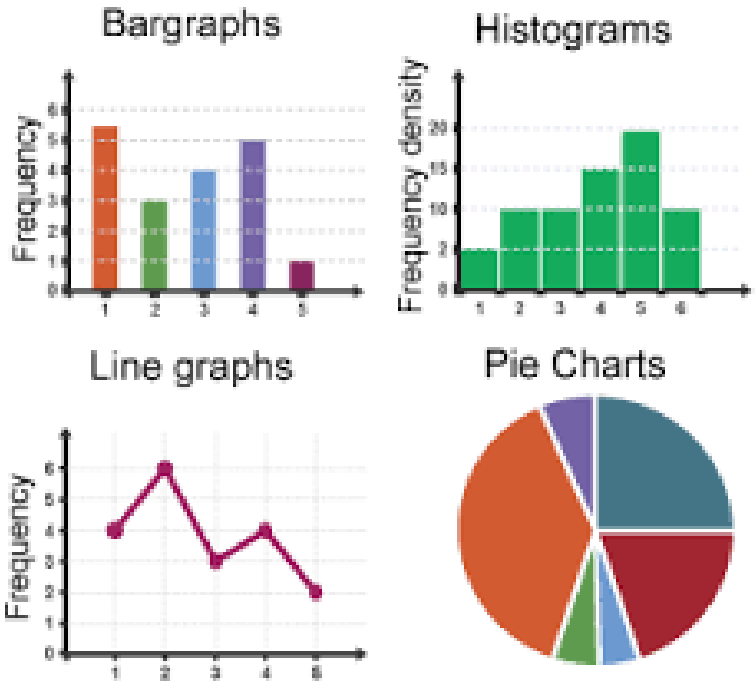
# *AP Statistics*
# *Summer Packet*



# AP Statistics Summer Assignment

Dear Incoming AP Statistics Student,

Thank you for your interest in this course. Please keep in mind that this is an advanced level course in which you will be expected to work independently and collaboratively on assignments and to keep up with course material at all times.

During the year we will look at a variety of ways to interpret and display data, to analyze results and to determine the validity of statistical findings. Hopefully, the knowledge gained in this course will help you with many of your future endeavors.

Throughout this course you will need access to a TI-84 graphing calculator, both in class and at home. If you do not have access to a calculator over the summer, you can use an online/app version, however these will not be sufficient for the duration of the course.

The material in this packet is designed to introduce you to some of the basic elements of statistics which will be the foundation of the AP Statistics course. You are expected to follow the directions and complete this packet using the resources provided. **Material from this packet will be tested within the first two weeks of class in the fall.**

I look forward to working with you next year. Enjoy the summer!

Best regards,

Mrs. Gutheil

# AP Statistics Summer Assignment

Welcome to AP Statistics, future statisticians! The purpose of this assignment is to make you familiar with the concepts of data analysis and to be able to hit the ground running on the first day of school.

The summer assignment is composed of two 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 . Follow the steps below:

   - Go to www.stattrek.com
   - Click on "AP Statistics" then on "AP Tutorial"
   - On the left side of the screen is a list of general topics. Under each general topic is a list of subtopics. You will explore the following subtopics to complete the vocabulary list. **Not all of the terms listed on the website are included on your list to define.**

| **EXPLORING DATA** | | | **EXPERIMENTATION** | |
|---|---|---|---|---|
| The Basics | Variables<br>Population vs. Sample<br>Central Tendency<br>Variability<br>Position | | Surveys | Data Collection<br>Sampling Methods<br>Bias in Surveys |
| Charts and Graphs | Patterns in Data<br>Dotplots<br>Histograms<br>Stemplots<br>Boxplots<br>Scatterplots<br>Comparing Data Sets | | Experiments | Intro to Experimentation |

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

**AP Statistics** **Summer Assignment**

Name _____

**PART 1 – VOCABULARY: Define each of the following terms from the information on the stattrek website. Words in BLUE on the website can be clicked on for more information. When asked to provide an example or a sketch of the word, please provide *a unique example NOT given on the website.***

1. Categorical Variable

   *Example:*

2. Quantitative Variable

   *Example:*

3. Discrete Variable

4. Continuous Variable

5. Univariate Data

6. Bivariate Data

7. Population

   *Example:*

8. Sample

   *Example:*

9. Median

10. Mean

    *Formula*:

11. Outlier

12. Parameter

13. Statistic

14. Range

15. Standardized Score (z-score)

    *Formula*

16. Center

17. Spread

18. Variance

    *Formula:*

19. Standard Deviation

    *Formula:*

20. Symmetry

*Sketch:*

21. Unimodal

*Sketch*:

22. Bimodal

*Sketch*:

23. Skewness

*Sketch Skewed Left:*

*Sketch Skewed Right:*

24. Uniform

*Sketch:*

25. Gaps

*Sketch:*

26. Outliers

*Sketch:*

27. Dot Plots

28. Bar Chart

29. Histogram

30. Difference between a Bar Chart and a 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 be used for Comparing Data

38. Census

39. Sample Survey

40. Experiment

41. Observational Study

42. Population Parameter

43. Sample Statistic

44. Voluntary Sample

45. Convenience Sample

46. Simple Random Sample (SRS)

47. Stratified Sampling

48. Cluster Sampling

49. Multistage Sampling

50. Systematic Random Sampling

51. Undercoverage

52. Non-response Bias

53. Voluntary Response Bias

54. Response Bias

55. Types of Response Bias

**PART 2 – PRACTICE PROBLEMS**

**A.) CATEGORICAL or QUANTITATIVE - Determine if the variables listed below are *quantitative* or *categorical*. (Q or C)**

1. Time it takes to get to school _____
2. Number of people under the age of 18 living in a household _____
3. Hair color _____
4. Temperature of a cup of coffee _____
5. Teacher salaries _____
6. Gender _____
7. Whether someone is a smoker or not _____
8. Height _____
9. Amount of oil spilled _____
10. Age of Academy Award winners _____
11. Type of depression medication _____
12. Jellybean flavors _____
13. Social Security Number _____
14. Type of meat _____
15. Number of pairs of shoes owned _____

**B.) STATISTICS – WHAT IS THAT?**

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

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 home runs Mark McGuire hit in each season from 1982 – 2001. (Round any decimal answers to 4 decimal places.)

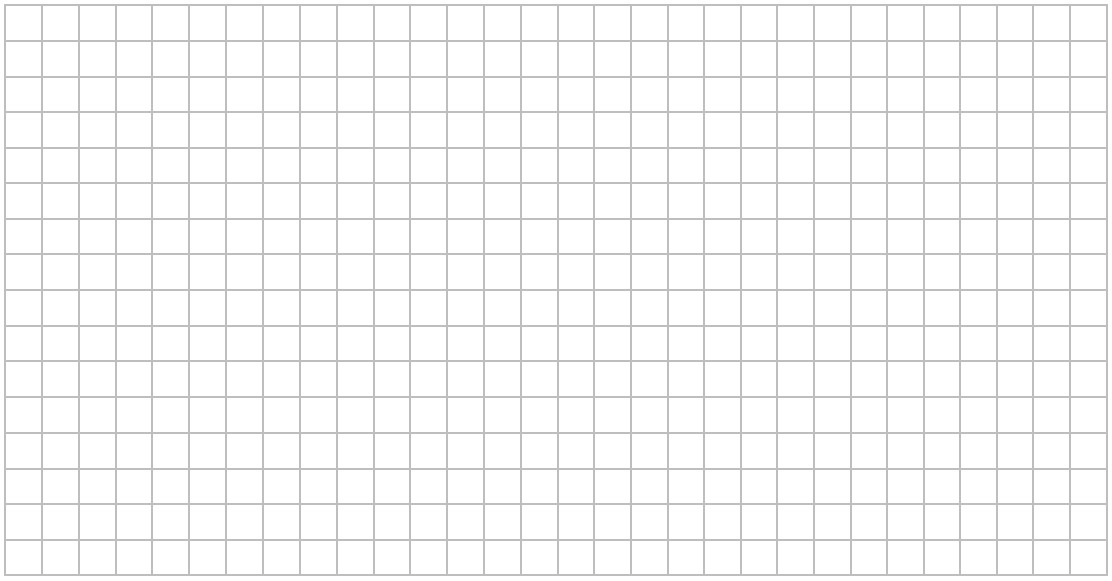
| | |
|---|---|
| Mean | |
| Minimum | |
| Q1 | |
| Median | |
| Q3 | |
| Maximum | |
| Range | |
| IQR | |

**C.) 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."

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

2. What percent of accidental deaths were from "other causes?"

3. NEATLY create a well-labeled **bar graph** of the distribution of causes of accidental deaths. Be sure to include an "other causes" bar.

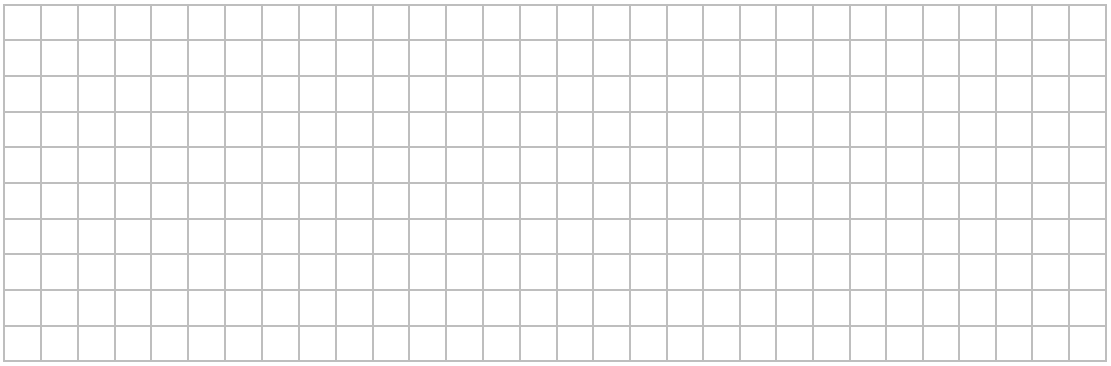


4. A pie chart is another graphical display used to show all the categories in a categorical variable relative to each other. Create a pie chart for the accidental death percentages. You may try using software or Internet source to make one and paste in the space below (*Microsoft Excel works well*).

**D.) HURRICANE'S A'COMIN' –** The data below give the number of Atlantic 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 |

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



**E.) 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 |

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

**F.) 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 |

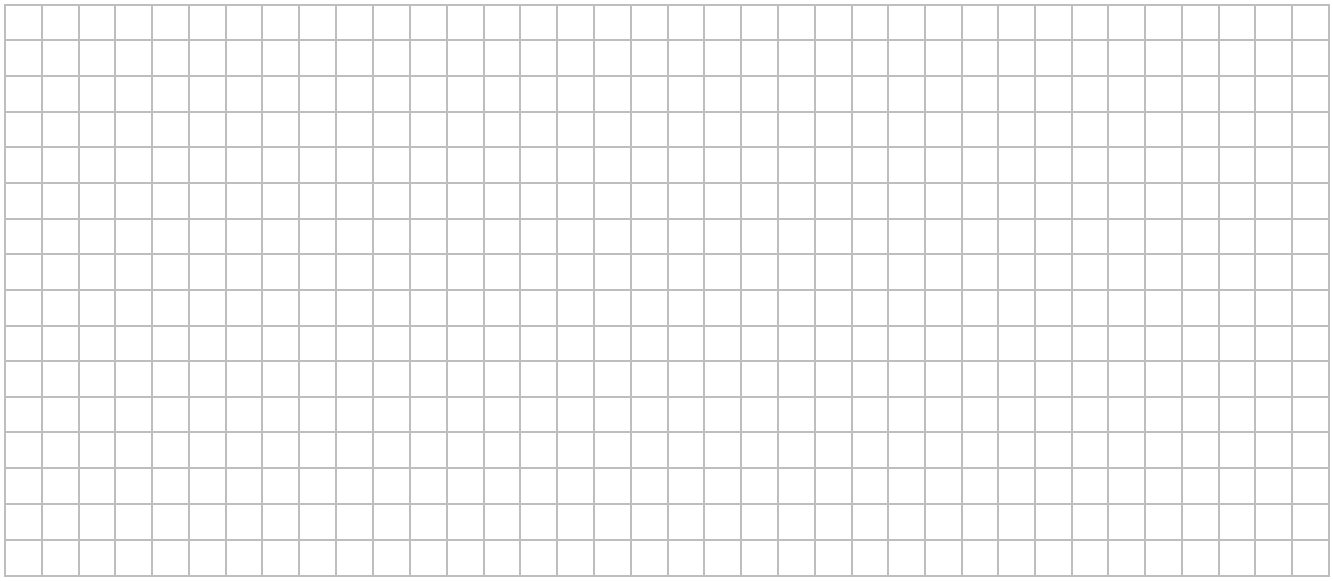
and for 20 first-year college men:

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

1. Put the data values in order for each gender. Compute the following numerical summaries for each gender. (Round any decimal answers to 4 decimal places.)

| **Women** | | | **Men** | |
|---|---|---|---|---|
| **Mean** | | | **Mean** | |
| **Minimum** | | | **Minimum** | |
| **Q1** | | | **Q1** | |
| **Median** | | | **Median** | |
| **Q3** | | | **Q3** | |
| **Maximum** | | | **Maximum** | |
| **Range** | | | **Range** | |
| **IQR** | | | **IQR** | |

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



**G.) To answer the following, refer to the readings on www.stattrek.com titled "Survey Sampling Methods".**

The 7 types of sampling designs are:

| | | | |
|---|---|---|---|
| A. Voluntary response | B. Convenience | C. Simple Random Sampling (SRS) | D. Stratified |
| E. Cluster | F. Multistage | G. Systematic | |

1. The Florida division of Weight Watchers (WW) is doing research to determine how many people on the Weight Watchers diet cheat at least once per week. They decide that anonymous surveys will give them an accurate representation but do not have time to get responses from ALL the Florida Weight Watchers participants.

   *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 Florida.

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

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

_____ IV. The Florida regional office is in Orlando, so they survey members at the WW center in Orlando.

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

_____ VI. From a numbered list of all Florida WW 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 Weight Watchers situation?

**H.) ALGEBRA!!!! – The prerequisite for AP Statistics is Algebra II. You will not 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 frequently in AP Statistics:*** $z = \dfrac{x - \bar{x}}{s}$ ***. Use your algebra skills…***

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

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

**I.) LINEAR FUNCTIONS & SCATTERPLOTS –** 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 claimed 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 situation.

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

3.) Angie 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. Include labels and scales.

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

**J.) PROBABILITY – You are expected to have a basic understanding of simple probability. If you find these problems less than intuitive, there numerous sites available online that provide basic probability explanations. I recommend YouTube – search "basic probability rules".**

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.) $\frac{1}{8}$ B.) $\frac{2}{9}$ C.) $\frac{2}{7}$ D.) $\frac{3}{8}$ E.) $\frac{1}{2}$

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 of getting heads on the first toss, and tails on the second toss? (HINT: Think about the possible outcomes when you toss a coin twice)

A.) 1/6
B.) 1/3
C.) 1/4
D.) 1/2
E.) 1

4.) If a coin is tossed twice, what in 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 (not counting "y"). _____

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