Project 1

# Question 1

## Excel Spreadsheet



## Calculate the mean, median, and the mode

|  |  |
| --- | --- |
| **Mean** | 13.69 |
| **Median** | 13.00 |
| **Mode** | 13.00 |
|  |  |

## Using the mean as your parameter estimate for your model, calculate the number of errors, the sum of the absolute errors, and the sum of the squared errors.

|  |  |
| --- | --- |
| **Mean Parameter** |  |
| Number of Errors | 13 |
| Sum of Absolute Errors | 23.69 |
| Sum of Squared Errors | 64.77 |

## Now use the median as your parameter estimate and calculate the same three error terms

|  |  |
| --- | --- |
| **Median Parameter** |  |
| Number of Errors | 10 |
| Sum of Absolute Errors | 23.00 |
| Sum of Squared Errors | 71.00 |

## Use the mode as your parameter estimate and calculate the same three error terms

|  |  |
| --- | --- |
| **Mode Parameter** |  |
| Number of Errors | 10 |
| Sum of Absolute Errors | 23.00 |
| Sum of Squared Errors | 71.00 |

## Look across your answers to (b), (c), and (d) to see for each error term, which estimate gives the lowest error? That is, which of the mean, median, and mode minimizes the sum of squared errors and by how much? Which one minimizes the sum of absolute errors and by how much? And which one minimizes the count of errors and by how much?

* The estimate that gives the lowest error is the mean parameter as the sum of squared errors is 64.77 compared to 71 for both the median and mode parameter estimates.
* The parameter that minimizes the sum of absolute errors are both the median and mode, which is 23 compared to 23.69 for the mean parameter.
* The parameter that minimizes the count of errors are both median and mode at 10.

# Question 3

“HR Data.csv” contains 311 rows and 36 columns. First, open this data file in RStudio or Jamovi. Next, obtain estimates of the mean, median, variance, and standard deviation of two variables (SALARY, ENGAGEMENTSURVEY). Report these along with a one sentence interpretation of what the values mean to you

## Answer

A screenshot of a graph

Description automatically generated

These values are meaningful as the number of unique rows in the data is 311. There seems to be outliers that are skewing the mean as the mean salary and engagement survey numbers seem statistically significantly larger than the median salary and engagement survey numbers.

# Question 4

Compare the mean SALARY for Men and Women. Briefly describe these results, being sure to indicate whether it APPEARS (we don't know how to do the formal test yet) that it would be useful to make predictions of SALARY conditional on employee sex (M or F).

## Answer

A screenshot of a paper

Description automatically generated

The mean male salary from this data set is $70,629 and the mean female salary is $67,787. The difference is small enough that it appears it may not be statistically significant, so we can’t say for sure that sex would be a useful predictor for salary.