AP Statistics / Statistics - Summer Review Packet – 2022

DUE **AUGUST 31, 2022** in the appropriate Google Classroom

In preparation for the Fall Semester, this assignment is required to prepare you for your Statistics course.

About Statistics:

Statistics is the branch of mathematics pertaining to the practice and science of collecting and analyzing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample. Students use a graphing calculator as an integral tool in analyzing data and modeling functions to represent real world applications. Saint Dominic Academy recommends a **TI-84 Plus or TI-84 Plus CE** graphing calculator. If a student have a graphing calculator other than a TI-84, she will be responsible for learning how it operates. The calculator will be used throughout your high school and college career.

Expectations of the Summer Packet:

The problems in this packet are designed to help you review topics that are important to your success in Statistics. ***All work must be shown for each problem***. The problems should be done correctly, not just attempted.

All work should be completed and ready to turn in to the Google Classroom.

Notes: The internet is a great resource… use it!
Some helpful sites:
www.mathisfun.com (click on data)
www.khanacademy.com
www.profrobbob.com (Tarrou's ChalkTalk)

.
Enjoy your summer!

Contents

Section 1. Introduction to Statistics

Sta-tis-tics

1: a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data (statistics with capital "S")

2: a collection of quantitative data (statistics with lowercase "s")

The study of Statistics is unlike any Math class that you have taken before. Advanced Placement Statistics acquaints students with the major concepts and tools for collecting, analyzing, and drawing conclusions from data. Ideas and computations presented in this course have immediate links and connections with actual events and real world situations.

There is heavy emphasis on the ability to think, reason, explain and support your conclusions as opposed to just performing rudimentary computations.

Writing and reading comprehension are important components of the course as a large part of this class is ***clear communication and interpretation of data.***

Statistics is separated into two branches:

Descriptive statistics uses the data to provide descriptions of the population, either through numerical calculations or graphs or tables.

Inferential statistics makes inferences and predictions about a population based on a sample of data taken from the population in question.

Statistics requires **skilled writing**. Please answer the following question in several descriptive and well-written sentences.

1. **What are your expectations coming into this class? What do you expect to learn?**

Section 2. Data in Context

Statistics is the ***most verbal*** (reading, writing, listening and speaking) of the mathematics disciplines. Expect read and write a lot more than in any of your other mathematics classes. Your ability to read and interpret what you read is extremely important.

2. Read the following passage and then answer the questions.

"Teen Automobile Crash Rates Are Higher When School Starts Earlier"

ScienceDaily (June 10, 2010) — Earlier school start times are associated with increased teenage car crash rates, according to a research abstract presented June 9, 2010, in San Antonio, Texas, at SLEEP 2010, the 24th annual meeting of the Associated Professional Sleep Societies LLC.

Results indicate that in 2008 the teen crash rate was about 41 percent higher in Virginia Beach, Va., where high school classes began at 7:20 a.m., than in adjacent Chesapeake, Va., where classes started more than an hour later at 8:40 a.m. There were 65.4 automobile crashes for every 1,000 teen drivers in Virginia Beach, and 46.2 crashes for every 1,000 teen drivers in Chesapeake.

"We were concerned that Virginia Beach teens might be sleep restricted due to their early rise times and that this could eventuate in an increased crash rate," said lead author Robert Vorona, MD, associate professor of internal medicine at Eastern Virginia Medical School in Norfolk, Va. "The study supported our hypothesis, but it is important to note that this is an association study and does not prove cause and effect."

The study involved data provided by the Virginia Department of Motor Vehicles. In Virginia Beach there were 12,916 drivers between 16 and 18 years of age in 2008, and these teen drivers were involved in 850 crashes. In Chesapeake there were 8,459 teen drivers and 394 automobile accidents. The researchers report that the two adjoining cities have similar demographics, including racial composition and per-capita income.

3. Please answer the questions: the five W's (Who, What, When, Where, Why) and How

a) ***Who*** is being studied?

b) ***What*** is being recorded? (In other words, what is/are the variable(s)?

c) ***When*** was the data collected?

d) ***Where*** was the data collected?

e) ***Why*** do you think this data was collected and analyzed?

f) ***How*** was the data collected and analyzed? In other words, what methods were used?

g) Why do you think the authors of the study mentioned that "it is important to note that this is an association study and does not prove cause and effect?"

4. Article/Journal Review

Find one (1) newspaper, magazine or Internet article that include statistical concepts.

This article should include things like graphs, charts or a discussion of studies. It should report conclusions made as a result of looking at data or studies.

Make note of the statistics mentioned and answer the questions listed below using at least one-two paragraphs for each question.

You must provide links to the articles you have chosen or scan/photograph the article you use and attach them to your written submission.

Questions:

1. What was the purpose of the article? Why was it written?

2. Were any conclusions stated? If so, what were they?

3. Is the article convincing? Do you believe the stated results? Explain

4. Include any other relevant or interesting information about the article.

Section 3. Displaying and Describing Qualitative Data

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.

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

4. What percent of accidental deaths were from "other" causes? Show how you determined your answer.

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

6. NEATLY create a well-labeled **pie chart** (circle graph) of the distribution of causes of accidental deaths. Be sure to include an "other causes" bar. Be sure the pie "wedges" are proportionally sized to each category.

Section 4. Displaying and Describing Quantitative Data

Quantitative data is **data** expressing a certain quantity, amount or range. Usually, there are measurement units associated with the **data**, e.g. meters, in the case of the height of a person. It makes sense to set boundary limits to such **data**, and it is also meaningful to apply arithmetic operations to the **data**.

Measures of Central Tendency

Statistic	Definition
Mean	Sum of the data divided by the number of items in the data set
Median	Middle number of the ordered data (if there are an odd number of data points), or the mean of the middle two numbers (if there are an even number of data point)
mode	Number or numbers that occur most often

Example 1: Jason recorded the number of hours he spent watching television each day for a week. Find the man, median, and mode for the number of hour.

Mon	Tues	Wed	Thurs	Fri	Sat	Sun
2	3.5	3	0	2.5	6	4

$$\boldsymbol{mean} = \frac{sum\ of\ hours}{number\ of\ days} = \frac{2+3.5+3+0+2.5+6+4}{7} = 3$$ The mean is 3 hours.

To find the median, first put the numbers in order from least to greatest.

0 2 2.5 **3** 3.5 4 6

Because there are an odd number of data points, the median is the number located in the middle.
The median is 3 hours.

Each of the data points occurs only once in the set so there is no mode.

Find the mean, median, and mode for each set of data.

7. Maria's test scores

92, 86, 90, 74, 95, 100, 90, 50

8. Rainfall last week in inches

0, 0.3, 0, 0.1, 0, 0.5, 0.2

Term	Definition
Minimum	The lowest value in the data set.
Maximum	The highest value in the data set.
Range	The difference between the maximum and minimum. Range = max - min
Median	The value that separates the data set in half.
Lower quartile	The median of the lower half of the data set. Also known as the first quartile (Q1)
Upper quartile	The median of the upper half of the data set. Also known as the 3rd quartile (Q3)
Interquartile range IQR	The difference between the upper quartile and lower quartile. IQR – Q3 – Q1
Outlier	Data that are more than 1.5 times the value of the IQR beyond the quartiles. Lower outliers < Q1 – 1.5 IQR upper outliers > Q3 + 1.5 IQR

A **box-and-whisker plot** is used to show the general layout of a set of data – where most of the numbers fall. It shows the median of the whole set, the median of both halves, and the highest and lowest numbers in the data set.

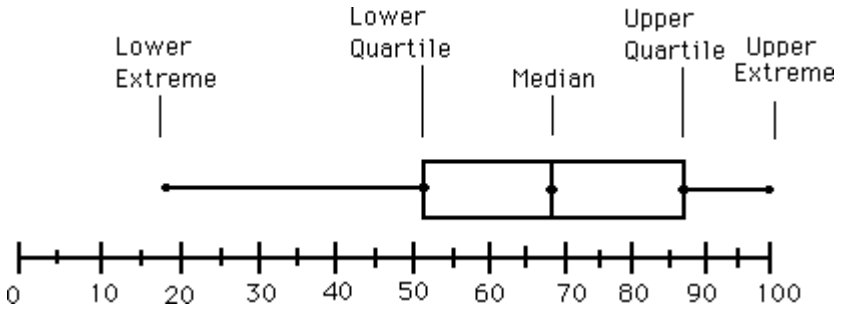
Data is divided into four parts, or quartiles. The median (Q_2) of the entire set of numbers is the center of the "box." The numbers less than the median are then divided into two sections by finding the median of those numbers. That new median is called the first (or lower) quartile (Q_1), and is the marker for the left side of the box. Finding the median of the values greater than the original median gives you the third (or upper) quartile value (Q_3); this is the marker for the right side of the box. The lowest value is the end of the left whisker, and the highest value is the end of the right whisker.

Example: 18, 27, 34, 52, 54, 59, 61, 68, 78, 82, 85, 87, 91, 93, 100

↑ ↑ ↑

First Quartile ($Q_1 = 52$) | Median ($Q_2 = 68$) | Third Quartile ($Q_3 = 87$)

To graph this data, we create a number line including at least all of the values in the set of data. The box-and-whisker plot is placed on or near the number line as indicated above.

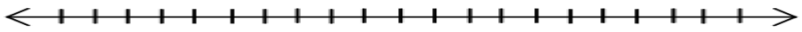


The Interquartile Range (IQR) is the value of $Q_3 - Q_1$. In the example, IQR = 87 – 52 = 35.

9. A survey was conducted by Mrs. Johnson in her math class. She asked, "How many hours did you spend on homework this past week?" The responses are below:

13	4	10	8	9
0	2	5	15	12
9	16	11	8	4
3	8	10	6	8
11	6	7	11	9

Draw a box and whisker plot for this set of data.



a. What is the median number of hours spent on homework? _____

b. What is the Interquartile Range? IQR = _____

c. What can you infer about the grades of the students in Mr. Johnson's mathematics class from the box and whisker plot representing the number of hours of homework done in a week?

Section 5. Combinatorics

Counting Outcomes A tree diagram is a visual display used to find the number of outcomes given a number of choices. Another method that relates the number of outcomes to the number of choices is the Fundamental Counting Principle.

Fundamental Counting Principle	If event *M* can occur in *m* ways and is followed by event *N* that can occur in *n* ways, then the event *M* followed by *N* can occur in *m n* ways.
	Example 1: The Shoe store sells 9 different styles of running shoes, each available in 2 colors. How many combinations of color and style are there? 9 x 2 = 18 combinations of color and style

Permutations An arrangement or listing in which ***order or placement is important*** is called a permutation. For example, the arrangement AB of choices A and B is different from the arrangement BA of these same two choices.

Factorial	Definition: n factorial = n! = (n)(n-1)(n-2)(n-3)…(1) **Example 2:** 7! = (7)(6)(5)(4)(3)(2)(1) = 5040
Permutations	The number of permutations of *n* objects taken *r* at a time is $P(n, r) = \frac{n!}{(n-r)!}$.
	Example 2: Find *P* (6, 2). $P(n, r) = \frac{n!}{(n-r)!}$ Permutation Formula $P(6, 2) = \frac{6!}{(6-2)!}$ $n = 6, r = 2$ $= \frac{6!}{4!}$ Simplify. $= \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1}$ Definition of factorial $= 6 \cdot 5$ or 30 Simplify. There are 30 permutations of 6 objects taken 2 at a time.

Combinations An arrangement or listing in which ***order is not important*** is called a combination.
For example, AB and BA are the same combination of A and B.

Combinations	The number of combinations of n objects taken r at a time is $C(n, r) = \frac{n!}{(n-r)!\,r!}$.
	Example 3: A club with ten members wants to choose a committee of four members. Six of the members are women, and four are men. How many different committees are possible? $C(n, r) = \frac{n!}{(n-r)!\,r!}$ Combination Formula $C(10, 4) = \frac{10!}{(10-4)!\,4!}$ $n = 10, r = 4$ $= \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!}$ Divide by the GCF 6!. $= 210$ Simplify. There are 210 ways to choose a committee of four when order is not important.

Fundamental Counting Principle

10. If a sandwich shop has 3 different types of meat, 4 types of bread, and 3 different type of cheese. How many types of sandwiches can you create if you must have meat, bread and cheese on each?

11. If a person has 4 pairs of shoes and 6 pairs of socks, then how many shoe-sock combos are possible?

Permutations: Order matters;

12. How many ways can we pick a President, Treasurer and Secretary from a club of 20 members?

Combinations: Order doesn't matter;

13. How many ways can we select a committee of 3 people from a club of 20 members?

14. How many ways can we pick 4 winners for 8 prizes if each prize is the same?

Section 6. Probability - Independent and Dependent Events

Compound events, or two or more simple events happening together, can be independent or dependent.

Events are **independent events** if the probability of one event does not affect the probability of the other.

Events are **dependent events** if one event in some way changes the probability that the other occurs.

The following are the **Multiplication Rules for Probability**.

Probability of Two Independent Events	$P(A \text{ and } B) = P(A) \cdot P(B)$
Probability of Two Dependent Events	$P(A \text{ and } B) = P(A) \cdot P(B \text{ following } A)$ $P(A \text{ and } B) = P(A) \cdot P(B\|A)$

Determine whether the events are *independent* or *dependent*. Then find the probability.

15. A king is drawn from a deck of 52 cards, then a coin is tossed and lands heads up.

16. A spinner with 4 equally spaced sections numbered 1 through 4 is spun and lands on 1, then a die is tossed and rolls a 1.

17. A red marble is drawn from a bag of 2 blue and 5 red marbles and not replaced, then a second red marble is drawn.

18. A red marble is drawn from a bag of 2 blue and 5 red marbles and then replaced, then a red marble is drawn again.

Section 7. General Mathematics Review

19. Write equations of the horizontal and vertical lines that pass through the point (-3, 4).

Horizontal line _____ Vertical line _____

20. Find the slope and y-intercept of the line.

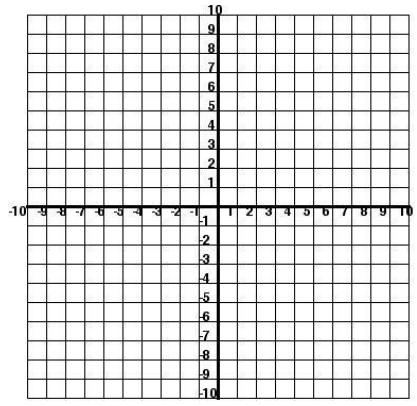
a. $y = \frac{2}{3}(2x - 4)$

b. $\frac{1}{3}y - 6x = 4$

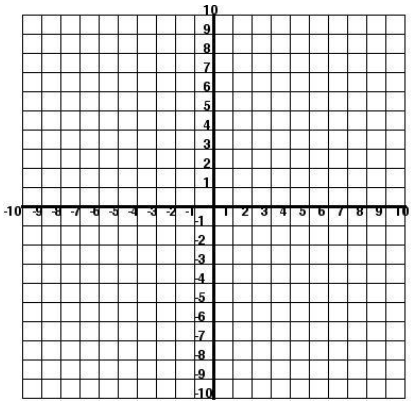
21. Find the slope and write the equation of the line containing the points (6, -2) and (0, 5)

22. Plot the data using a scatter plot then decide if the data is linear, exponential, quadratic, or absolute value. Sketch the graph here.

a. (-3, 4) (-2, 3.5) (-1, 3) (0, 2.5) (1, 2) (2, 1.5) (3, 1)



b. (-3, 4) (-2, 3) (-1, 2) (0, 1) (1, 2) (2, 3) (3, 4)



23. For the function $f(x) = 3x^2$ find the requested values.

f (– 3) =

f (0) =

f (2) =

24. Solve each equation:

a. $2\sqrt{x} + 9 = 21$

b. $\sqrt{2x+10} = x+1$

c. $2|x-1| = 14$

d. $\frac{1}{3}n + 3 = n - 2$

h. $x^2 - 8x + 7 = 0$

i. $\frac{m}{12} + \frac{5}{6} = \frac{5}{24}$

Section 8. How to Lie with Statistics

In 1954, former *Better Homes and Gardens* editor and active freelance writer Darrell Huff published a slim (142 page) volume which overtime would become the most widely read statistics book in the history of the world. Although this 1954 book is old, it is a classic (now in its 50th printing) and very entertaining.

Your assignment is to read and take notes on this short book, How to Lie with Statistics, by Darrell Huff. We will be discussing the book in class when you return in September and then you will be quizzed on the book. This book is available at the library, through Amazon.com, or as a PDF file through the link below:

http://www.horace.org/blog/wp-content/uploads/2012/05/How-to-Lie-With-Statistics-1954-Huff.pdf

25. Read the introductory quotes just past the cover page and before the Introduction. What do you think that H.G. Wells meant when he said "**Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write**."

26. Read the **Introduction**. Select a quote in the introduction and write a paragraph reflecting on what it means to you.

27. What do you think the word "semiattachment" means on page 8 when Huff wrote:

> "Like many a more sophisticated statistic it was guilty of semiattachment: It assumed that newspaper space given to crime reporting is a measure of crime rate."

(If you need some help, look at Chapter 7 on **The Semiattached Figure** starting on page 74.)

Article/Journal Review

Collect 1 newspaper, magazine or Internet article that include statistical concepts.

This article should include things like graphs, charts or a discussion of studies. It should report conclusions made as a result of looking at data or studies.

Make note of the statistics mentioned and answer the questions listed below using at least one-two paragraphs for each question.

You must provide links to the articles you have chosen or scan/photograph the article you use and attach them to your written submission.

This assignment will be collected on the first day of class.

Questions: 1. What was the purpose of the article? Why was it written?
2. Were any conclusions stated? If so, what were they?
3. Is the article convincing? Do you believe the stated results? Explain
4. Include any other relevant or interesting information about the article.