**Statistics - PART 1**

**Categorizing Data**

Data - facts and statistics collected together for reference or analysis.

**Type of Data :**

Data can appear in several forms:  
   • Data values can be numbers, referred to as quantitative data.  
   • Data values can be names or labels, referred to as qualitative data.  
   • Data values can be numbers which act as names instead of numbers

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | Quantitative Data | | • Deals with numbers. • Also referred to as Numerical Data. • Data which can be measured. • Height, weight, area, volume, length, time, temperature, speed, cost, etc. • Quantitative → Quantity | | |  |  | | --- | --- | | ***Example 1:***  **Candy Bar** | snickers |   *Quantitative Data:*   |  | | --- | | • weight 1.83 ounces • 280 calories • length 10 cm • width 3 cm • height 1.8 cm | | | |  | | --- | | Qualitative Data | | • Deals with names, labels, descriptions. • Also referred to as Categorical Data. • Data which can not measured. • Eye color, smells, car models, textures, tastes, favorites, candy bars, etc. • Qualitative → Quality | | |  |  | | --- | --- | | ***Example 1:***  **Candy Bar** | snickers |   *Qualitative Data:*   |  | | --- | | • dark chocolate • contains peanuts • caramel smell • brown wrapper • nougat center | | |

**Number of Variables in Data:**

Univariate data means "one variable" (one type of data).  
Bivariate data means "two variables" (two types of data).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | Univariate Data | | • Deals with one variable. • Major purpose is to describe. • No relationships or causes. | | *Statistical Analysis:* • measures of central tendency - mean, mode, median • outliers and interquartile range • range, maximum, minimum, variance, quartiles, mean absolute deviation, standard deviation • shape, center, spread or distributions | | *Displays:* • Dot Plots • Histograms • Box Plots *•*Quartiles | | *Example:* How many students in the freshman class own a skateboard? | | |  | | --- | | Bivariate Data | | • Deals with two variables. • Major purpose is to explain. • Relationships and causes. | | *Statistical Analysis:* • correlations • comparison, causes, relationships, explanations • analysis of 2 variables simultaneously • tables showing one variable depending upon the other variable • independent and dependent variables | | *Displays:* • Two-Way Frequency Tables • Scatter Plots • Line of Best Fit • Linear/Quadratic Regressions | | *Example:*Is there a relationship between the number of skateboards a freshman owns and his/her final test score in Algebra 1? | |

**Representing Data Graphically**

Let's take a look at the same set of data displayed using all three of these graphical methods.

Data set to be used for ALL graphs: (Ratings from 1-10 on the TV program "Psych".)

9, 3, 10, 4, 6, 5, 7, 3, 6, 5, 2, 4, 6, 7, 5, 10, 4, 1, 7, 3, 5, 7, 4, 3, 9, 5, 2, 3, 9, 5

|  |  |
| --- | --- |
| Histogram: Histograms subdivide data into intervals (bins), and use rectangles (usually columns) to show the frequency (count) of observations in each interval. The choice of the size of the intervals may vary. stathistguy |  |

You cannot determine if a specific value is in the data set by looking only at this histogram. The dot plot will show that the "8" is missing. The median of the data displayed in a histogram is not readily seen, as it is in a box plot

|  |  |
| --- | --- |
| Dot Plot: Dot plots include ALL values from the data set, with one dot for each occurrence of an observed value from the set. psychdotguy |  |

Dot plots work well for small sets of data, but become difficult to construct for large data sets. A histogram or box plot will deal more efficiently with large data sets. Dot plots show all values in the set. The median, however, is not readily seen, as it is in the box plot

|  |  |
| --- | --- |
| Box and Whisker Plot: Box plots show a "five statistical summary" of the data set, dividing the data into quarters (25%). From left to right on the diagram: minimum, first quartile, median (or second quartile), third quartile, and maximum. Outliers, when present, are shown as a separate dot or asterisk. psychguy3 | The median is readily seen in this box plot. Box plots work well for large data sets. You cannot, however, determine if a specific value is in the data set by looking only at this box plot. |

**Shapes of Distributions**

When graphed, the data in a set is arranged to show how the points are distributed throughout the set. These distributions show the spread (dispersion, variability, scatter) of the data. The spread may be stretched (covering a wider range) or squeezed (covering a narrower range).

The shape of a distribution is described by its number of peaks and by itspossession of symmetry, its tendency to skew, or its uniformity. (Distributions that are skewed have more points plotted on one side of the graph than on the other.)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PEAKS:Graphs often display peaks, or local maximums. It can be seen from the graph that the data count is visibly higher in certain sections of the graph.  **1.** one clear peak is called a *unimodal* distribution. **2.** two clear peaks are called a *bimodal*distribution. (Here, the term "mode" is used to describe a local maximum in a chart (such as the midpoint of the a peak interval in a histogram). It does not necessarily refer to the most frequently appearing score, as in the "central tendency mode". | |  | | |
| **3.** single peak at the center is called *bell shaped*distribution. *Note:* A bell shaped graph (bell curve), is a frequency distribution that resembles the outline of a bell when plotted on a graph. | | bellz | | |
| Histograms | Dot Plots | | Box Plots |
| *Note:*The graphs shown below demonstrate the *shapes* of various sets of data. The histogram, dot plot and box plot in each separate section represent the same data set. | | | |
| Symmetric (bell shaped) - when graphed, a vertical line drawn at the center will form mirror images, with the left half of the graph being the mirror image of the right half of the graph. In the histogram and dot plot, this shape is referred to as being a "bell shape" or a "mound". The *most typical* symmetric histogram or dot plot has the highest vertical column in the center. This shape is often referred to as being a "normal curve" (or normal distribution). Not all symmetric graphs, however, have this shape (see Symmetric U-shaped below). | | | |
|  |  | |  |
| Symmetric (U-shaped) - as mentioned above, a symmetric graph forms a mirror image of itself when reflected in its vertical center line. Unlike the previous graphs, these histograms and dot plots have more of a U shape. | | | |
|  |  | |  |
| Skewed Right (positively skewed) - fewer data plots are found to the right of the graph (toward the larger numeric values). The "tail" of the graph is pulled toward higher positive numbers, or to the right. The mean typically gets pulled toward the tail, and is greater than the median. (There may be exceptions to the this last statement.) | | | |
|  |  | |  |
| Skewed Left (negatively skewed) - fewer data plots are found to the left of the graph (toward the smaller numeric values). The "tail" of the graph is pulled toward the lower or negative numbers, or to the left. The mean typically gets pulled toward the tail, and is less than the median. (There may be exceptions to the this last statement.) | | | |
|  |  | |  |
| Uniform - The data is spread equally across the range. There are no clear peaks in these graphs, since each data entry appears the same number of times in the set. Notice in the boxplot how each section is of equal length: min to Q1, Q1 to median, median to Q3, and Q3 to max. These graphs are also symmetric. | | | |
|  |  | |  |

**Measures of Center**

A measure of central tendency (measure of center) is a value that attempts to describe a set of data by identifying the central position of the data set (as representative of a "typical" value in the set).

Graphically speaking, the center of a distribution is located at the median of the distribution. The median is the point where half of the data points are found on its left side and half on its right side. While the median indicates the "center", it may not always represent the most typical value in the data set. Let's see which measures of center represent the most typical values of the data given various situations.

|  |  |  |
| --- | --- | --- |
| Which measures of center are representative of the most typical values in the data set? | | |
| Distribution | Central Tendency | Typical Graph |
| Bell-shaped Symmetric Distribution | The mean, mode and median will be the same value.  • Best measure of center: mean |  |
| Distribution Skewed Right | Typically has mean > median There may be exceptions to this statement.  • Best measure of center: median |  |
| Distribution Skewed Left | Typically has mean < median There may be exceptions to this statement.  • Best measure of center: median |  |
| U-shaped Symmetric Distribution | Neither the mean nor the median is a good indicator of typical values in the set.  • Best measure of center: midrange - but the*mean* as a "balance point" is also a descriptor of the center of the distribution |  |

**Population vs Sample Data**

When dealing with statistical data, it is important to distinguish between  
"population" data sets and "sample" data sets.

A population data set contains all members of a specified group (the entire list of possible data values). [Utilizes the count n in formulas.]  
Example: The population may be "ALL people living in the US."

A sample data set contains a part, or a subset, of a population. The size of a sample is always less than the size of the population from which it is taken. [Utilizes the count *n* - 1 in formulas.]  
*Example:* The sample may be "SOME people living in the US."

When calculating the formulas for mean absolute deviation (MAD), variance, and standard deviation, it is important to know if you are working with an entire population (where you have all of the possible data), or if you are working with only a sample (a part) of the data. In addition, if you are using a sample of the data, you need to know if you will be making generalizations about the entire population, based upon this sample.

|  |  |
| --- | --- |
| formula1 | formula2 |
| formula3 | |

Note: When working with "sample data sets", statisticians use n for the number of data entries and meansmall for the mean, however, when working with "population data sets", they use N for the number of data entries and meanother for the mean

**Variance and Standard Deviation**

Variance measures how far a set of data is spread out. A variance of zero indicates that all of the data values are identical. All non-zero variances are positive.

A small variance indicates that the data points tend to be very close to the mean, and to each other. A high variance indicates that the data points are very spread out from the mean, and from one another. Variance is the average of the squared distances from each point to the mean.

Thee process of finding the variance is very similar to finding the MAD, mean absolute deviation.The only difference is the squaring of the distances.

Process: (1) Find the mean (average) of the set.

(2) Subtract each data value from the mean to find its distance from the mean. (3) Square all distances.

(4) Add all the squares of the distances.

(5) Divide by the number of pieces of data (for population variance).

One problem with the variance is that it does not have the same unit of measure as the original data. For example, original data containing lengths measured in feet has a variance measured in square feet.

|  |  |
| --- | --- |
| formula2 | formulakey1 |

Standard deviation shows how much variation (dispersion, spread, scatter) from the mean exists. It represents a "typical" deviation from the mean. It is a popular measure of variability because it returns to the original units of measure of the data set.

A low standard deviation indicates that the data points tend to be very close to the mean. A high standard deviation indicates that the data points are spread out over a large range of values.  
The standard deviation can be thought of as a "standard" way of knowing what is normal (typical), what is very large, and what is very small in the data set.

Standard deviation is a popular measure of variability because it returns to the original units of measure of the data set. For example, original data containing lengths measured in feet has a standard deviation also measured in feet

|  |  |
| --- | --- |
| sdformula | formulakey1 |

|  |  |
| --- | --- |
| **bullet** To compute standard deviation by hand: The standard deviation is simply the square root of the variance. This description is for computing population standard deviation. If sample standard deviation is needed, divide by *n* - 1 instead of *n*. Since standard deviation is the square root of the variance, we must first compute the variance. | |
| **1.** Find the mean. | mean12 |
| **2.** Subtract the mean from each data value and square each of these differences (*the squared differences*). | sd2 |
| **3.** Find the average of the squared differences (add them and divide by the count of the data values). This will be the variance. | sd3 variance |
| **4.** Take the square root. This will be the population standard deviation. Round the answer according to the directions in the problem. | sd4 standard deviation |

**Normal Curve :**

A normal curve is a symmetric, bell-shaped curve. The center of the graph is the mean, and the height and width of the graph are determined by the standard deviation. When the standard deviation is small, the curve will be tall and narrow in spread. When the standard deviation is large, the curve will be short and wide in spread. The mean and median have the same value in a normal curve.

|  |  |
| --- | --- |
| Normal Curve Empirical Rule: *Approximately* ...  • 68% of the data lie ithin one standard deviation of the mean.  • 95% of the data lies ithin two standard deviations of the mean.  • 99.7% of the data lies within three standard deviations of the mean.  IQR for a normal curve is 1.34896 x standard deviation. | normalgraphe |

**Measures of Spread**

**(May also be called: Measures of Variability,Measures of Dispersion, or Measures of Scatter)**

A measure of spread (variability, dispersion, scatter) refers to how the data within the set is "spread out" (or "dispersed", or "scattered") about the mean

If the data is clustered around the center value, the "spread" is small.  
The further the distances of the data values from the center value, the greater the "spread".

|  |  |
| --- | --- |
|  |  |

1.Range:

The first method of measuring "spread" of a data set that you learned was finding the range. Range is the differene between the largest data value and the smallest data value in the set. While the range is simple to compute, it is often unreliable as a measure of variability. The range is based on only two values within the set, which may tell very little about "how" the remaining values are distributed in the set. For this reason, range is used as a supplement to other measures of spread, instead of being the only measure of spread.

  
This range of 43 tells us very little about how the data in this set is scattered. The range alone cannot tell us, for example, if the data is clustered to one end of the set, or if there is an outlier in the data set.

2.Interquartile Range (IQR):

The interquartile range is another form of range which divides the set into four equal parts (or quarters). The three values that form the four divisions are called quartiles: first quartile, Q1; second quartile (median), Q2, and third quartile, Q3. The interquartile range is the difference between the third quartile and the first quartile. You can think of the IQR (also called the *midspread* or *middle fifty*) as a "range" between the third and first quartiles. The IQR is considered a more stable statistic than the typical range of a data set, as seen in the first section. The IQR contains 50% of the data, eliminating the influence of outliers.

|  |  |  |
| --- | --- | --- |
| rangequartile   |  |  | | --- | --- | | IQRformula | quartile key | |

3.Mean Absolute Deviation (MAD):

The mean absolute deviation is the average (mean) of the absolute value of the differences between each piece of data in the data set and the mean of the set. It measures the average distances between each data element and the mean.  
*Process:*(1) Find the mean (average) of the set. (2) Subtract each data value from the mean to find its distance from the mean. (3) Turn all distances to positive values (take the absolute value). (4) Add all of the distances. (4) Divide by the number of pieces of data (for population MAD).

|  |  |
| --- | --- |
| formula1 | formulakey1 |

4.Variance:

The variance is the average of the squared differences from the mean. A small variance indicates that the data points tend to be very close to the mean and to each other A high variance indicates that the data points are very spread out from the mean and from each other. One problem with the variance is that it does not have the same unit of measure as the original data. For example, original data containing lengths measured in feet has a variance measured in square feet.

*The process is very similar to finding the MAD. The only difference is the squaring of the distances.  
Process:* (1) Find the mean (average) of the set. (2) Subtract each data value from the mean to find its distance from the mean. (3) Square all distances. (4) Add all the squares of the distances. (4) Divide by the number of pieces of data (for population variance).

|  |  |
| --- | --- |
| formula2 | formulakey1 |

*Note:* the notation used to represent "variance" is actually the square of the notation for standard deviation. The notation always reminds us of the relationship betweeen these two quantities.  
  
varSDnotation

5.Standard Deviation:

The standard deviation is the average distance of each data point from the mean. The standard deviation is the square root of the variance. Taking the square root will return the same units as expressed in the original data, thus eliminating this problem as found with variance. Now, original data containing lengths measured in feet has a standard deviation measured in feet.  
A low standard deviation indicates that the data points tend to be very close to the mean. A high standard.deviation indicates that the data points are spread out over a large range of values.

*The process is simply square rooting the variance. The process for finding variance is shown above.  
Process:* Find the square root of the variance

|  |  |
| --- | --- |
| sdformula | formulakey1 |

Which methods work best?

|  |  |
| --- | --- |
| Symmetrical data spread is best summarized by: MAD, variance, and Standard Deviation FYI: The standard deviation is affected more by values with large deviations from the mean (more spread) than is the MAD. | Skewed data spread is best summarized by: range and IQR |

**Comparing Statistical Graphs**

Let's take a look at ways to compare statistical data by examining their graphs.

Remember:

When comparing graphs of data, you should consider the Shape, any possible Outliers, the Center, and the Spread of each distribution.  
"Remember to look at each graph's SOCS!"

ex1

|  |
| --- |
| TASK: Prepare box plots and compare the following test scores: Kyle's Scores: 85, 76, 92, 88, 75, 90, 80, 88 Amira's Scores: 91, 84, 86, 76, 71, 82, 80, 90 |

|  |  |
| --- | --- |
| • Prepare a box plot for each student This can be done by hand or with a graphing calculator. • A few of the possible findings: **(S)(S)** Both graphs are slightly skewed left, with the data on the left of the medians being more spread out. The data to the right of the medians are closer together. **(O)** There are no outliers in either plot, which says neither student really "bombed" a test. **(C)** Kyle's median score is higher than Amira's median score (86.5 vs 83). **(S)**The range of Kyle's scores is 92-75=17 and the range of Amira's scores is 91-71=20. Kyle's scores are slightly more consistent. **(other observations)** | comparebox11 |
| • Kyle's lowest (75) and highest (92) scores are both better (higher) than Amira's scores (71 and 91). **•**50% of Kyle's scores are above 86.5, whereas 50% of Amira's scores are above 83. • Conclusion: You can conclude that Kyle's scores are slightly better than Amira's scores. | |

ex2

|  |
| --- |
| TASK: Sven is preparing for a 15 mile Snowshoe Marathon. Prepare a histogram **and** a box plot and compare what is being shown. Miles covered during each practice run: 2, 2, 3, 4, 5, 5, 6, 6, 7, 5, 8, 9, 2, 10, 11, 4, 12, 2, 13, 4, 15, 5, 11, |

|  |  |
| --- | --- |
| • Prepare a histogram and box plot This can be done by hand or with a graphing calculator. • A few of the possible findings: **The histogram:** •**(S)**shows peaks in the data, telling the number of times Sven repeated the runs in each interval (not seen in the box plot). • gives a clearer picture of the number of miles spent in each practice run interval (not seen in box plot). **The box plot:** • **(C)** clearly shows the median (middle) of the data (more calculation would be needed to determine this value from the histogram). | marathonn |

|  |
| --- |
| **• (S)**shows that 50% of the lengths of the runs falls between 4 miles and 10 miles(more calculation needed to obtain this information from the histogram). **Both graphs:** • **(S)** show that the data is skewed to the right, indicating that most of the practice runs fell far short of the 15 miles needed for the marathon. • **(O)** neither graph shows the presence of outliers. • Conclusion: You can conclude that the majority of Sven's practice runs are far less than the actual 15 miles he will need to traverse for the upcoming marathon. Perhaps these are his "beginning" practice runs, with more extended practice runs to follow. If not, it would appear that Sven is not ready for the marathon. |

ex2

|  |
| --- |
| TASK: Kate and Rick submit mileage reports every two weeks as seen by the graphs of their reports. Compare what is being shown. |

|  |  |  |
| --- | --- | --- |
| Kate's Report |  | Ricks Report |
| • Some findings from Kate's Report: **• (S)**the graph is skewed to the left due to the outlier. |  | • Some findings from Rick's Report: **• (S)** the graph is symmetric (a mirror image of itself centered on 40). |
| •**(O)**the 0 mileage is an outlier. The interquartile range is 10, and 0 is less than the first quartile (40) minus 1.5 x IQR (0 < 25). |  | **• (O)**the data is clustered around 40 and there are no outliers. |
| • **(C)** The median is 45. The mode is 45. The mean is 41 (the value of the mean is being pulled to the left by the outlier)**.** |  | • **(C)** The median is 40. The mode is 40. The mean is 40. The measures are all the same due to the symmetry of the graph. |
| **• (S)**The range is 50. The interquartile range is 10. The population standard deviation is 14.11. Kate's standard deviation is larger because her data spread is larger due to the outlier. |  | **• (S)** The range is 20. The interquartile range is 10. The population standard deviation is 5.48. |
| Conclusion:Their employer notices that Kate averaged 41 miles per day, and Rick averaged 40 miles per day, concluding their efficiency to be similar. We know that the outlier is affecting Kate's report. What we do not know is whether Kate's report minus the outlier would be typical of her daily mileage, making her average daily mileage 45.6, or whether she increased her mileage over 9 days since she had no mileage for one day. We also do not know if more mileage is viewed favorably by their employer. | | |

Practice :

**Practice with Box Plots**

**1.** Jason saves a portion of his salary from his part-time job in the hope of buying a used car. He recorded the number of dollars he was able to save over the past 15 weeks.

Dollars saved: 19, 12, 9, 7, 17, 10, 6, 18, 9, 14, 19, 8, 5, 17, 9

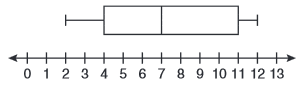
Which box and whisker plot represents this data?

Choose:

Top of Form

|  |  |  |
| --- | --- | --- |
|  | jason1 | jason2 |
|  | jason3 | jason4 |

Bottom of Form

2.  Examine the box plot: 

**(a)** What is the value of the third quartile?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 2 | 4 | 7 | 11 |

Bottom of Form

**(b)** What is the median score?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 2 | 4 | 7 | 11 |

Bottom of Form

**(c)** What percent of the values are *less than 4*?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 25 | 50 | 75 | 100 |

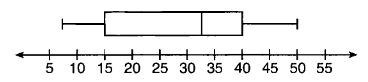
Bottom of Form

**(d)** What is the range of the data presented in the box plot?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 2 | 7 | 10 | 12 |
| **3.**  A movie theater recorded the number of tickets sold daily for a popular vampire movie during the month of December. The box plot below shows the data for the number of tickets sold, in hundreds. | | | | |
| boxticketmoviebox Determine whether the following statements are TRUE or FALSE regarding this box plot. | | | | |
| |  |  |  | | --- | --- | --- | |  | TRUE | FALSE | | 25% of the data is between 300 and 400 | Top of Form   |  |  | | --- | --- | |  |  |   Bottom of Form | | | the second quartile is 600 | Top of Form   |  |  | | --- | --- | |  |  |   Bottom of Form | | | the median is 400 | Top of Form   |  |  | | --- | --- | |  |  |   Bottom of Form | | | the first quartile is 300 | Top of Form   |  |  | | --- | --- | |  |  |   Bottom of Form | | | the range is 300 | Top of Form   |  |  | | --- | --- | |  |  |   Bottom of Form | | | 50% of the data is between 300 and 600 | Top of Form   |  |  | | --- | --- | |  |  |   Bottom of Form | | | | | | |

**4.** The box plot shows the ages of people attending a music concert.



**(a)** Which interval contains exactly 50% of the ages?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 7-15 | 15-33 | 33-40 | 15-40 |

Bottom of Form

**(b)** What percentage of the ages are 15 or older?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 25 | 35 | 75 | 85 |

Bottom of Form

**(c)** What is the median age?

Top of Form

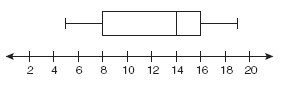
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 25 | 30 | 33 | 35 |

Bottom of Form

Ans : a - (15-40) area betweenQ1 And Q3, b – 75( q1 to maximum), c – 33 ( Q2)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **5.** The box plot shows the lengths of cell phone calls made by three students in a one week period.  Which of the following statements is true? Choose:  Top of Form   |  |  | | --- | --- | |  | All of Nikki's calls were longer than all of Paula's calls. | |  | Paula made the fewest number of calls. | |  | Mario made 50% more calls than Paula. | |  | The median length of Mario's calls is equal to the median length of Nikki's calls. |   Bottom of Form | cellbox |

Ans : last option

**6.** Examine the box plot:  

**(a)** Which interval represents the interquartile range, IQR?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 8-14 | 14-16 | 8-16 | 16-19 |

Bottom of Form

**(b)** What percentage of the data lies in the IQR?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 25 | 50 | 75 | 100 |

Bottom of Form

**(c)** Which interval could be described as being the most "clustered" section of the data?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 5-8 | 8-14 | 14-16 | 16-19 |

**Practice with Outliers**

**1.** Given data set: -3, 5, 10, 12, 14, 18, 24, 26, 49, 60

Which of the following statements is true regarding this data?  
Choose:

Top of Form

|  |  |
| --- | --- |
|  | The value -3 is the only outlier. |
|  | The value 60 is the only outlier. |
|  | No outliers exist. |
|  | Multiple outliers exist. |
|  |  |

Bottom of Form

**2.** The following temperatures were recorded (in Fº) each day for two weeks. 82, 72, 83, 75, 80, 78, 82, 73, 60, 79, 80, 78, 83, 81

**(a)** What is the outlier in this data, if one exists?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 84 | 83 | 60 | no outliers |

Bottom of Form

**(b)** What is the mean for this set of data?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 75 | 77.6 | 78.9 | 79.5 |

Bottom of Form

**(c)** What is the mean for this set of data, if the outlier is removed?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 75 | 77.6 | 78.9 | 79.5 |

Bottom of Form

**(d)** Which of the box plots gives the best representation of this data set?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Top of Form  Choose:   |  |  | | --- | --- | |  | Box A | |  | Box B | |  | Both give the same information. | |  | Neither represent this data. |   Bottom of Form | outlier1 |

**(e)** Which term best describes this data?  
Choose:

Top of Form

|  |  |  |
| --- | --- | --- |
|  | symmetric | skewed right |
|  | skewed left | uniform |

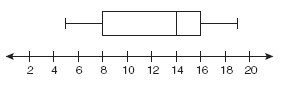
|  |
| --- |
| **3.**The data set shown below has an outlier. Determine the outlier and then answer the questions as to what happens to the median, mean, mode, range and standard deviation when the outlier is removed.  Data: 29, 19, 35, 27, 21, 48, 23, 12, 24, 26, 20, 28, 30, 22, 19, 32, 22 |
|  |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | If the outlier is excluded, what happens to: | |  |  |  | | --- | --- | --- | | Increase | Decrease | No effect | | | the median? | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | the mean? | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | the mode? | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | the range? | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | the standard deviation? | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | |

**4.**  Which measure of central tendency is most affected by outliers?

Choose:

Top of Form

|  |  |
| --- | --- |
|  | mean |
|  | median |
|  | mode |

**6.** Given the box plot: 

**(a)** What is the first quartile, Q1, value?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 5 | 8 | 14 | 16 |

Bottom of Form

**(b)** What is the third, Q3, value?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 5 | 8 | 14 | 16 |

Bottom of Form

**(c)** What is the interquartile range, IQR?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 5 | 8 | 14 | 16 |

Bottom of Form

**(d)** What is the upper outlier fence, Q3 + (1.5 • IQR) ?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 16 | 20 | 24 | 28 |

Bottom of Form

**(e)** *True* or *False*: The value 28 is an outlier for this data set.

Top of Form

|  |  |  |
| --- | --- | --- |
| Choose: | TRUE | FALSE |

**Measures of Center and Shapes of Distributions**

|  |  |
| --- | --- |
| 1. The number of snowboarding accidents reported weekly during one winter season at the Fun Mountain Resort:        12, 15, 6, 8, 12, 17, 10, 8, 7, 13, 14, 16, 8, 18, 11  (6,7,8,8,8,10,11,12,12,13,14,15,16,17,18 = ASC) |  |

**(a)**Find the mean.

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 11.7 | 12 | 12.7 | 12.8 |

Bottom of Form

**(b)**Find the median.

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 11 | 12 | 13 | 14 |

Bottom of Form

**(c)**Find the mode.

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | 8 | 11 | 12 | no mode |

Bottom of Form

**(d)** A teenager wants to emphasize the safety of snowboarding. Which measure of central tendency should the teen use?

Top of Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Choose: | mean | median | mode |  |

Bottom of Form

**(e)** A parent wants to emphasize the dangers of snowboarding. Which measure of central tendency should the parent use?

Top of Form

|  |  |  |  |
| --- | --- | --- | --- |
| Choose: | mean | median | mode |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **1.**  Questions pertain to the histogram shown below. | | | Top of Form  **(a)** How many students are in the Fencing Club?   |  |  |  | | --- | --- | --- | | Choose: | 17 | 18 | |  | 19 | 20 |   Bottom of Form  Top of Form  **(b)**How many students' heights are greater than or equal to 66 inches tall?   |  |  |  | | --- | --- | --- | | Choose: | 5 | 6 | |  | 7 | 11 |   Bottom of Form | Fencing Club | | **(c)** What percentage of the students have a height greater than or equal to 60 inches but less than 66 inches, to the *nearest tenth*of a percent?  Top of Form   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Choose: | 44.4% | 55.2% | 66.7% | 80.0% |   Bottom of Form | | | **(d)** In which interval, will you find the median student height, *h*? Choose:  Top of Form   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | 60 ≤ *h* < 62 | 62 ≤ *h* < 64 | 64 ≤ *h* < 66 | 66 ≤ *h* < 68 |   Bottom of Form | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **2.**  The box and whisker plots below represent the number of points scored in each game for the 2013 fall football season for two area high school teams. | | | Top of Form  **(a)** In general, which team scored more points per game?   |  |  |  | | --- | --- | --- | | Choose: | Team A | Team B |   Bottom of Form  **(b)**Which team's data shows more variability in the points scored?  Top of Form   |  |  |  | | --- | --- | --- | | Choose: | Team A | Team B |   Bottom of Form  **(c)** What is the difference between the median points scored by these two teams?  Top of Form   |  |  |  | | --- | --- | --- | | Choose: | 4 | 5 | |  | 8 | 9 |   Bottom of Form | practicecomparebox1 football kid2 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **3.**  In the same morning, fifteen people spend money on $5 lottery tickets at a local gas station. The dot plot shows the amount spent by each customer on the tickets. | | | Top of Form  **(a)** What is the median of the dollars spent?   |  |  |  | | --- | --- | --- | | Choose: | $20 | $30 | |  | $40 | $45 |   Bottom of Form | pracdotplot | |  | |  | | **(b)** What is the mean (average) of the dollars spent?  Top of Form   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Choose: | $20 | $30 | $31 | $40 |   Bottom of Form  **(c)** What monetary value represents the first quartile of this data?  Top of Form   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Choose: | $15 | $25 | $40 | $45 |   Bottom of Form | | | **(d)** What is the interquartile range?  Top of Form   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Choose: | 15 | 25 | 30 | 45 |   Bottom of Form | | | **(e)** What is the population standard deviation for this data (to the *nearest hundredth*)?  Top of Form   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Choose: | 12.46 | 13.96 | 14.05 | 14.54 |   Bottom of Form | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **4.**  The box and whisker plots shown below represent monthly temperatures (º F) for three cities. | | | Top of Form  **(a)**Which city has temperatures most evenly spread over the greatest range?   |  |  |  |  | | --- | --- | --- | --- | | Choose: | A | B | C |   Bottom of Form  Top of Form  **(b)**Which city has a potential outlier among its data?   |  |  |  |  | | --- | --- | --- | --- | | Choose: | A | B | C |   Bottom of Form | PracticeTemp | | Top of Form  **(c)**Which city has temperatures consistently above freezing (32º F)?   |  |  |  |  | | --- | --- | --- | --- | | Choose: | A | B | C |   Bottom of Form | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **5.**  Examine the histograms shown below. Complete the chart below by choosing the graph that best shows the stated result. | | orachistogreen2 | | |  |  |  |  | | --- | --- | --- | --- | |  | Graph A | Graph B | Both Graphs | | Larger Median | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Symmetric Graph | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Smaller Range | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Larger Data Count | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Interval Scale of 2 | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Unimodal | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **6.**  Two game players have been recording their best scores each day during a one week video tournament. Based on the box plots shown below, determine if the following statements are true or false. | | | **(a)**The median score of Player A is greater than the median score of Player B.  Top of Form   |  |  |  | | --- | --- | --- | | Choose: | True | False |   Bottom of Form  **(b)**The highest score was obtained by Player B.  Top of Form   |  |  |  | | --- | --- | --- | | Choose: | True | False |   Bottom of Form | practiceplayers | | **(c)**The interquartile range of the scores of Player B is less than the interquartile range of the scores of Player A.  Top of Form   |  |  |  | | --- | --- | --- | | Choose: | True | False |   Bottom of Form | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **7.**  Examine the dot plots shown below. Complete the chart below by comparing the two plots. | | datax     datay | | |  |  |  |  | | --- | --- | --- | --- | |  | Larger in Data X | Same in both data sets | Larger in Data Y | | Median | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Range | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Mean | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Standard Deviation | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Mode | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | Maximum | Top of Form   |  |  |  | | --- | --- | --- | |  |  |  |   Bottom of Form | | | | |

Bottom of Form

Bottom of Form

Bottom of Form

Bottom of Form

Bottom of Form