

Part 1: Vocabulary Hunt / Review

- Use the Online Resource: <https://www.statisticshowto.com/probability-and-statistics> o On the next page, you will see a list of vocabulary terms. Use the above website to write definitions and then provide an example or diagram where asked.
- Bookmark this site for later use: it's a pretty solid study resource.

Part 2: Practice Problems

- Any charts or graphs can be done by hand or created using software (e.g., Excel / Google Sheets).

Parts 1 and 2 ARE DUE THE FIRST FRIDAY IN SEPTEMBER. IT WILL BE CHECKED FOR

COMPLETENESS. Complete means you did your best and you made an honest attempt to set yourself up for success this year. I do not expect perfection! “Essentially complete and correct” is the standard. In fact, it is more than acceptable to note a few words that you couldn’t find or a definition you didn’t understand. Make note of it and then move on. We will be going over all this during the school year, but I don’t want us getting tripped up by basics. So, THIS ASSIGNMENT ESTABLISHES MUCH OF THE VOCABULARY I AND YOUR TEXTBOOK WILL BE USING TO DESCRIBE MORE ADVANCED STATISTICS.

Part 3: Procure a graphing calculator

- This course and the AP exam requires a graphing calculator. I recommend the TI-84Plus model and will be teaching the course assuming you have one. There will be a class set available in the room, but it is beneficial (though not required) to have your own. If buying a graphing calculator is a financial problem, please let me know.

Part 4: Get interested in statistics:

- Watch the following presentation about the power of a well-constructed argument that uses statistics: https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen ● Explore the website <http://gapminder.org>. Click on the option to “Play with our data.” Can you use it to learn something unexpected/interesting about the world? How can you draw conclusions from this presentation of data?

I will be intermittently available over the summer via email at pmawn@woburnps.com for questions.

DO NOT LEAVE THIS ASSIGNMENT UNTIL THE WEEK BEFORE SCHOOL STARTS.

Part 1: Vocabulary

Directions: Define the following terms from the information on the statisticshowto website. For some terms, please provide an example (different from any on the website). For other terms, I may ask for a formula or a diagram.

1. Categorical Variables:

a. Example:

2. Quantitative Variables:

a. Example:

3. Discrete Variables:

a. Example:

4. Continuous:

5. Univariate Data:

6. Bivariate Data:

7. Population:

a. Example:

8. Sample:

a. Example:

9. Median:

10. Mean:

a. Formula:

11. Parameter:

12. Statistics:

13. Range:

14. Center:

15. Spread:

16. Variance:

a. Formula:

17. Standard Deviation:

a. Formula:

18. Standard Score (z-score)

a. Formula:

19. Symmetry:

a. Diagram:

20. Unimodal:

a. Diagram:

21. Bimodal

a. Diagram:

22. Skewness:

a. Diagram of data skewed left: b.

Diagram of data skewed right: 23.

Uniform:

a. Diagram:

24. Gaps:

a. Diagram:

25. Outliers

a. Diagram:

26. Dot plots:

27. Bar Chart:

28. Histogram:

29. Difference between a bar chart and a histogram:

30. Stemplots:

31. Box plots:

32. Quartiles:

33. Range:

34. Interquartile Range (IQR):

35. Difference between a statistic and a parameter: 36.

Difference between an assumption and a condition:

Part 2: Practice Problems

CATEGORICAL OR QUANTITATIVE VARIABLES

Directions: Determine if the variables listed below are quantitative or categorical.

1. Time it takes to get to school
2. Number of people under 18 living in a household

3. Hair color
4. Temperature of a cup of coffee
5. Teacher salaries
6. Gender
7. Smoking vs. Non-Smoking
8. Height
9. Amount of oil spilled
10. Age of Oscar winners
11. Type of Depression medication
12. Jellybean flavors
13. Country of origin
14. Type of meat
15. Number of shoes owned

STATISTIC –WHAT IS THAT?

1. A statistic is a numerical summary of data. Quantitative data has many different statistics that can be calculated. Determine the given statistics from the data below on the number of home runs Mark McGuire hit in each season from 1982 – 2001 (with one missing season).

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

Mean

Minimum

Maximum

Median

Q1

Q3

Range

Interquartile Range (IQR)

Based on the statistics above, give a plausible range for what you expect the value of the ??? cell to be. Explain your reasoning.

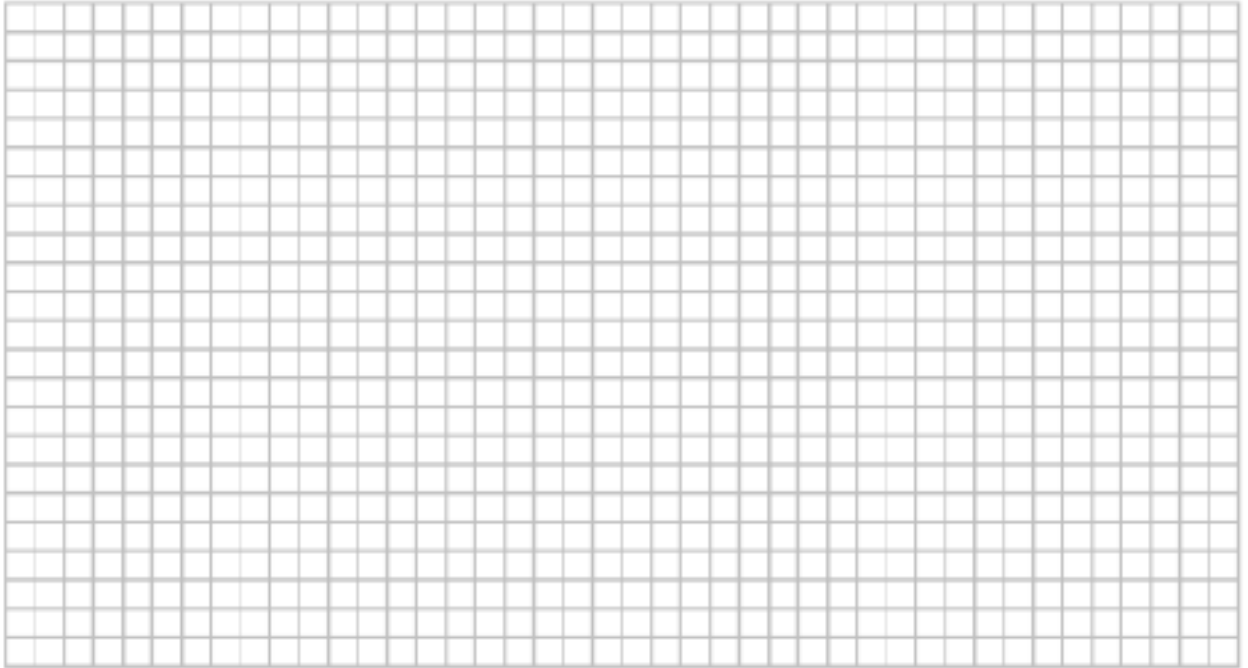
DISPLAYING DATA: GRAPHS AND CHARTS

2. *ACCIDENTAL DEATHS*: In 1997 there were 92,353 deaths from accidents in the United States. Among these were 42,340 deaths from motor vehicle accidents, 11,858 from falls, 10,163 from poisoning, 4051 from drowning, and 3601 from fires. The rest were listed as “other” causes.

- a. Find the percent of accidental deaths from each of these causes, rounded to the nearest percent.

b. What percent of accidental deaths were from “other” causes?

c. NEATLY create a well-labeled bar graph of the distribution of causes of accidental deaths. Be sure to include an “other causes” bar.



d. A pie chart is another graphical display used to show all the categories in a categorical variable relative to each other. Create a pie chart for the accidental death percentages. You may try using software (*Excel, Google sheets*) to make one and paste in the space below.

3. *HURRICANE SEASON*: The data below gives the number of hurricanes that happened each year from 1944 through 2000 as reported by Science magazine.

3	2	1	4	3	7	2	3	3	2	5	2	2	4	2	2	6	0	2	5	1	3	1	0
3	2	1	0	1	2	3	2	1	2	2	2	3	1	1	1	3	0	1	3	2	1	2	1
1	0	5	6	1	3	5	3																

a. Make a dot plot to display these data. Make sure you include appropriate labels, title, and scale. The graph paper should ensure you space your markings consistently.



What does this dot plot show you about the data that was not evident just by looking at the series of numbers give?

Suppose you wanted to evaluate the claim that global warming has caused an increase in the frequency of hurricanes. How could you evaluate this claim? Assuming that the numbers in the above chart are in chronological order, do you agree or disagree with this claim? Explain.

4. *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 stem plot using tens of dollars as the stem and dollars as the leaves. Make sure you include appropriate labels, title and key.

- b. What does the stem plot show you about the data that was not evident just by looking at the raw data?

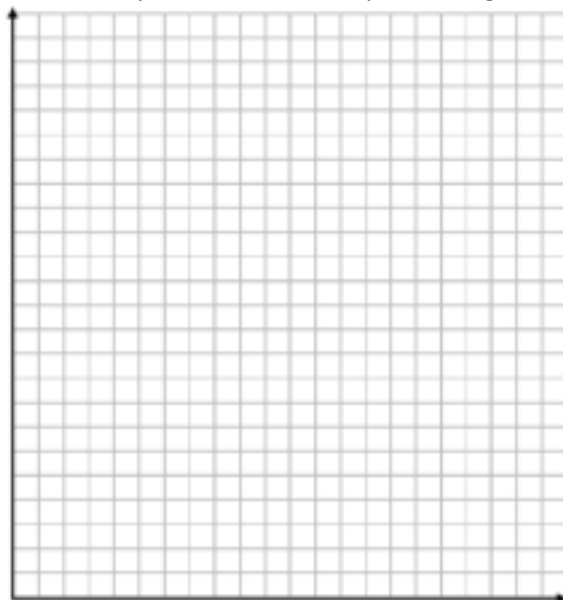
5. *WHERE DO OLDER FOLKS LIVE?* This table gives the percentage of residents aged 65 or older in each of the 50 states.

State	Percent	State	Percent	State	Percent
Alabama	13.1	Louisiana	11.5	Ohio	13.4
Alaska	5.5	Maine	14.1	Oklahoma	13.4
Arizona	13.2	Maryland	11.5	Oregon	13.2
Arkansas	14.3	Massachusetts	14.0	Pennsylvania	15.9
California	11.1	Michigan	12.5	Rhode Island	15.6
Colorado	10.1	Minnesota	12.3	South Carolina	12.2
Connecticut	14.3	Mississippi	12.2	South Dakota	14.3
Delaware	13.0	Missouri	13.7	Tennessee	12.5
Florida	18.3	Montana	13.3	Texas	10.1
Georgia	9.9	Nebraska	13.8	Utah	8.8
Hawaii	13.3	Nevada	11.5	Vermont	12.3
Idaho	11.3	New Hampshire	12.0	Virginia	11.3
Illinois	12.4	New Jersey	13.6	Washington	11.5
Indiana	12.5	New Mexico	11.4	West Virginia	15.2
Iowa	15.1	New York	13.3	Wisconsin	13.2
Kansas	13.5	North Carolina	12.5	Wyoming	11.5
Kentucky	12.5	North Dakota	14.4		

a. Complete the table of bins for the percent (p) of residents aged 65 or older started below on the left and then create a histogram. Make sure you include appropriate labels, titles and scale.

What can you conclude from your histogram?

Bin Widths	Frequency
$4 < p < 6$	1
$6 < p < 8$	
$8 < p < 10$	



6. SSHA Scores: The Survey of Study Habits and Attitudes gives surveys to incoming college freshmen. After asking them a series of questions, the survey evaluates each person regarding the number of minutes spent on homework every week.

Below is a sample of their data for 18 year old freshmen females:

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

Below is a sample of their data for 18 year old freshmen males:

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

- a. Put the data in order for each gender and complete numeral summaries for each gender.

Women	Men
-------	-----

Mean	Mean
Minimum	Minimum
Q1	Q1
Median	Median
Q3	Q3
Maximum	Maximum
Range	Range
IQR	IQR

b. Using the Min, Q1, Med, Q3, Max from each gender, make parallel box plots in order to compare the two genders.



What can you conclude about the study habits of each gender?

Surveying and Obtaining Data:

Directions: To answer the following questions, refer to the “Sampling in Statistics” page. There you will find descriptions of many types of and considerations needed for sampling. The ones listed below are those that we will focus on this year. Try to use this page to find information on the following types of surveying and write a definition or description for each.

a) Voluntary Response

b) Convenience

c) Simple Random

d) Stratified

e) Cluster

f) Multistage

g) Systematic

7. The Massachusetts Division of Weight Watchers is doing research to determine how many people cheat on their diets (go over on daily points) at least once a week. They decide that anonymous surveys will give them the best chance for honest answers, but they do not have time to survey ALL people enrolled in Weight Watchers. Read the scenarios below and determine which of the 7 survey methods above best describe their survey:

I. Randomly select 10 members from each of the Weight Watchers centers in Massachusetts. _____

II. Use an alphabetical listing of all Massachusetts members. Randomly choose a starting person on the list. Then select every 20th person thereafter to survey.

III. Randomly select 2 or 3 branches of the Massachusetts division and survey every member of that center. _____

IV. Send out the survey to every member of the Massachusetts division. Place drop boxes in each Weight Watchers center. Anyone who returns a survey will be in the sample.

V. The Massachusetts regional office is in Waltham, so they survey members at the Weight Watchers center in Waltham only. _____

VI. From a numbered list of all Massachusetts division members, use a computer to randomly select 100 numbers and survey all members with those corresponding numbers. _____

What is the population of interest in the above situation?

Algebra Review:

While we won't be encountering a ton of "solving for x" type problems this year, there will be instances where we will need to set up an equation and solve for a variable from a word problem. A formula

frequently used in statistics is $z = \frac{x - \bar{x}}{s}$. (note: $\bar{x}$ is a common notation for the average of all x values) 8. If $z = 2.5$, $x = 102$ and $\bar{x} = 100$, what is s ? Show your work.

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

10. If , and what is z ? Show your work.

It is expected that you have a thorough understanding of linear functions and scatter plots. 11.

The USDA reported that in 1990 each person in the United States consumed an average of 133 pounds of natural sweeteners. They also claim this amount has decreased by about 0.6 pounds, on average, each year.

a. If 1990 could be considered “year 0”, which of the above numbers represents the slope and which represents the y-intercept?

b. What is the equation of the line of best fit using the slope and y-intercept above?

c. Predict the average consumption of sweeteners per person for the year 2020. Does this prediction make sense? What if you tried to make a prediction for the year 2222?

12. The following equation can be used to predict the average height of boys anywhere

between birth and 15 years old: , where x is the age (in years) and is the predicted height (in inches).

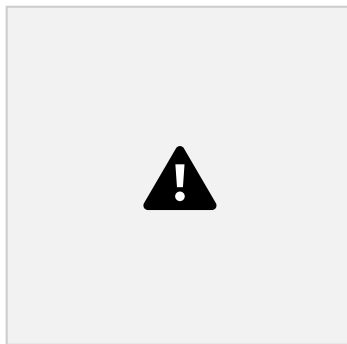
a. What does the slope represent in this problem? Interpret it in the context of this problem/situation.

b. What does the y-intercept represent in this problem? Interpret it in context.

13. Hilary wonders if people of similar heights tend to date each other. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (heights are in inches):



a. Construct a scatter plot of the data.



- b. Describe the association between the heights of the women and the men they date.
- c. Estimate a best fit line for the above data. What variables did you chose to use?
14. Write the equation of a line in slope –intercept form that has the following characteristics: a.
Every time the independent variable increases by 10, the dependent variable increases by 15
and the dependent variable has an initial value of 74.
- b. It passes through the points (4, 0.21) and (9, 1.73)

You are expected to have a basic understanding of simple probability. If you find these problems “less than intuitive”, review the “Intro to Probability” section on stattrek.

15. A special lottery is to be held to select the student who will live in the only deluxe room in a dormitory. There are 100 seniors, 150 juniors, and 200 sophomores who applied. Each senior's name is placed in the lottery 3 times; each junior's name, 2 times; and each sophomore's name, 1 time. What is the probability that a senior's name will be chosen? Show your work.
16. Which of the following has a probability closest to 0.5?
- a. The sun will rise tomorrow.
 - b. It will rain tomorrow.
 - c. You will see a dog with only three legs when you leave the room.
 - d. A fair die will come up with a score of 6 four times in a row.
 - e. There will be a plane crash somewhere in the world within the next five minutes.

Explain your answer:

17. If a coin is tossed twice, what is the probability that on the first toss the coin lands heads and on the second toss the coin lands tails? (Hint: What are the possible outcomes when you toss a coin twice?)
- a. $\frac{1}{8}$
 - b. $\frac{1}{6}$
 - c. $\frac{1}{4}$
 - d. $\frac{1}{2}$
 - e. 1

18. Calculate the following probabilities and arrange them in order from least to greatest.

1) The probability that a fair die will produce an even number. _____

2) A random digit from 1 to 9 (inclusive) is chosen, with all digits being equally likely. The

probability that when it's squared the answer will contain the digit 1. _____ 3) The

probability that a letter chosen from the alphabet will be a vowel. _____ 4) A

random number between 1 and 20 (inclusive) is chosen. The probability that its square root

will not be an integer. _____

ORDER: _____, _____, _____, _____

19. (Optional): What do you want me (Mr. Mawn) to know about you? About how you learn?
About your hopes/interests/goals? Why are you taking AP Stats? Favorite high school
teacher/experience so far? Your favorite joke/ice cream flavor/ hobby...etc...

20. (Optional): Do you have any questions for me (Mr. Mawn)? Anything you'd like to know
about me?