Data Analyst Nanodegree

Statistic ,Project1:

1. What is our independent variable? What is our dependent variable?

   The independent variable is the colour of words ether it's same or not. The dependent variable is the time it takes to name the ink colors.

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

   Null hypotheses: The average time it takes to name the ink colors are the same for congruent words and incongruent words.

   Note: (μ_{cong} is time to name the ink colours in congruent words condition and μ_{incong} is the time to name incongruent words condition).

   H_0: $\mu_{cong} = \mu_{incong}$

   Alternative hypotheses : : The average time it takes to name the ink colors in incongruent words more than the average time for the congruent words .

   H_1: $\mu_{incong} > \mu_{cong}$

   I will use T-test because there is no standard deviation of population are provide in this question and the sample is less than 30,so we use the pried t-test because we test time for the same sample in congruent words and after we change the colores to be incongruent words.[1]

[1] http://www.statisticshowto.com/when-to-use-a-t-score-vs-z-score

3. .Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

   By using Excel :

   measure of central tendency:

congruent words	in congruent words
Mean 14.051125	Mean 22.01592
Median 14.3565	Median 21.0175
Mode #N/A	Mode #N/A

   measure of variability:

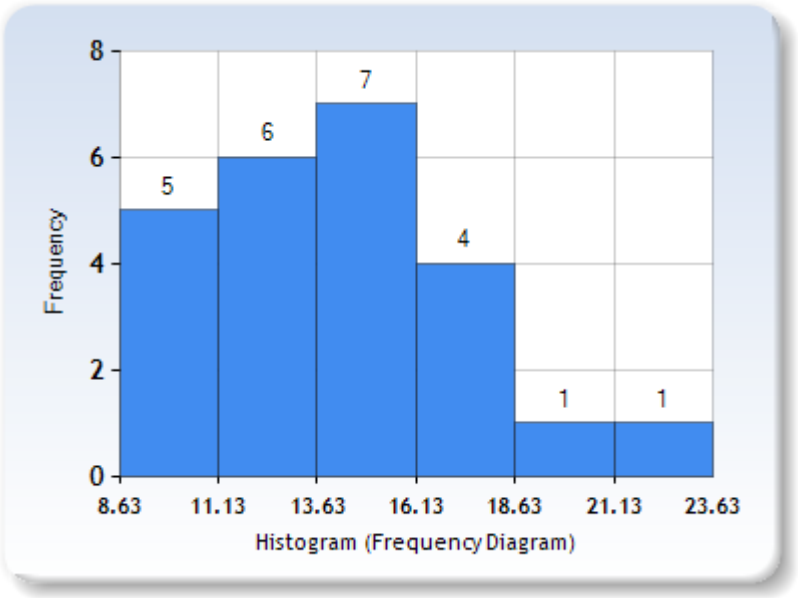
congruent words	in congruent words
Standard Deviation 3.559358	Standard Deviation 4.797057
Sample Variance 12.66903	Sample Variance 23.01176
Maximum 22.328	Maximum 35.255

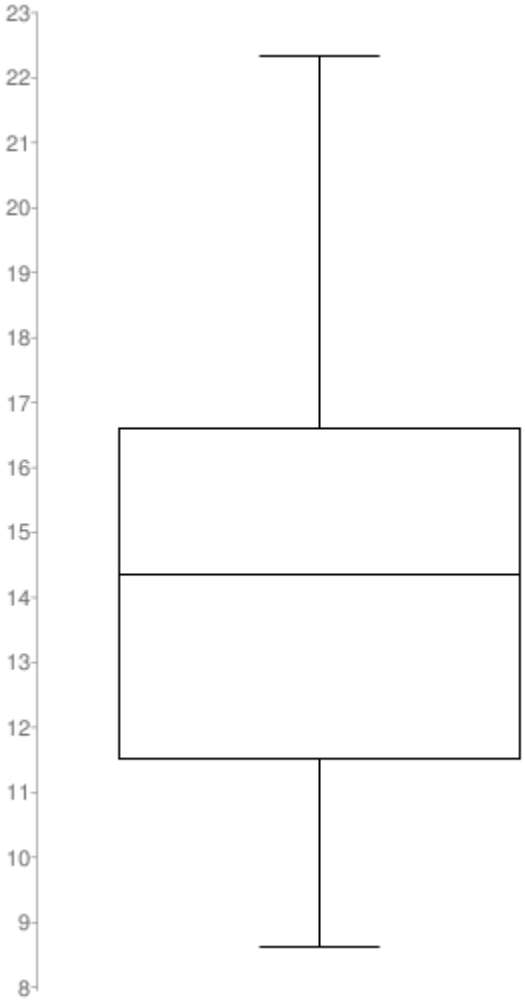
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

   For histograms I use :
   http://www.socscistatistics.com/descriptive/histograms/Default.aspx

   For Box plot I use:
   http://www.alcula.com/calculators/statistics/box-plot/

Congruent word visualization :

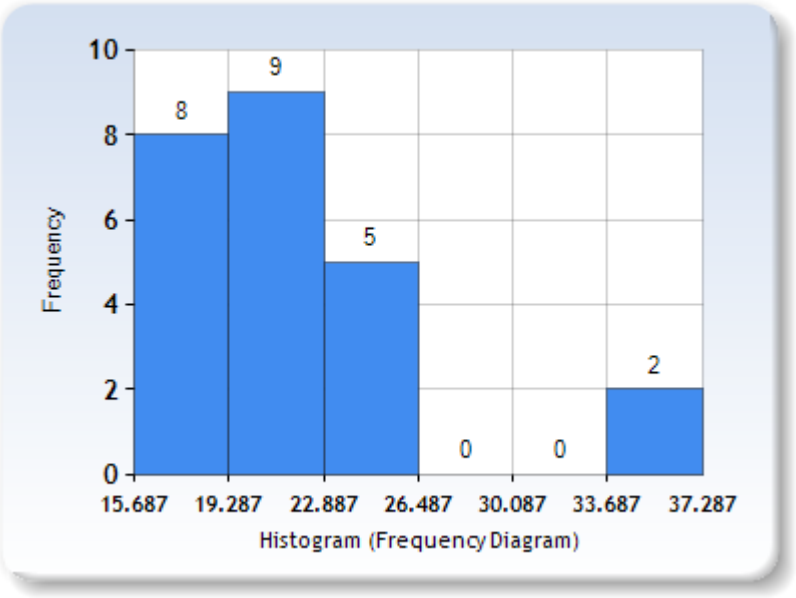


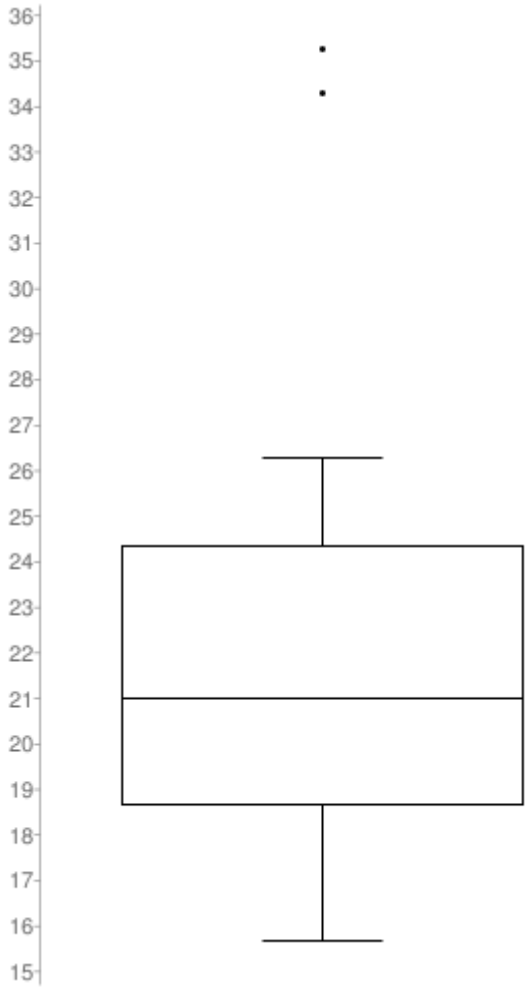


The histogram are asymmetrical it's positive skewed .

Form the Box plot we can see there is no outlier and the mean is little less than the median.

incongruent word visualization :





The histogram are asymmetrical it's positive skewed .

Form the Box plot we can see there outlier and the mean is larger than the median.

5. Now, perform the statistical test and report your results. What is your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?

   By Excel:

   Confidence interval = 95%

   An alpha level= 0.05

   t-Stat=-8.02070694410996

   P-value = 4.10300058571119E-08

   t-critical = 2.068658

   We found :

   t-Stat< t-critical and P-value< alpha

   so we reject the null hypothesis: The average time it takes to name the ink col colors ours are not same for congruent words and incongruent words.

   And this is what we expect because people read the word faster than Identify the colors.