Data Analyst Nanodegree
Project 1
Senay Goitom
29 August 2015

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

The independent variable is the time it takes to respond to the question of which color ink the word is written in. The dependent variable is whether the word is congruent or incongruent.

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

Suppose that response times for the incongruent and congruent lists (R_I and R_C, respectively) are random variables that are normally distributed as follows:

$$R_I \sim N(\mu_I, \sigma_I)$$

$$R_C \sim N(\mu_C, \sigma_C)$$

In other words, suppose that the underlying variances for each are not necessarily equal. One appropriate null hypothesis for this task is that the mean response time is equal for the two groups. In this case, the hypothesis would be written as follows:

$$H_0: \mu_I - \mu_C = 0$$

In this case, assuming that each list contains a sufficient number of observations, then a reasonable test would be a two-tailed test of equal means, with a p-value of 0.05.

Alternatively, one may reasonably suspect that the mean response time for the incongruent word list is greater than that of the congruent word list, since the color mismatch may require the respondent to think harder about the correct response. In this case, we would formulate the null hypothesis as:

$$H_0: \mu_I - \mu_C \leq 0$$

In this case, a single-tailed test of equal means, with a p-value of 0.05 would be appropriate.

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

Summary Stat	Congruent	Incongruent
Min.	8.63	15.69
1st Qu.	11.9	18.72
Median	14.36	21.02
Mean	14.05	22.02
3rd Qu.	16.2	24.05
Max.	22.33	35.26
Std. Dev.	3.56	4.80

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.

The three plots below summarize the distributions of completion times for the congruent and incongruent word lists. After presenting the distributions for each list separately, the final figure overlays the two plots to show that the distribution of the completion times for the incongruent list appears to be centered at a higher value, which is in line with the summary statistics presented above. The incongruent list also appears to be more dispersed than the congruent list.
.

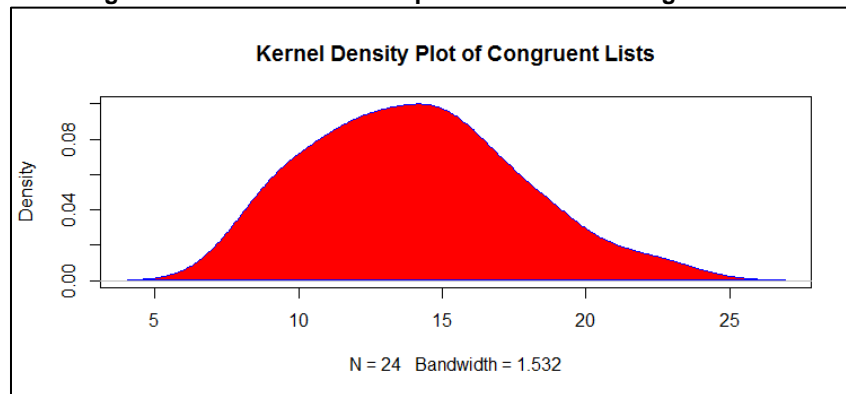**Figure 1: Distribution of Completion Times for Congruent List**

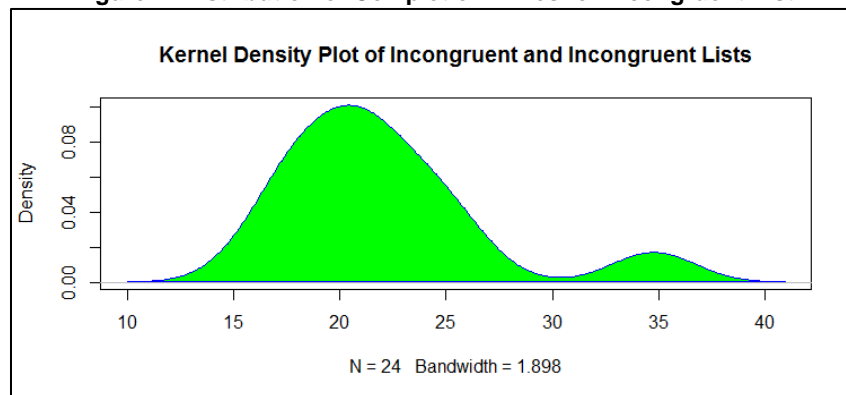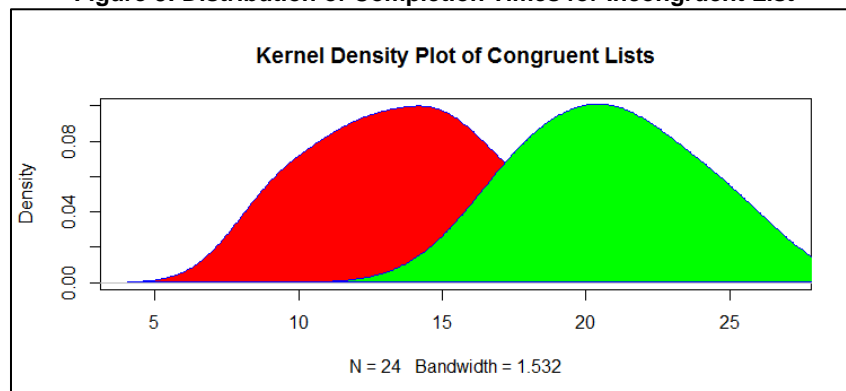**Figure 2: Distribution of Completion Times for Incongruent List**

Figure 3: Distribution of Completion Times for Incongruent List

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?**

Testing the second null hypothesis, that the response time for the congruent list is greater than or equal to the response time for the incongruent list, with a confidence level of 0.05, the critical value is roughly -1.71. The t-value is -6.53, and thus we can reject the null hypothesis that the mean completion time for the congruent list is at least as long as the mean completion time for incongruent list.

```
Welch Two Sample t-test

data:  congruent_list and incongruent_list
t = -6.5323, df = 42.434, p-value = 3.255e-08
alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:
      -Inf -5.914456
sample estimates:
mean of x mean of y
 14.05113  22.01592
```

Sources Used:
http://www.statmethods.net/graphs/density.html
http://www.statmethods.net/stats/ttest.html
http://scistatcalc.blogspot.com/2013/11/online-critical-t-value-calculator.html