Statistics: The Science of Decisions Project Instructions

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

   Independent variable: the type of conditions: a congruent words condition, and an incongruent words condition.

   Dependent variable: the time it takes to name the ink colors in equally-sized lists.

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

   The null hypothesis (H0): the time it takes to name the ink colors in the congruent words condition is not different from the incongruent words condition (i.e. time_congruent = time_incongruent).

   Given that theoretically it should take more time when the word and ink color is incongruent, the alternative hypothesis (HA) can be that the time used to name the ink color in the congruent words condition is less than the incongruent words condition (i.e. time_congruent < time_incongruent).

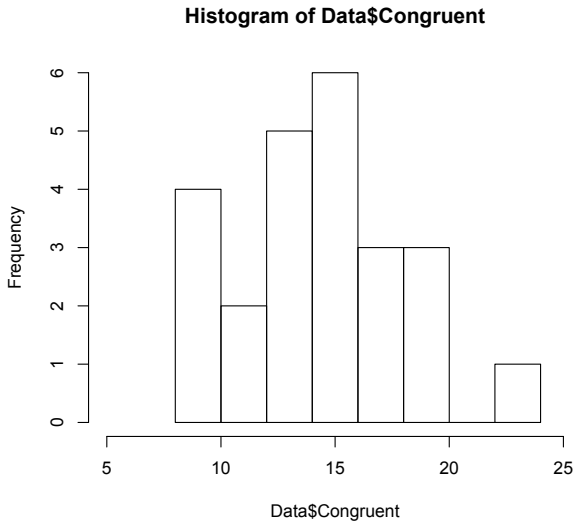
   The dependent t-test for paired sample I will expect to perform.

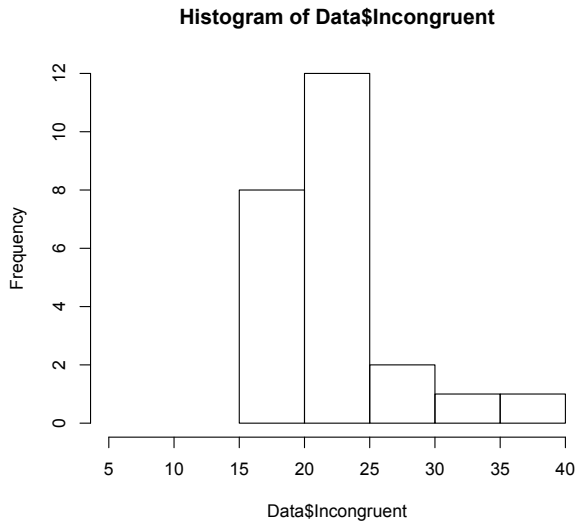
3. **Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.**

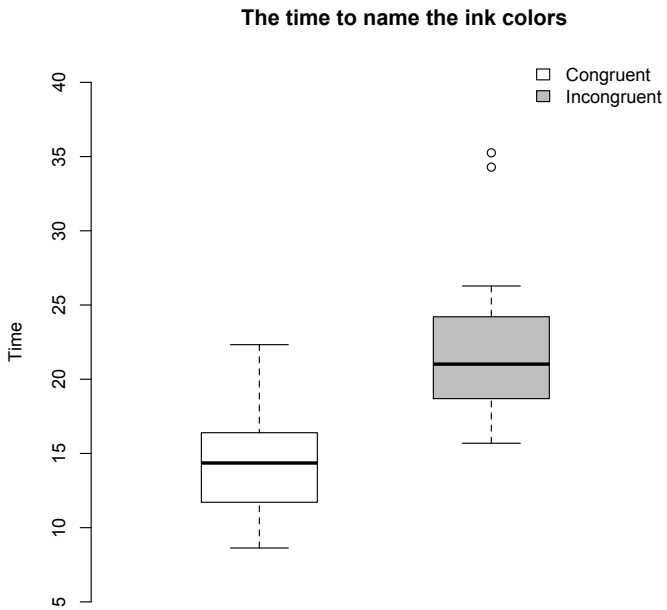
   The mean time_congruent is 14.05. The mean time_incongruent is 22.02.
   The median time_congruent is 14.36.The median time_incongruent is 21.02.
   The variance of time_congruent is 12.67. The variance of time_incongruent is 23.01.
   The sample standard deviation of time_congruent is 3.56.
   The sample standard deviation of time_incongruent is 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 distribution of the sample in congruent words condition is approximately normal, with the mode of 15, ranges from around 8 to 23.
   - The distribution of the sample in incongruent words condition is positively skew, with the mode between 20-25, ranges from around 15 to 40.
   - The box-plot indicates there are two outliers in the incongruent words condition.







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?

Paired t-test: at 95% confidence level, $t(23) = -8.02$, $p<.0001$, one-tailed.
Confidence interval on the mean difference; 95%CI = (-10.02, -5.91).
Effect size: $d = -1.64$; $r\text{-squared} = .74$
Because of $p<.0001$, we reject the null hypothesis. The result suggests that the time used to name the ink color in the congruent words condition is significantly less than in the incongruent words condition, which matches my expectation.

6. **Optional: What do you think is responsible for the effects observed? Can you think of an alternative or similar task that would result in a similar effect? Some research about the problem will be helpful for thinking about these two questions!**

   There are several theories to explain the stroop effects observed, including the theories related to processing speed, selective attention, automaticity, parallel distributed processing and cognitive development. Some similar tasks are modified versions of the original stroop task, including the Warped words task, the emotional stroop task, the spatial stroop task and the numerical and reverse stroop tasks.

   Reference source: https://en.wikipedia.org/wiki/Stroop_effect