**AP Statistics Summer Assignment**

Welcome AP Statistics' scholars to AP Statistics. In this class, you will learn to describe sets of data and use your analysis to draw conclusions, in the context of the situation that produced the data. In fact, this course is like no other mathematics course you have ever taken because it concentrates your scholastic abilities much more heavily on critical thinking, logical reasoning, explaining, writing, communicating, and supporting your conclusions with appropriate calculations and visuals. In sum, this course is a combination of mathematics, English, and science concepts; communication and the ability to collaborate with others is key.

This is a college-level course and I will teach it that way. With that being said, you need to be a proactive participant in this course. This means you must be willing to work with me, and your fellow classmates. There are challenging concepts and a great deal of material that must be covered before April's end. With this in mind, you need to be committed to giving it your absolute best effort daily. In order to be ready in January, you have some preliminary work to do. In sum, this work is to get you comfortable exploring data analysis and reviewing statistical concepts.

AP Statistics is focused on four topics:

• **Exploring Data.** What patterns in a set of data do you observe? What do these patterns tell us about the data? What can we learn from analyzing these patterns?
• **Sampling and Experimenting.** What plan of action would you implement to conduct a study? Can you effectively write a research proposal for simulating a real-world situation?
• **Anticipating Patterns.** What inference can you make about random phenomena by using probability? How can you extrapolate your model into the future? What might you hope to have happen?
• **Statistical Inference.** How can you apply given parameters to your test hypothesis to determine if it is valid?

Students who take AP Statistics and perform successfully must have:

• A committed work ethic. The amount of time that should be set aside for work outside the classroom is at least equal to the amount of time spent in the classroom.
• Great listening skills. Some of the best learning comes from hearing other students ask the questions that are troubling you.
• Initiative. Students must be willing to go beyond what is asked of them and do what is expected of an AP student.
• A good ability to work with others. The course requires students, to work cooperatively to problem solve on a daily basis.
• The willingness to connect AP Statistics to other subjects, and an appreciation for how statistics connects to the world outside the classroom.
• Their own TI-84+ calculator or a scientific calculator in which he or she has become comfortable using.

Attached is your summer assignment (make sure to read the directions **thoroughly**). Although you will not see me until January, I want the completed summer assignment submitted to me the first day of school in August (Room: 324). This assignments will serve as your first test grade in the course, and will be graded based on accuracy and completeness, so please answer all questions and show all work.

Once again, I look forward to working with each and every one of you. I hope, you have a safe and great summer. Remember to bring your completed assignment to me (Room: 324) on the first day of school.

*Enjoy Your Summer!*

Educationally Yours, Dr. Randy Lattimore, Ph.D.

## Part One: Key Vocabulary List

**Directions:** Define each of the following terms in your own words. If a term requires an example or sketch, provide a unique one.

1. Categorical Variables

   *Example*:

2. Quantitative Variables

   *Example*:

3. Discrete Variables

4. Continuous Variables

5. Univariate Data

6. Bivariate Data

7. Population

   *Example*:

**8.** Sample

*Example*:

**9.** Median

**10.** Mean

*Formula*:

**11.** Outlier

**12.** Parameter

**13.** Statistics

**14.** Range

**15.** Standard 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.** What is the differences between Bar Charts and Histograms?

**31.** Stem Plot

**32.** Box Plot

**33.** Quartiles

**34.** Range

**35.** Interquartile Range (IQR)

**36.** Four ways to describe data sets (SOCS)

**37.** What types of Graphs can be used for comparing Data?

## Part Two: Practice Problems

### CATEGORICAL OR QUANTITATIVE?

**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 years old living in a household
3. Hair color
4. Temperature of a cup of coffee
5. Teachers' 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 Major League Baseball (MLB) season from 1982-2001.

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

| | |
|---|---|
| Mean | |
| Minimum | |
| Maximum | |
| Median | |
| $Q_1$ | |
| $Q_3$ | |
| Range | |
| IQR | |

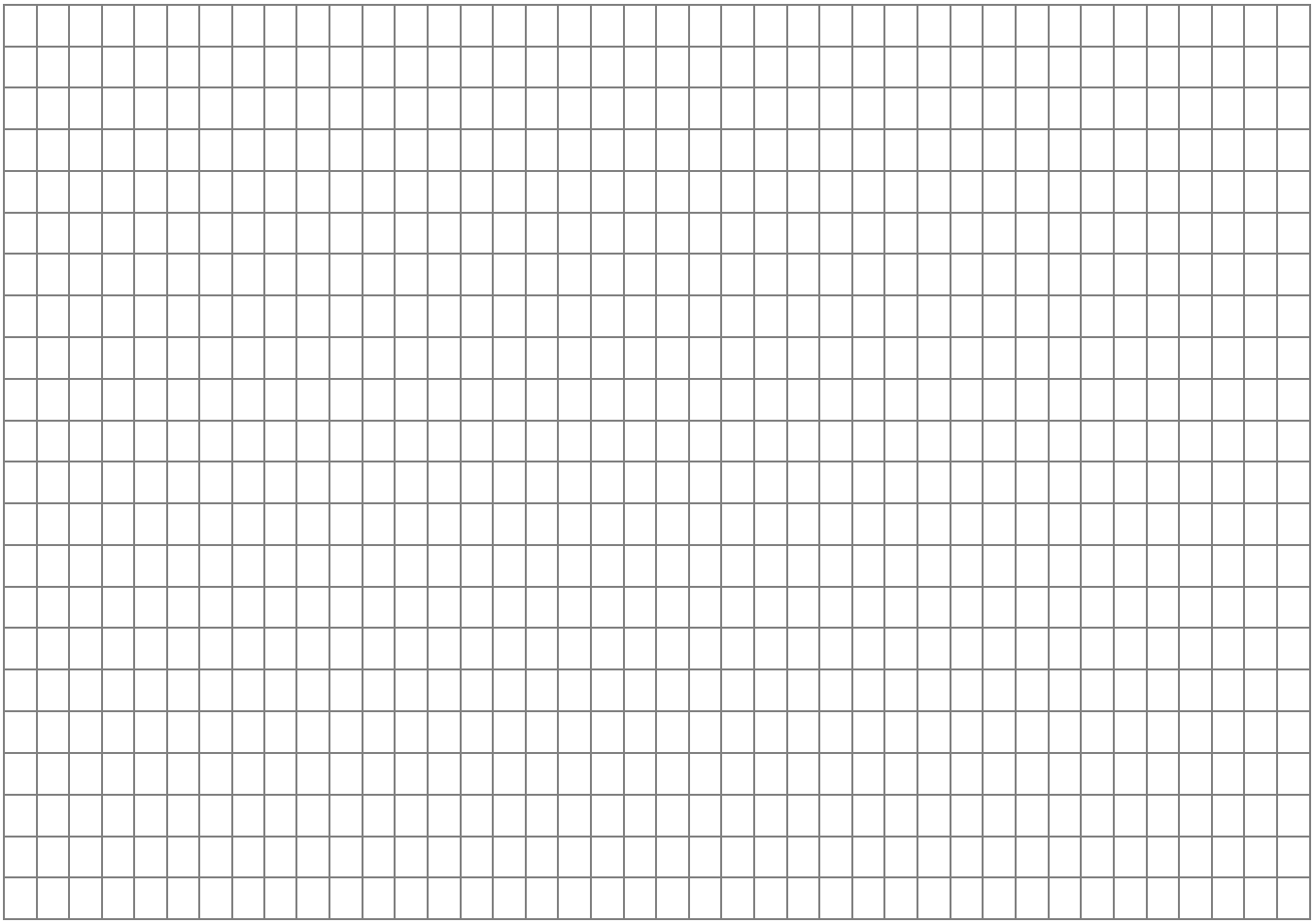
## ACCIDENTAL DEATHS

In 1997 there were 92353 deaths from accidents in the United States. Among these were 42340 deaths from motor vehicle accidents, 11858 from falls, 10163 from poisoning, 4051 from drowning, and 3601 from fires. The rest were "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.** Create a neat and well-labeled bar graph of the distributions of causes of accidental deaths below. Be sure to include an "other causes" bar.



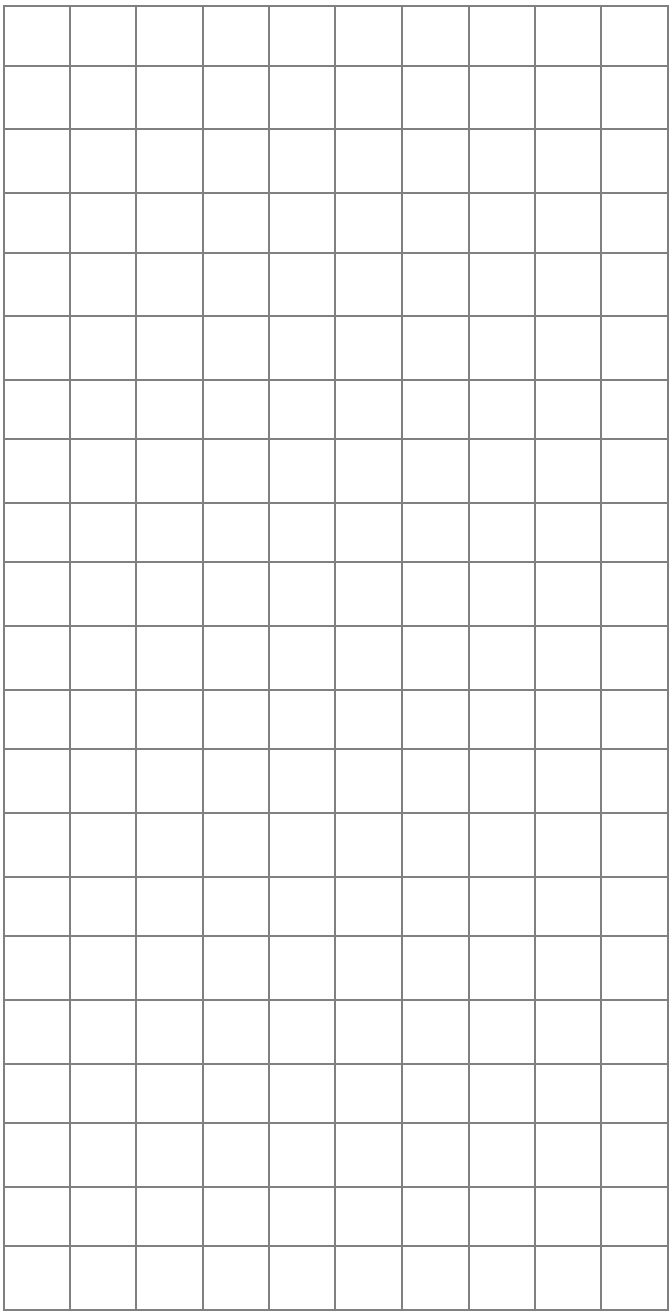
**d.** A pie chart is another graphical display used to show all the categories in a categorical variable relative to each other. Create a pie chart for the accidental death percentages. You may use a software or Internet source to make one and attach it. Microsoft Excel works well.

**IT IS A TWISTA**

The data below gives the number of hurricanes that happened each year from 1944 through 200 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 | | | | | | | | | | | | | | | | |

Make a dot plot 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 (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 |

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

## WHERE DO ELDERLY FOLKS LIVE?

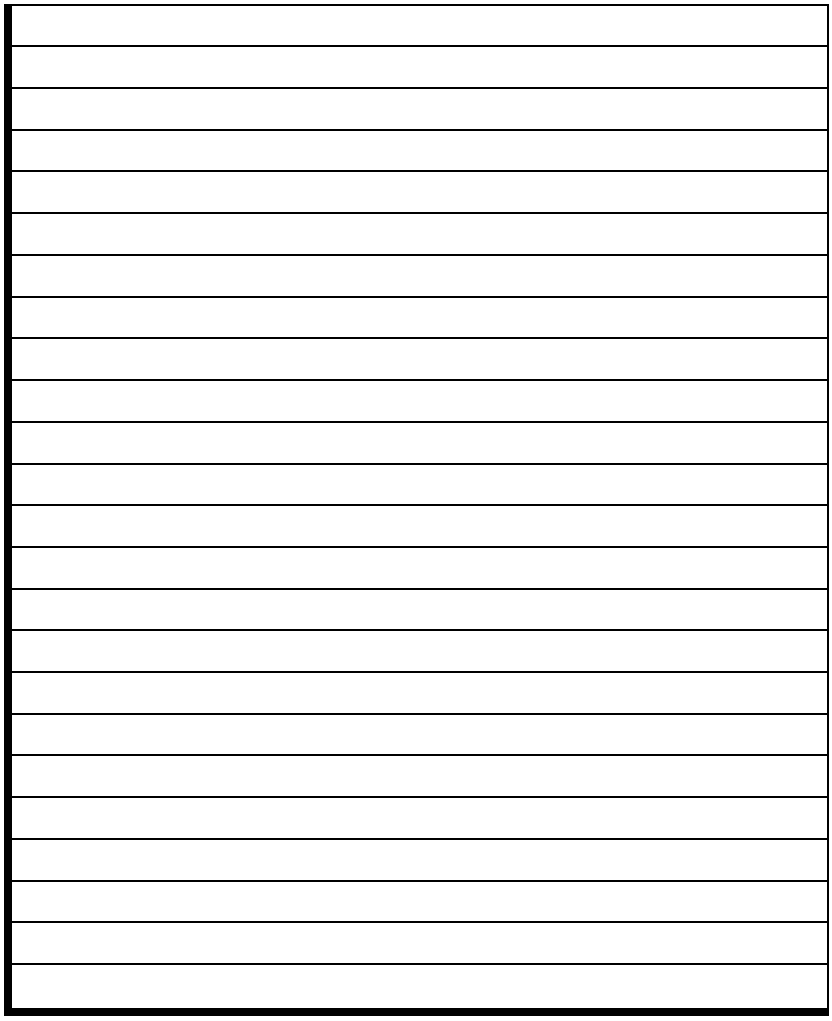
This table gives the percentage of residents aged 65 or older in each of the 50 states.

| 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 each other 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 the 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*.

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



**Note:** The prerequisite for AP Statistics is Algebra 2. FYI: You will not find very much equation solving in this course. I suggest that you do a quick review of Algebra 1 and Algebra 2 content that review will be helpful.

***To answer the following refer to the readings on www.stattrek.com, and the "Survey Sampling Methods" which is a short video.***

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 for 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 $20^{th}$ 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 members and survey all members with those corresponding numbers.

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

Here is a formula that is used often in AP Statistics: $z = \frac{x-\bar{x}}{s}$. Use your Algebraic skills, to solve these problems.

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

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

**Note:** 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 per 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.

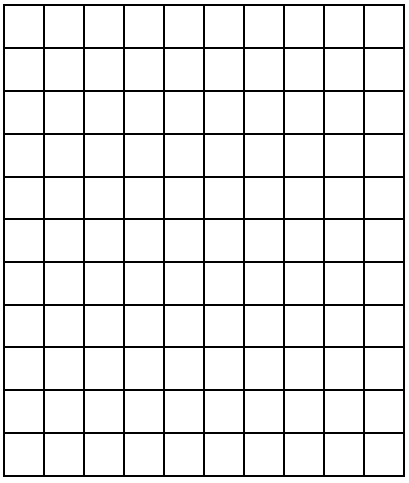
**2.** The following equation can be used to predict the average height of boys anywhere bet anywhere between birth and 15 years old: $y = 2.79x + 25.64$, where $x$ is the age (in years) and $y$ is the height (in inches).

**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 of the situation.

**3.** Hillary 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 scatter plot of the data.



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

***Note:*** 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 three times; each junior's name two times; and each sophomore's name one 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.** $\frac{1}{8}$

**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.** $\frac{1}{6}$ **B.** $\frac{1}{3}$ **C.** $\frac{1}{4}$ **D.** $\frac{1}{2}$ **E.** 1

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

**A.** $\frac{1}{8}$ **B.** $\frac{1}{6}$ **C.** $\frac{1}{4}$ **D.** $\frac{1}{2}$ **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 is 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**: _____, _____, _____, _____

If you are done with the summer assignment, then take a deep breath, pat yourself on the back

**AND**

*Enjoy the rest of your summer vacation and I will see each of you on the first day of school*!