West Essex High School
AP Statistics Summer Assignment 2023-2024

AP Statistics is the high school equivalent of a one semester, introductory college statistics course. In this course, you will be expected to learn:

- To produce convincing oral and written statistical arguments, using appropriate terminology, in a variety of applied settings.
- When and how to use technology to aid you in solving statistical problems
- Essential techniques for producing data (surveys, experiments, observational studies), analyzing data (graphical & numerical summaries), modeling data (probability, random variables, sampling distributions), and drawing conclusions from data (inference procedures – confidence intervals and significance tests)
- To become critical consumers of published statistical results by heightening your awareness of ways in which statistics can be improperly used to mislead, confuse, or distort the truth.

Supplies

The course is cumulative and we will cover the entire textbook. It is imperative that you stay organized. You will need to retain warm ups/homework quizzes, formula sheets, guided notes and homework. You should have a 3-ring binder and 100 3 by 5 inch index cards to practice your interpretations of key terms and analyses. You must have a TI-84 calculator.

Google Classroom

Join the summer classroom to post your assignment and ask questions. Class code __________.

This assignment is due BEFORE school starts and counts as a homework grade.

Part 1

Watch the following TED Talks and comment on each one using no more than 100 words per talk.

Arthur Benjamin: Teach statistics before calculus! | TED Talk

How statistics can be misleading - Mark Liddell | TED-Ed

Can you spot the problem with these headlines? (Level 1) - | TED-Ed

Part 2 Take a Survey

Go to the ***This is Statistics*** page and take a quiz to determine what kind of statistician you could be based on your interests. Include the results in your response to Part 1.

What Kind of Statistician Could You Be?

Part 3 Start Interpreting

In AP Statistics, it is important to be able to interpret what statistical values mean in context. Below are sentence frames with examples that we will be using in the beginning of the year. Familiarize yourself with these terms (feel free to look them up if you do not know the basic definition) and their interpretations.

Intepretation Guide

Standard Deviation (SD): The **context** typically varies by SD from the mean of mean.

*Example: The **height of power forwards in the NBA** typically varies by 1.52 inches from the mean of 80.1 inches.*

Percentile: percentile % of **context** are less than or equal to value.

*Example: 75% of **high school student SAT scores** are less than or equal to 1200.*

z-score: Specific value with **context** is z-score standard deviations above/below the mean.

*Example: A **quiz score** of 71 is 1.43 standard deviations below the mean. (z = -1.43)*

Describe a distribution: Be sure to address shape, center, variability, and outliers (in **context**).

*Example: The **distribution of student height** is unimodal and roughly symmetric. The mean height is 65.3 inches with a standard deviation of 8.2 inches. There is a potential upper outlier at 79 inches and a gap between 60 and 62 inches.*

Correlation (r): The linear association between **x-context** and **y-context** is weak/moderate/strong (strength) and positive/negative (direction).

*Example: The linear association between **student absences** and **final grades** is fairly strong and negative. (r = −0.93)*

Residual: The actual **y-context** was residual above/below the predicted value when **x-context** = #.

*Example: The actual **heart rate** was 4.5 beats per minute above the number predicted when **Matt ran** for 5 minutes.*

y-intercept: The predicted **y-context** when x = 0 **context** is y-intercept.

*Example: The predicted **time to checkout at the grocery store** when there are 0 **customers in line** is 72.95 seconds.*

Slope: The predicted **y-context** increases/decreases by slope for each additional **x-context**.

*Example: The predicted **heart rate** increases by 4.3 beats per minute for each additional **minute jogged**.*

Standard Deviation of Residuals (s): The actual **y-context** is typically about s away from the value predicted by the LSRL.

*Example: The actual **SAT score** is typically about 14.3 points away from the value predicted by the LSRL.*

Coefficient of Determination (r^2): About $\underline{r^2}$% of the variation in **y-context** can be explained by the linear relationship with **x-context**.

*Example: About 87.3% of variation in **electricity production** is explained by the linear relationship with **wind speed**.*

Describe the relationship: Be sure to address strength, direction, form and unusual features (in **context**).

*Example: The scatterplot reveals a moderately strong, positive, linear association between the **weight and length of rattlesnakes**. The point at (24.1, 35,7) is a potential outlier.*

Probability $P(A)$: After many many **context**, the proportion of times that **context A** will occur is about P(A).

Example: P(heads) = 0.5.
*After many many **coin flips**, the proportion of times that **heads** will occur is about0.5.*

Conditional Probability $P(A|B)$: Given **context B**, there is a P(A|B) probability of **context A**.

Example: P(red car | pulled over) = 0.48.
*Given that **a car is pulled over**, there is a 0.48 probability of **the car being red**.*

Expected Value (Mean, μ): If the random process of **context** is repeated for a very large number of times, the average number of **x-context** we can expect is expected value. (decimals OK).

*Example: If the random process of **asking a student how many movies they watched this week** is repeated for a very large number of times, the average number of **movies** we can expect is 3.23 movies.*

Sentence Frames and Examples from STATS MEDIC