



**SUMMARY OF PART 1**

* Collect data and construct a well detailed histogram, making sure to describe the shape and any unusual features (if any exist)
* Calculate mean, median, mode, range, 5-number summary, standard deviation.
* Check your data for any outliers.
* Sketch and label a normal model for your data regardless of what your distribution looks like.
* Check to see what percentage of your data is within 1, 2, and 3 standard deviations of the mean and compare those percentages to the empirical rule.
* Make sure to describe your distribution (histogram) Shape, Center, Spread and explain why you chose those measures.
* Do the statistics calculated (mean/median etc.) support the shape of your histogram or would you have expected the shape to be different.

**Project Part 2: Linear Regression**

**The Project:** You will perform a linear regression analysis with residual diagnostics using a data set that you collect. With this data you will create and analyze a scatter plot and answer several questions about the data

**Gathering Data:**

* Select a sample size of at least 25 (x,y) values
* Make sure to specify your explanatory and response variables.

**Analyzing Data**

* Explain if you believe your two quantitative variables should have an underlying (cause/effect) relationship
* Use this data to create and analyze a scatter plot and a residual plots using DESMOS (or some similar spreadsheet program) and the steps we discussed in class.

**Summary**

* Explain your data:
* Who/What it represents?
* Where/How did you collect it?
* What units is the data in?
* Analysis of the scatter plot (to include):
  + An image of the scatter plot with appropriate scales and labels, including a regression line and equation and r2 value.
  + A description of the scatter plot (Shape, Strength, Type of Correlation)
  + A calculation and interpretation of r
  + A calculation and interpretation of the slope and r2
  + A residual plot and an analysis of whether a linear model is appropriate for the data
  + Identification of outliers and influential data points
  + Explanation as to how you feel about using your regression equation to make predictions.

|  |  |
| --- | --- |
| **Rubric for Statistics Project Part 2** | **Points Possible** |
| The process of collecting data from the web or other was completed correctly | 6 |
| The amount of data gathered was appropriate (at least 25 points) | 4 |
| Quantitative and not categorical variables were used | 4 |
| The data hypothesized to have or not have a relationship is reasonable | 12 |
|  | **26** |
| Scatter plot is drawn correctly | 6 |
| LSRL lines are calculated correctly and included on the scatter plot | 8 |
| R-squared and “r” are calculated correctly and included | 6 |
| Residual plots are drawn correctly | 6 |
|  | **26** |
| Slope is interpreted correctly for the graph | 8 |
| R squared is interpreted correctly for the graph | 8 |
| The appropriateness of a linear model is discussed correctly | 8 |
| Individual residuals are calculated and interpreted correctly | 8 |
| Comments on possible outliers or influential points | 8 |
| Explanation about using equation for predictions | 8 |
|  | **48** |
| **Total:** | 100 |

**Part III of Project**

Conducting an actual hypothesis test including all aspects of the test

(Details to come later)