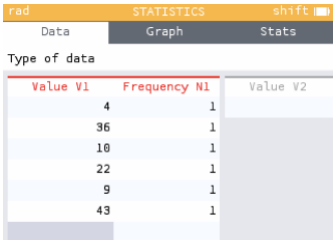
$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

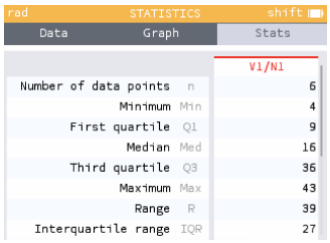
Ex. {4, 36,10,22, 9, 43} $$s = \sqrt{\frac{(4-16.2)^2 + (26-16.2)^2 + \ldots + (9-16.2)^2}{5-1}} = 12.89$$

This means that the typical number differs from the mean by about 12.89 units. The smaller the standard deviation the closer the data should be clustered around the mean.

To see statistical results including quartiles and standard deviation we can us the Statistics App on the Numworks calculator:







Use this link to watch tutorials for the Statistics APP: https://www.numworks.com/calculator/tutorials/statistics/

Here is a list of parent's ages at the time their children were born.

Dad:	41	27	23	31	30	33	26	32	43	25	34	27	25	34
	27	26	28	32	32	35	27	33	34	34	34	35		
Mom:	39	26	23	30	28	33	23	32	38	23	35	24	24	33
	24	23	24	32	23	30	24	29	34	35	26	31		

Using you calculator find the following:

13. Find the mean and median for the Dad data:

Mean _____ Median _____ Which is larger? _____

14. Find the mean and median for the Mom data:

Mean _____ Median _____ Which is larger? _____

15. Now compare the two means you calculated. Which is larger? _____

Is this what you expected? _____. Explain why or why not.

_____

_____

16. Calculate the standard deviations for both sets of data: Dad_____ Mom_____

Why might these values be different? Explain.

_____

17. Find Q_1 and Q_3 and the IQR for the Dad data. Q_1_____ Q_3_____ IQR_____

18. Find Q_1 and Q_3 and the IQR for the Mom data. Q_1_____ Q_3_____ IQR_____

19. A company has two machine that fills cans of soft drinks. Samples from each machine show the following number of ounces per can:

Machine A: 11.1, 12.0, 11.4, 12.1, 11.7, 11.5, 12.2, 11.4, 11.3, 11.9

Machine B: 10.9, 12.4, 12.7, 11.8, 12.3, 11.9, 12.0, 12.5, 12.7, 11.6

Find the mean and standard deviation for both machines.

x_A =_____ s_A =_____ x_B =_____ s_B =_____

20. Using you answer to #19, explain which machine is" better" at filling soft drink cans.

_____

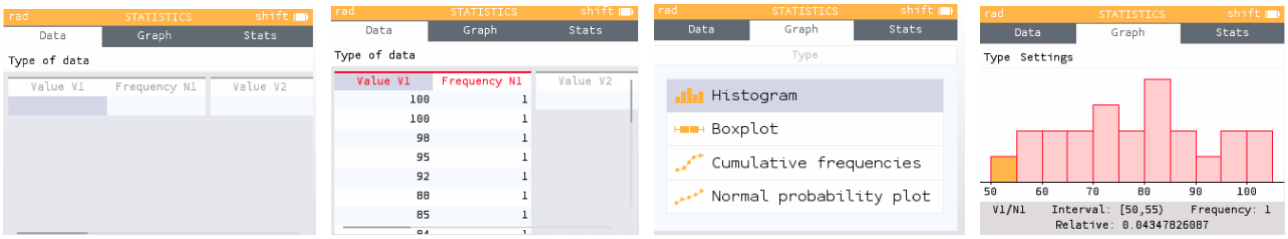
_____

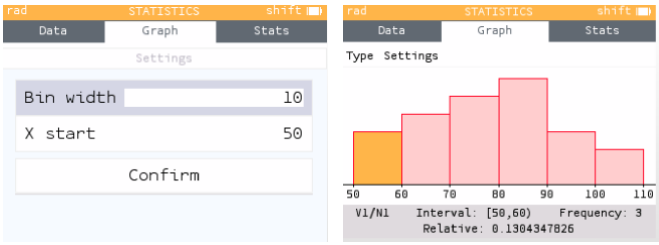
Activity 3 Graphical Displays of One Variable Data

A set of quantitative data is often separated into groups or intervals. For example, test scores may be separated into 90-99, 80-89, and so on. The frequency refers to the count for each interval. A histogram is a visual representation of the frequencies plotted against the interval.

Example 1:

Enter the following test scores into a list and create a histogram using intervals of width 10: 100, 100, 98, 95, 92, 88, 85, 84, 84, 81, 80, 77, 75, 72, 70, 70, 66, 65, 63, 60, 59, 55, 50



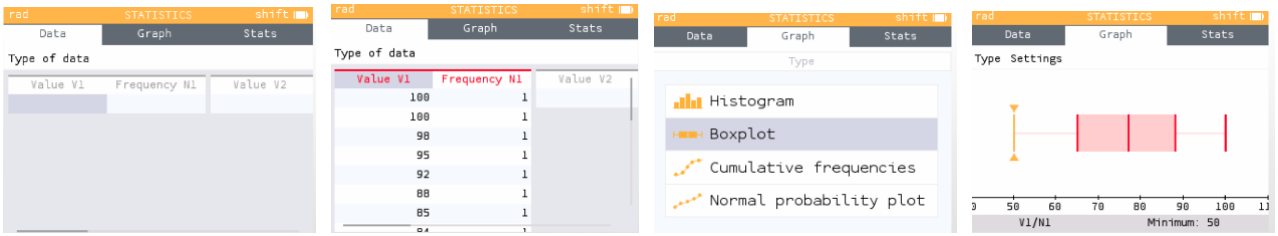


A ***box plot*** is a graphical display for a set of quantitative data that only displays the ***five number summary***:

1- **minimum**
2- **Q1** or **first quartile** (the data point that separates the bottom 25% of scores from the top 75%)
3- **median** (the data point that separates the bottom 50% of scores from the top 50%)
4- **Q3** or **third quartile** (the data point that separates the bottom 75% of scores from the top 25%)
5- **maximum**

Example 2:

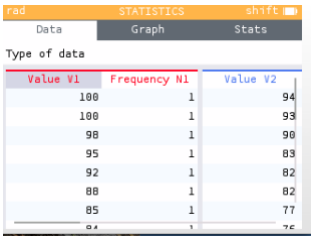
Create a box plot of the test scores.

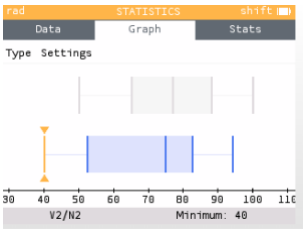


Box plots are particularly effective for comparison of two or more sets of data.

Example 3:

Suppose another class receives these test scores: 94, 93, 90, 83, 82, 82, 77, 76, 75, 70, 64, 60, 53, 52, 48, 44, and 40. Create two box and whiskers plots to compare their performance with the original group.



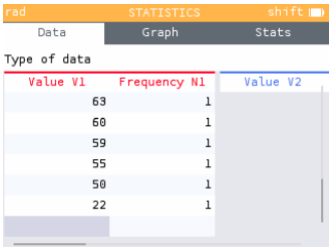


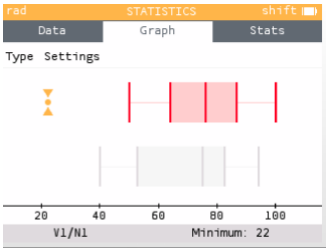
It appears as though the first set of scores has a higher median and a higher maximum score. It also appears that the two sets of scores have approximately the same range (about 50) but the second set has a larger IQR (30 compared to 23).

In AP Statistics a "modified box plot" is used. This shows any ***outliers***. An outlier is defined to be a data point that does not fit the pattern of the rest of the data. When your calculator graphs a modified box plot an algorithm is used to determine what it takes to "not fit the pattern of the rest of the data". This algorithm is 1.5⊕IQR away from the "box" part of the graph (above or below). These outliers are shown with dots or stars or any other small symbol.

Example 4:

Suppose one student from the original class takes the test late and scores 22. Create a box plot for the original class.





21. Using the histogram from example 1.

A. Change the interval size (Bin Width) to 5 then graph. Describe how this histogram differs from the first histogram. Which histogram would best describe the data?

B. Then go back and change the interval size (Bin Width) to 15 then GRAPH. Describe how this histogram differs from the first two. Again, state which histogram would best describe the data.

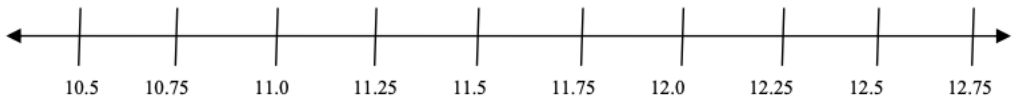
C. From these two problems what conclusions can you draw about how the interval affects the histogram?

22. A company has two machines that fill cans of soft drinks. Samples from each machine show the following number of ounces per can:

Machine A: 11.1, 12.0, 11.4, 12.1, 11.7, 11.5, 12.2, 11.4, 11.3, 11.9

Machine B: 10.9, 12.4, 12.7, 11.8, 12.3, 11.9, 12.0, 12.5, 12.7, 11.6

Create a box plot for each machine. Sketch one above the other and place them in the empty space here with appropriately labeled and scaled axes:



Find the range for: Machine A: _____ Machine B: _____

Find the median for: Machine A: _____ Machine B: _____

Which machine is more "consistent"? Why? _____

Say that Machine A gives one more sample with a fill of 10 oz. Is this an outlier? _____

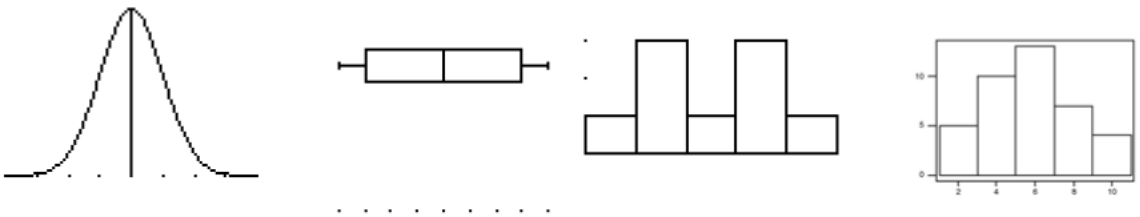
Say that Machine B gives one more sample with a fill of 10 oz. Is this an outlier? _____

Activity 4: Assessing the Shape of a Graph

When describing a set of data we look at the following features:
Shape center spread clusters and gaps outliers

We have several terms that we use to describe the shape, but this packet will concentrate on only two: ***symmetric*** and ***clustered***.

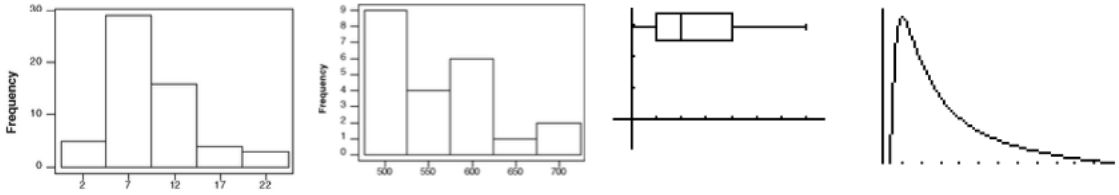
One can tell if a graph is ***symmetric*** if a vertical line in the "center" divides the graph into two fairly congruent shapes. The following sets of data can be described as symmetric



The mean and the median are approximately the same in a symmetric set of data.

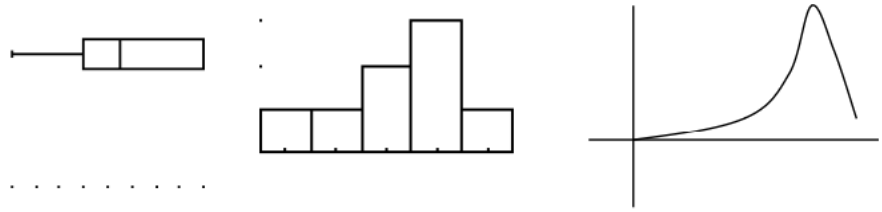
One can tell if a graph is ***skewed*** if the graph has a big clump of data on either the left (skewed right) or the right (skewed left) with a tendency to get flatter and flatter as the values of the data increase (skewed right) or decrease (skewed left). A common misconception is that the "skewness" occurs at the big clump; skewness follows the direction of the "tail." The following sets of data can be described as skewed:

skewed right:



The mean is larger than the median in a skewed right set of data. The mean is always further along the "tail".

skewed left:

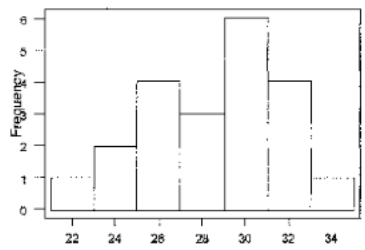


The mean is smaller than the median in a skewed right set of data. The mean is always further along the "tail".

For the following graphs, find the shape, center (**just find the median**), and spread (**find only the range**). If there any other notable features evident in the graph (clusters, gaps, or outliers), then say where they are. Otherwise do not comment on clusters, gaps or outliers.

(Note: To find the center of these graphs, use the frequencies found on the y-axis. Count how many are in each bar. Add these up and divide by two. This tells you where the median is located. Which bar is this value in? That's the median. For graph A, n = 21, so the middle value is 10.5. Starting with the first bar count 1 + 2 + 4 + 3 + 6... So the median is in the bar that contains the 10.5 value (bigger than 10 anyway). That's 30. So, the median is 30.

23.



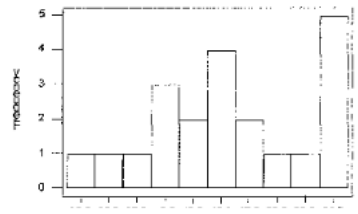
Shape_____

Center_____

Spread_____

Clusters?_____

24.



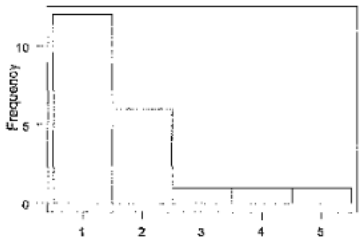
Shape_____

Center_____

Spread_____

Clusters?_____

25.



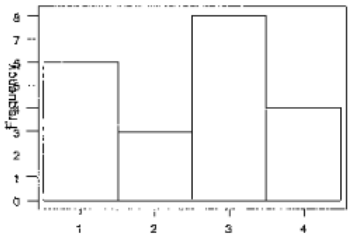
Shape_____

Center_____

Spread_____

Clusters?_____

26.



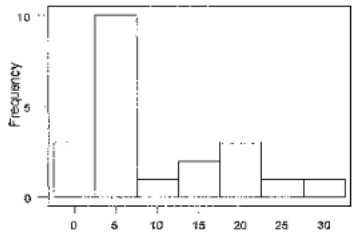
Shape_____

Center_____

Spread_____

Clusters?_____

27.



Shape_____

Center_____

Spread_____

Clusters?_____