Name: _____

Welcome to AP Statistics!
(2022-2023)

You have selected a college-level math course unlike your others. The material is extremely relevant to almost all aspects of life. I expect you to put forth a strong, consistent effort.

The purpose of this *Summer Assignment* is to:

1. Give you information on what to expect, and how this course is different from other math courses.
2. Refresh your knowledge of statistics topics that you should know prior to this course.
3. Give you a chance to demonstrate your ability to analyze data.

Please list some of the sites on the Internet that you found helpful when defining terms, explaining concepts, using the graphing calculator, etc.

⇨ _____

⇨ _____

⇨ _____

❖ **THIS ASSIGNMENT IS DUE AT OUR FIRST CLASS MEETING AT THE END OF AUGUST.**
❖ **THIS WILL COUNT AS YOUR FIRST ASSESSMENT GRADE**

If you have any questions or concerns while completing this assignment, email Mrs. Ames (aames@hamden.org).

1) What is the definition of a **statistic?**

2) Explain the difference between a **categorical variable** and a **quantitative variable.**

3) Determine if the variables listed below are *quantitative* or *categorical*. Neatly print "Q" for quantitative and "C" for categorical.

    _____ 1.Time it takes to get to school
    _____ 2. Number of shoes owned
    _____ 3. Hair color
    _____ 4. Temperature of a cup of coffee
    _____ 5. Teacher salaries
    _____ 6. Gender
    _____ 7. Height
    _____ 8. Amount of oil spilled
    _____ 9. Age of Oscar winners
    _____ 10. Type of pain medication
    _____ 11. Jelly Bean flavors
    _____ 12. Country of origin

4) Identify one type of graph you would make when working with a categorical variable? _____

   A quantitative variable? _____

5) What is the meaning of **distribution** in statistics?

6) When describing a distribution, what 4 things should you always mention?
   -
   -
   -
   -

7) Explain the difference between **range** and **interquartile range**.

8) Create a data set (with 5 numbers) in which the mean is *equal to* the median.

9) Create a data set (with 5 numbers) in which the mean is *greater than* the median.

10) **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 (rounded 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

Make a **stemplot** using tens of dollars as the stem and dollars as the leaves.

11) 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 numerical summaries for each gender. *This can be done using a graphing calculator or by hand.*

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. Make sure to label the horizontal axis.

12)   How is a **parameter** different from a **statistic**?

13)   List the steps required to **standardize** a value.

        What is the purpose of standardizing?

14)   **SAT versus ACT** – Eleanor scores 680 on the mathematics part of the SAT. The distribution of SAT scores in a reference population is symmetric and single-peaked with mean 500 and standard deviation 100. Gerald takes the American College Testing (ACT) mathematics test and scores 27. ACT scores also follow a symmetric, single-peaked distribution but with mean 18 and standard deviation 6. Find the standardized scores for both students. Assuming that both tests measure the same kind of ability, who has the higher score? _____ *Show your calculations.*

15) Briefly explain the **68-95-99.7% Rule**.

16) **Men's heights** – The distribution of heights of adult American men is approximately Normal with mean 69 inches and standard deviation 2.5 inches.

    a) Draw a Normal curve. Be sure to label the mean, as well as the points one, two, and three standard deviations away from the mean on the horizontal axis.

[#16 cont.]

     b) Use the 68-95-99.7% Rule to answer the following questions

- What percent of men are taller than 74 inches? _____

- Between what heights do the middle 95% of men fall? _____

- What percent of men are shorter than 66.5 inches? _____

17) What type of graph is used to show the relationship between 2 quantitative variables? _____

18) What does the sign (+/-) of the **correlation coefficient** tell you about the association between 2 quantitative variables? [Don't just say "positive" vs. "negative."]

19) What does an '**r**' value near 1 or -1 indicate?

20) List the steps required to perform a **linear regression** on data in L_1 and L_2 on the graphing calculator. (*A graphing calculator is strongly recommended for this class.)

21) **Meet the Archaeopteryz** – *Archaeopteryz* is an extinct beast having feathers like a bird but teeth and a long bony tail like a reptile. Only six fossil specimens are known. Because these specimens differ greatly in size, some scientists think they are different species rather than individuals from the same species. We will examine some data. If the specimens belong to the same species and differ in size because some are younger than others, there should be a positive linear relationship between the lengths of a pair of bones from all individuals. An outlier from this relationship would suggest a different species. Here are data on the lengths in centimeters of the femur (a leg bone) and the humerus (a bone in the upper arm) for the five specimens that preserve both bones.

Femur	38	56	59	64	74
Humerus	41	63	70	72	84

a) Make a **well-labeled** scatterplot. Do you think that all five specimens come from the same species? _____
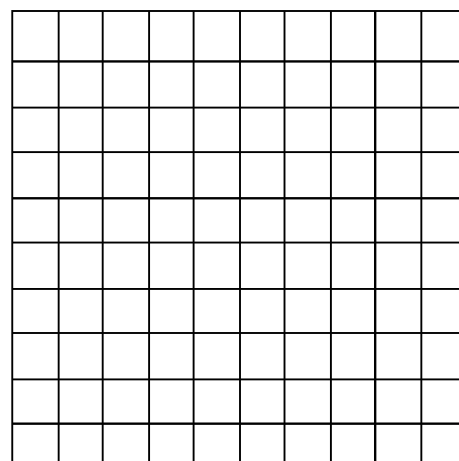
b) Describe the overall pattern. Include the direction, form and strength of the relationship.

c) Draw a line of best fit. Give an *approximation* of the slope and y- intercept. Write an equation for your line. _____

22) Define **residual**. Looking back at your graph in #21, do you think there would be any large residuals? _____ *Explain why or why not.*

23) Explain the difference between an **observational study** and an **experiment**.

24) List the **3 Principles of Experimental Design**.

-

-

-

25) The 7 types of sampling designs are:

A. voluntary response   B. convenience   C. simple random

D. stratified   E. cluster   F. multistage

G. systematic

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.