AP Statistics Summer Assignment 2019

Welcome to AP Statistics future statisticians! The purpose of this assignment is to make you comfortable exploring data analysis.

The summer assignment is composed of three parts.

1. Reading and Vocabulary: You will use a free online Statistical tutoring site that will give you information on variables and data displays. While reviewing the information on the site you will be completing a vocabulary list (see page 2 and 3). Follow the steps below:

- Go to www.stattrek.com
- Click on "AP Statistics" then "AP Tutorial"
- On the left side of the screen is a list of general topics. Under each general topic are a list of subtopics. You will read the following subtopics to complete the vocabulary list.

General Topic: Exploring Data	
Subtopics:	Variables
	Population vs Sample
	Central Tendency
	Variability
	Position
General Topic: Charts and Graphs	
Subtopics:	Charts and Graphs
	Patterns in data
	Dotplots
	Histograms
	Stemplots
	Boxplots
	Scatterplots
	Comparing data sets

2. Practice Problems: After reading all the material above you should be able to complete the questions in the remaining pages of this packet. You should do so in the spaces provided.

3. A graphing calculator is a required tool for this class. The TI-84+ (preferred), or TI-83+ is recommended. As you complete the Practice Problems refer to the TI Guidebook to become familiar with the list and statistical functions. For an online calculator go to http://www.alcula.com/calculators/statistics/. Other tutorials are available on youtube.

Upon your return to school in September, you are expected to turn in this packet completed. You are expected to complete each part of the problems and to construct all data displays neatly. This assignment will be graded for correctness. If you have any questions, please feel free to e-mail me at ddorval@nrpsk12.org

Have a great summer!!

Part 1: Vocabulary List

Please read the definitions of each of the following terms from the information on the textbook or the stattrek website. We will review these as we learn them and use them often enough that you will understand them.

1. Categorical Variables

2. Quantitative Variables

3. Discrete Variables

4. Continuous

5. Univariate Data

6. Bivariate Data

7. Population

8. Sample

9. Median

10. Mean

11. Outlier

12. Parameter

13. Statistics

14. Range

15. Standard Score (z-score)

16. Center

17.Spread

18. Variance

19. Standard Deviation

20.Symmetry

21. Unimodal

22. Bimodal

23.Skewness

24.Uniform

25.Gaps

26.Outliers

27.Dot plots

28. Bar Chart
29. Histogram

30.Difference between bar chart and histogram

31.Stemplots

32. Boxplots

33.Quartiles

34.Range

35.Interquartile Range

36. Four Ways to Describe Data Sets

37. Types of Graphs that can used for comparing data

Part 2: Learn About R using Datacamp

R is a free but very powerful programming language used in statistical analysis. Its flexibility and ease of use make it accessible to beginning learners but also very powerful in its ability to develop and implement new procedures.

I will use R frequently throughout the year to demonstrate statistical properties; you will use R throughout the year on projects.

R and R Studio are available free and may be downloaded from the sites below. These downloads are not necessary for the summer assignment; all work will be completed using DataCamp. R programs may be used on either interface; choosing 1 over the other is personal preference. I learned on R so I typically use that; R Studio looks more like the DataCamp activities (it also has some additional features that we won't be using).

Download R (optional)

Download R Studio (optional)

Please complete the following beginning after August 1st

- I will enter emails you provide to form a class AP Statistics 2019-2020, after that you should get an email.
- Go to Datacamp.com and create an account using your school email.
- Go to the "Courses" tab along the top and find "Introduction to R"
- Under "Introduction to R" complete the following chapters:
    - Intro to Basics
    - Vectors
    - Matrices
    - Factors
    - Data Frames**
- Go to the "Courses" tab along the top and find "Data Visualization in R"
- Under "Data Visualization in R" complete the following chapters:
    - A Quick Introduction to Base R Graphics
    - Different Plot Types**
    - Adding Details to Plots**

**These chapters are only available through the group.

Part 3: Practice Problems

CATEGORICAL OR QUANTITATIVE

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
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?

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 hit in each season from 1986 - 2001.

3	49	32	33	39	22	42	9	9
39	52	58	34	24	70	65	32	29

Mean	
Minimum	
1st Quartile	
Median	
3rd Quartile	
Maximum	
Range	
IQR	

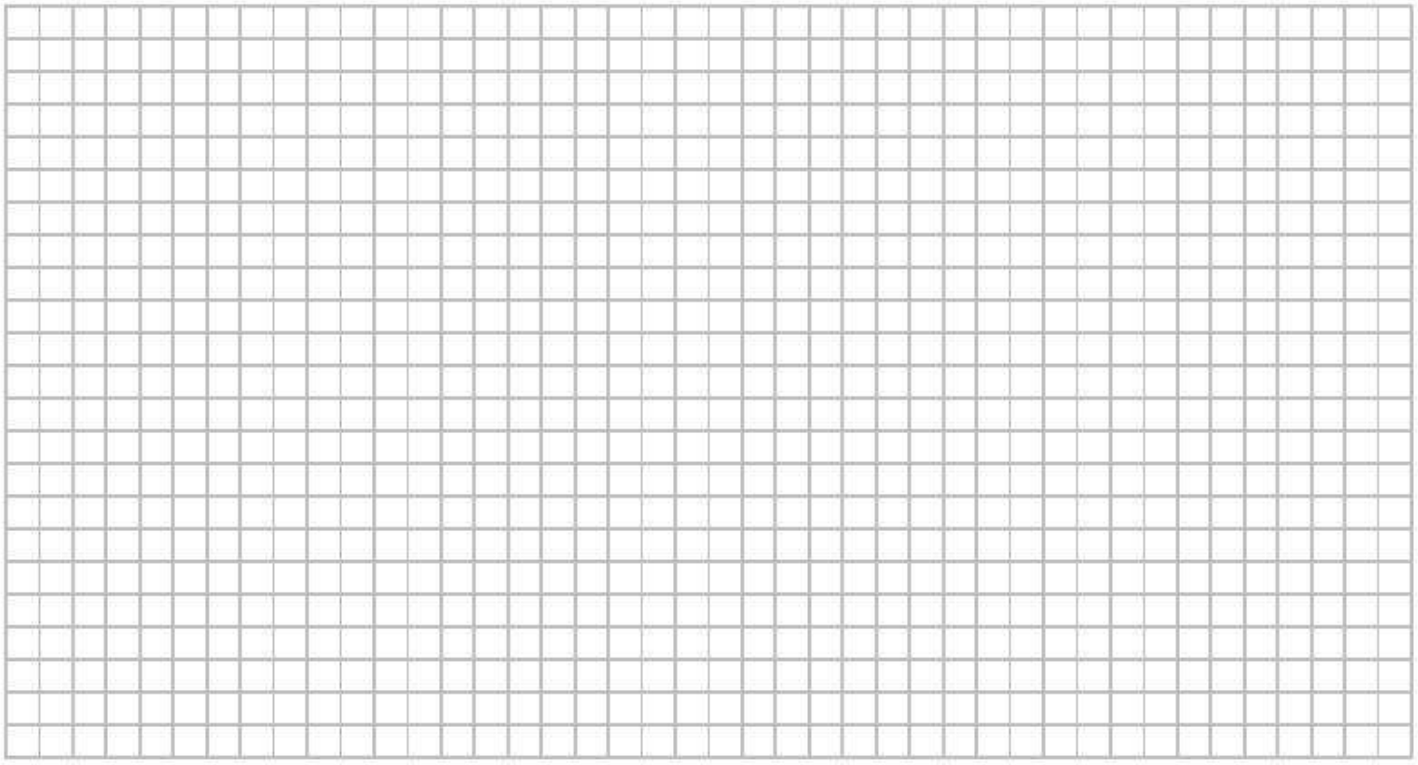
ACCIDENTAL DEATHS

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 and titled bar graph of the distribution of causes of accidental deaths. Be sure to include an "other causes" bar.



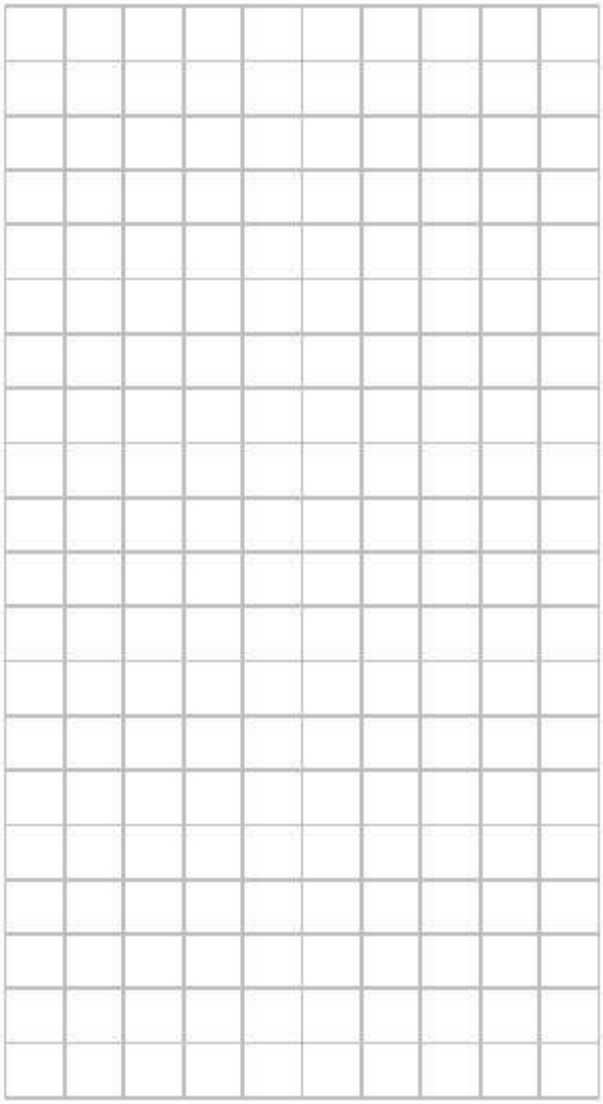
d. A pie chart is another graphical display used to show all the categories in a categorical variable relative to each other. Neatly create a pie chart for the accidental death percentages by hand. Try to replicate your pie chart using software or an internet source to make one and attach it.

IT'S A TWISTA

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 dotplot to display these data. Make sure you include appropriate labels, title, and scale. The graph paper should help ensure you space your markings (you may use x's or dots) consistently.



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.

WHERE DO OLDER FOLKS LIVE?

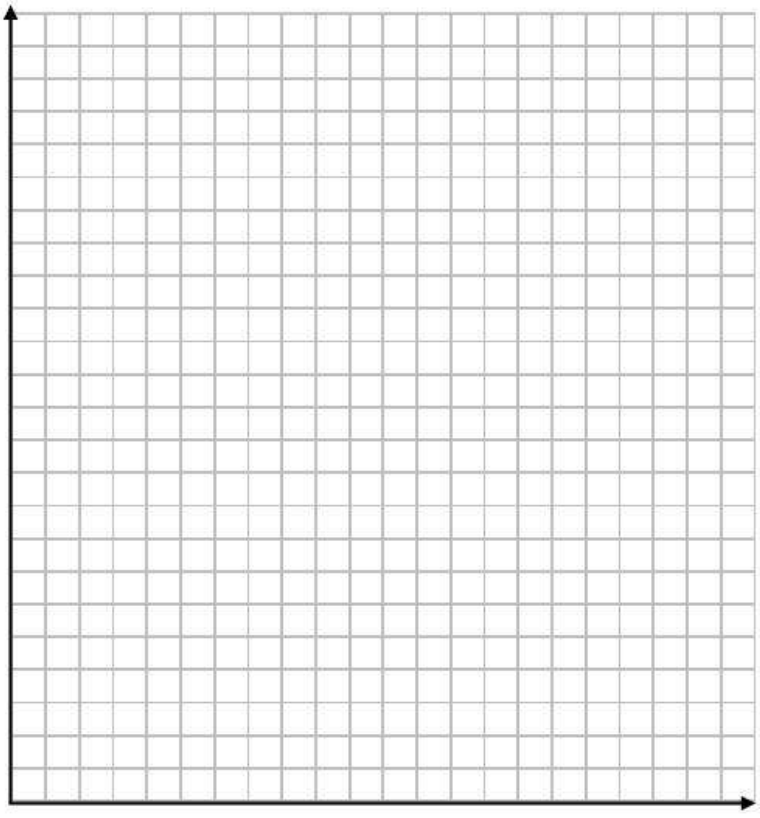
This table gives the percentage of residents aged 65 or older in each of the 50 states.

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		

Histograms are a way to display groups of quantitative data into bins (the bars). These bins have the same width and scale and are touching because the number line is continuous. To make a histogram you must first decide on an appropriate bin width and count how many observations are in each bin. The bins for percentage of residents aged 65 or older have been started below for you.

a. Finish the chart of Bin widths and then create a histogram using those bins on the grid below. Make sure you include appropriate labels, title and scale. Create the same histogram in both using software and using your calculator)

Bin Widths	Frequency
4 to < 6	1
6 to < 8	
8 to < 10	



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

b. Using the minimum, Q1, Median, Q3, and Maximum from each gender, make parallel boxplots to compare the distributions. Create the same graph using your calculator and computer software.

To answer the following refer to the readings on www.stattrek.com "Survey Sampling Methods".

The 7 types of sampling designs are:

A. voluntary response B. convenience C. simple random D. stratified

E. cluster F. multistage G. systematic

1. The Maryland division of Weight Watchers is doing research to determine how many people on the Weight Watchers diet cheat at least once a week. They decide that anonymous surveys will give them an accurate representation but do not have time to get responses from ALL the Maryland Weight Watchers people. Read the scenarios below and determine which of the 7 sampling methods best describes it.

_____ I. Randomly select 10 members from each of the WW centers in the Maryland division.

_____ II. Use an alphabetical listing of all Maryland division members. Randomly choose a starting person on the list. Then select every 20th person thereafter.

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

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

_____ V. The Maryland regional office is in Baltimore so they survey members at the WW center in Baltimore.

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

2. What is the population of interest in the WW situation?

Here is a formula often used in 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. If z = -3.35, $\bar{x}$ = 60, and s = 4, what is x ? Show your work.

It is expected that you have a thorough understanding of linear functions and scatterplots.

1. 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 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 2005.

2. The following equation can be used to predict the average height of boys anywhere between birth and 15 years old: $\hat{y}$= 2.79x + 25.64, where x is the age (in years) and $\hat{y}$ is the predicted height (in inches).

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

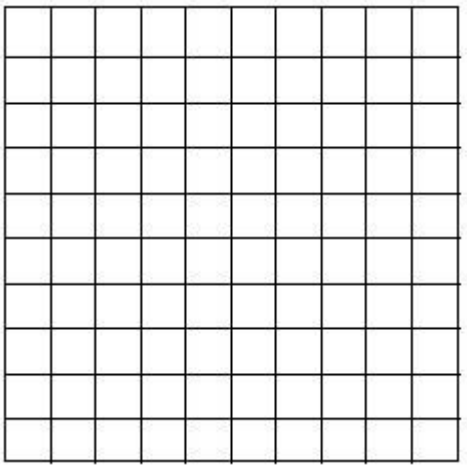
b. What does the y-intercept represent in this problem? Interpret it in context.

3. 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 in inches):

Women:	66	64	66	65	70	65
Men:	72	68	70	68	74	69

A. Construct a scatterplot of the data. Create the same scatterplot on both computer software and your calculator.

B. Describe the association between the heights of the women and the men they date.



You are expected to have a basic understanding of simple probability. If you find these problems "less than intuitive", there are numerous sites available online that provide basic probability explanations.

1. 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.

A. 1/8 B. 2/9 C. 2/7 D. 3/8 E. 1/2

2. 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.

3. 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. 1/6
B. 1/3
C. ¼
D. ½
E. 1

4. If a coin is tossed twice what is the probability that it will land either heads both times or tails both times?

A. 1/8
B. 1/6
C. 1/4
D. 1/2
E. 1

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

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

II. 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. _____

III. The probability that a letter chosen from the alphabet will be a vowel. _____

IV. A random number between 1 and 20 (inclusive) is chosen. The probability that its square root will not be an integer. _____

ORDER: _____, _____, _____, _____