Part I: Vocabulary List

Please define each of the following terms from the information on the stattrek website. When asked to provide an example of the word, provide a unique example of the word NOT given on the website.

1. **Categorical Variables** Also known as qualitative variables. These take on names or labels.
   Example: Ice cream flavor, brand of shoes, color pencil
2. **Quantitative Variables** Numerical variables. These variables represent a measurable quantity.
   Example: Number of students who play sports, how many pets someone owns
3. **Univariate Data** A study that only focuses on one variable
4. **Bivariate Data** A study that focuses on the relationship between two variables.
5. **Median** Arrange observations from smallest to largest and select the middle value.
6. **Mean** Adding all observation values and divide by number of observations.
7. **Population** Includes each element from the set of observations that can be made. (Parameter)
   Example: All people living in the U.S.
8. **Sample** Include one or more observations from the population (statistics)
   Example: Some people living in the U.S.
9. **Center** Located at the median of the distribution.

Center turns out to be a kind of slippery thing in statistics. The median is often a useful measure of the "center" of data. Both there are other "measures of central tendency" such as median that might be useful in certain contexts. You can think of center as "the most typical point of data."

10. **Spread** Variability of the data

11. **Symmetry** Distribution can be divided in half so that each half is a mirror image of the other half.

12. **Unimodal and Bimodal** Unimodal: when there is one clear peak
Bimodal: Two clear peaks
Single peak at the center = Bell-shaped

13. **Skewness** When there's more observations on one side of the graph than another.

Sketch Skewed left:
Fewer observations on the left

Sketch Skewed right:
Fewer observations on the right

14. **Uniform** Observations are equally spread across the range of the distribution.

15. **Gaps** Areas of a distribution with no data/observations

16. **Outliers** Extreme values that differ greatly from other observations

17. **Dot plots** A type of graphic display used to compare frequency counts within categories or groups.

18. **Difference between a bar chart and histogram**
Bar charts have spaces between columns, while histograms usually do not.
Bar charts = qualatative column variable Histograms = quantitative column

19. **Stem plots**
A type of chart that shows how individual values are distributed within a set of data. (stems & leaves)

20. **Boxplots** Type of graph used to display patterns of quantitative data.

21. **Quartiles** Used to divide a rank-ordered data set into four equal parts. Q1, Q2, Q3

Typically a continuous, quantitative variable.

22. Range The difference between the largest and smallest values in a set of values.

23. Interquartile Range The measure of variability, based on dividing a data set into quartiles.

Q1 Q2 Q3
1 2 3 4 5 6 7 8

Q3 - Q1 = IQR

24. Parallel boxplots

When data from two groups is displayed on the same chart using the same measurement scale.

Maybe it goes without saying, but the data sets are displayed as boxplots.

Data are given as a simple number count

25. Difference between a frequency table and relative frequency table

Frequency tables shows the popularity or mode of a certain type of data

Relative frequency tables show the number of times a specific event occurs compared to the total number of events.

26. Parameter An element from a population

Often given as percentages or proportions

The key here is the POPULATION. Parameters are measurable characteristics of the WHOLE population

27. Statistic Observations taken from a sample.

The key here is the SAMPLE. Statistics are measurable characteristics of the SAMPLE.

28. Marginal Distribution The probability distribution of the sums of rows or columns expressed as percentages of the grand total.

In frequency tables.

29. Conditional Distribution Probability distribution for a sub-population

In relative frequency tables.

30. Segmented Bar Chart Stacked bar chart with two axes. One shows a discrete value and the other shows values w/ different bars in different categories.

31. What are the W's of data (might need to do your own digging for this!)

Who, What, When, Where, Why, and How

Part 2: Practice Problems

CATEGORICAL OR QUANTITATIVE
Determine if the variables listed below are *quantitative* or *categorical*.

1. Time it takes to get to school Q
2. Number of people under 18 living in a household Q
3. Hair color C
4. Temperature of a cup of coffee Q
5. Teacher salaries Q
6. Gender C
7. Smoking C
8. Height Q
9. Amount of oil spilled Q
10. Age of Oscar winners Q
11. Type of Depression medication C
12. Jellybean flavors C
13. Country of origin C
14. type of meat C
15. number of shoes owned Q

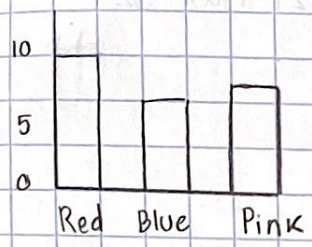
STATISTIC – WHAT IS THAT?
A statistic is a number calculated from data. Quantitative data has many different statistics that can be calculated. Determine the given statistics from the data below on the number of homeruns Mark McGuire has hit in each season from 1982 – 2001.

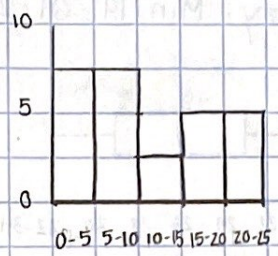
70	52	22	49	3	32	58	39
39	65	42	29	9	32	9	33

Mean	36.4375
Minimum	3
Maximum	70
Median	36
Q1	25.5
Q3	50.5
Range	67
IQR	25

3 9 9 22 29 32 32 33 39 39 42 49 52 58 65 70
↓ 25.5 ↓ 36 ↓ 50.5

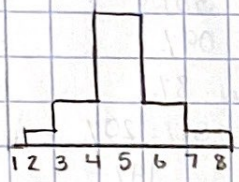
16. Bar charts - qualitative column variable Histogram - quantitative column variable





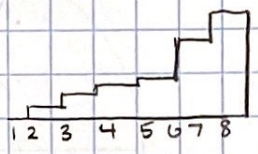
Sketch

Unimodal

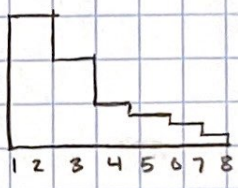


Bimodal

Skewed Left



Skewed Right



Define & Formulas μ = pop mean X_i = ith element N = number of elements

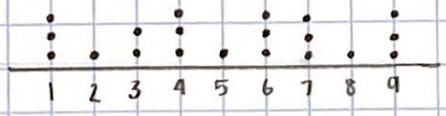
Standard Deviation
↳ Square root of the variance
$$\sigma = \sqrt{\sigma^2} = \sqrt{\Sigma(X_i - \mu)^2 / N}$$

A measure of variation in a data set.

Variance
↳ Average squared deviation from the population mean
$$\sigma^2 = \Sigma(X_i - \mu)^2 / N$$

Standard Score (Z-score)
↳ Indicates how many standard deviations an element is from the mean.
$$Z = (X - \mu) / \sigma$$

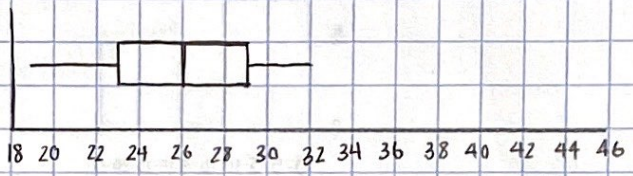
17.



Mean : 5.1
Median: 5.5
Standard Deviation : 2.71

Standard Deviation is a frustrating calculation to do by hand. We will usually use technology to aid us.

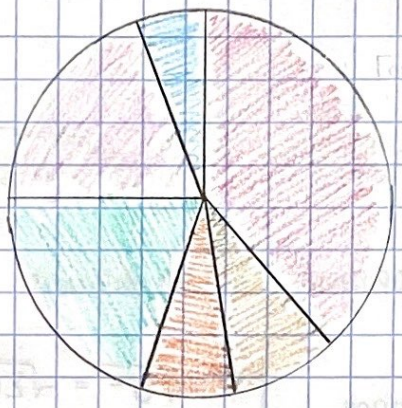
18. Mean : 26
Standard Deviation : 3.83
5-Number Summary: Min: 19 Q1 = 23 Median = 26 Q3 = 29 Max = 32
IQR: 6 = 29 - 23



19.

Titles are important!

Diego's Video Game Usage

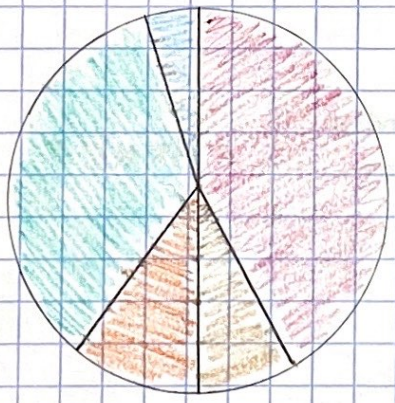


Fortnite: 38%
Pubg : 09%
The Show: 8%
Madden 2017: 20%
God of War: 19%
COD: 6%

20.

Titles are important!

Food Items Sold



Cheeseburger: 40%
Milkshake: 10%
Chicken Fingers: 13%
Fries: 33%
Egg Salad: 5%

SHOPPING SPREE!

A marketing consultant observed 50 consecutive shoppers at a supermarket. One variable of interest was how much each shopper spent in the store. Here are the data (round to the nearest dollar), arranged in increasing order:

3	9	9	11	13	14	15	16	17	17
18	18	19	20	20	20	21	22	23	24
25	25	26	26	28	28	28	28	32	35
36	39	39	41	43	44	45	45	47	49
50	53	55	59	61	70	83	86	86	93

a. Make a stemplot using tens of dollars as the stem and dollars as the leaves. Make sure you include appropriate labels, title and key.

Shopping Spree Money ($) Spent in Store

Stem	Leaves
9	3
8	3 6 6
7	0
6	1
5	0 3 5 9
4	1 3 4 5 5 7 9
3	2 5 6 9 9
2	0 0 0 1 2 3 4 5 5 6 6 8 8 8 8
1	1 3 4 5 6 7 7 8 8 9
0	3 9 9

Key: 8|0 = 80

SSHA SCORES

Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

154 109 137 115 152 140 154 178 101 103 126 126 137 165 165 129 200 148

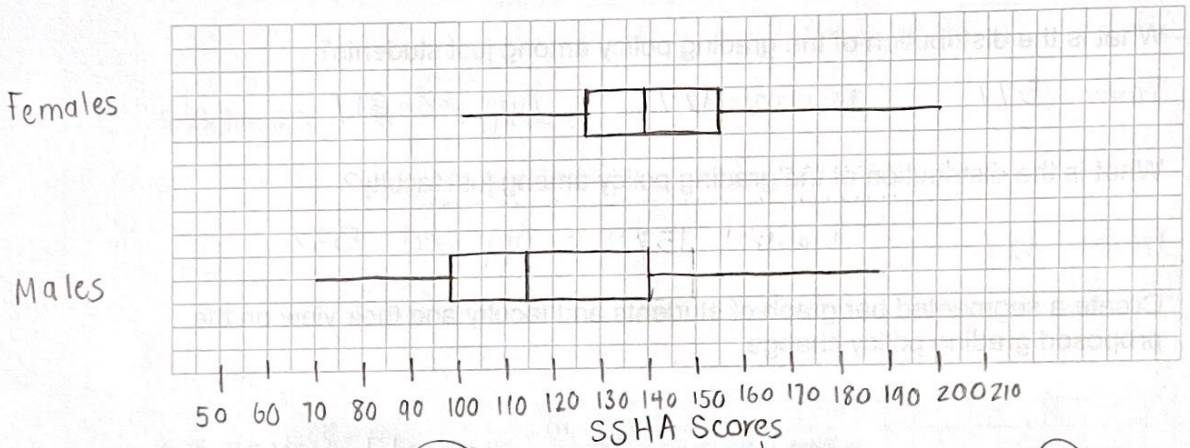
and for 20 first-year college men:

108 140 114 91 180 115 126 92 169 146 109 132 75 88 113 151 70 115 187 104

a. Put the data values in order for each gender. Compute numeral summaries for each gender.

	Women			Men
Mean	141.055		Mean	121.25
Minimum	101		Minimum	70
Q1	126		Q1	98
Median	138.5		Median	114.5
Q3	154		Q3	143
Maximum	200		Maximum	187
Range	99		Range	117
IQR	28		IQR	45

b. Using the minimum, Q1, Median, Q3, and Maximum from each gender, make parallel boxplots to compare the distributions.



101 103 109 115 126 126 129 137 137 | 140 148 152 154 154 165 165 178 200

138.5

70 75 88 91 92 | 104 108 109 113 114 | 115 115 132 140 140 151 164 180 187

98 114.5 126 146 143

New Grading Policy

A new grading policy has been proposed by the dean of the College of Education for all education majors. All faculty and students in the college were asked to give their opinions about the new policy. The results are given below.

	Favor	Neutral	Opposed	Row Total
Students	353	75	191	619
Faculty	11	5	18	34
Column Total	364	80	209	653

a. State the variables and if they are categorical or quantitative.

Votes - quantitative
People voting - categorical
Opinions - categorical

I know it can be confusing, but the MAIN variable of interest here is categorical: people's opinions.

b. What percent of responses were from students favoring the policy change? 353/653=54.05%
What percent of students favored the policy change? 353/619=57.02%
What percent favoring the policy change were students? 96.98%.

c. What is the marginal distribution of the grading policy change?

Favor = 56%. Neutral = 12%. Opposed = 32%.

d. What is the distribution of the grading policy among just students?

Favor = 57%. Neutral = 12%. Opposed = 31%.

e. What is the distribution of the grading policy among just faculty?

Favor = 32%. Neutral = 15%. Opposed = 53%.

f. Create a segmented bar graph of students and faculty and their view on the proposed grading policy change.

See other student sample

Algebra Section:
The prerequisite for AP Statistics is Algebra II. You will find very much equation solving in this course, but some quick review of Algebra I and Algebra II content will be helpful.

Here is a formula that is used often in AP Statistics: $z = \frac{x - \bar{x}}{s}$

1. If z = 2.5, x = 102, and $\bar{x}$ =100. What is s? Show your work.

$2.5 = \frac{102 - 100}{s}$ $\quad 2.5 = \frac{2}{s}$ $\quad \boxed{s = 0.8}$

2. If z = -3.35, x = 60, and s = 4, what is $\bar{x}$? Show your work.

$-3.35 = \frac{60 - \bar{x}}{4}$ $\quad -13.4 = 60 - x$ $\quad -60 \quad +60$ $\quad -73.4 \quad 0 \quad -1$ $\quad -1$ $\quad \boxed{\bar{X} = 73.4}$

3. Solve $\frac{0.05}{1.96} = \frac{1.96\sqrt{\frac{0.5^2}{n}}}{1.96}$ for n.

$0.0255^2 = \sqrt{\frac{0.3^2}{n}}^2$

$6.51 \times 10^{-4} = \frac{0.5^2}{n}$ $\quad \boxed{n = 384.16}$

4. If $-1.64 = \frac{60 - \mu}{\sigma}$ and $1.96 = \frac{95 - \mu}{\sigma}$, solve for μ and σ.

$60 - \mu$

See other student sample