Stroop Effect Exploratory Analysis

Javier Monterrubio

11/14/2015

Background information

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the color of the ink in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the congruent words condition, the words being displayed are color words whose names match the colors in which they are printed: for example RED (red ink), BLUE (blue ink). In the incongruent words condition, the words displayed are color words whose names do not match the colors in which they are printed: for example PURPLE (green ink), ORANGE (purple ink). In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

Questions For Investigation

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

The dependant variable is the words *condition* (congruent an incongruent in this case). The independent variable is the *time* it takes to name the ink colors of the words measured in seconds.

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

For each participant we are measuring two times. So in this case we are analyzing a set of paired data. "Two set of observations are *paired* if each observation in one set has a special correspondence or connection with exactly one observation in the other data set." And in particular we have a pre-test post-test set.

```
##   Congruent Incongruent
## 1    12.079      19.278
## 2    16.791      18.741
## 3     9.564      21.214
## 4     8.630      15.687
## 5    14.669      22.803
## 6    12.238      20.878
```

```
## N:  24
```

To analyze paired data, it is often useful to look at the difference in outcomes of each pair of observations. In this data set, we look at the difference in times. We are going to represent it as the $diff$ variable.

```
##   Congruent Incongruent   diff
## 1    12.079      19.278  7.199
## 2    16.791      18.741  1.950
## 3     9.564      21.214 11.650
## 4     8.630      15.687  7.057
## 5    14.669      22.803  8.134
## 6    12.238      20.878  8.640
```

To analize a paired data set, we simply analize the differences, using the $t-distribution$ techniques, this is, a t-test. In general, use the $t-distribution$ for inference of the sample mean when observations are independent and nearly normal. We can relax the nearly normal condition as the sample size increases. But the general rule of thumb for when to use a t score is when the sample size meets the following two requirements:

- The sample size is below 30.
- The population standard deviation is unknown (estimated from your sample data).

In advance we know that the stroop effect affects to the time of reaction for incongruents words increasing it. So we are considering two scenarios: there is no difference or there is some difference in average times (using one-tailed test).

$$H_0 : \mu_{diff} = 0$$

$$H_A : \mu_{diff} > 0$$

These formulas means that, in the first case (the null hypothesis H_0) we infer with our sample means $\bar{x}_{Incongruent}$ and $\bar{x}_{Congruent}$ that the difference of the population means are zero. This also could be seen as $H_0 : \mu_{Incongruent} = \mu_{Congruent}$. The second formula (the alternative hypothesis H_A) says that we are infering with our sample means that the difference of the population means are greather than 0. Another representation could be $H_0 : \mu_{Incongruent} > \mu_{Congruent}$.

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

A summary of this difference is:

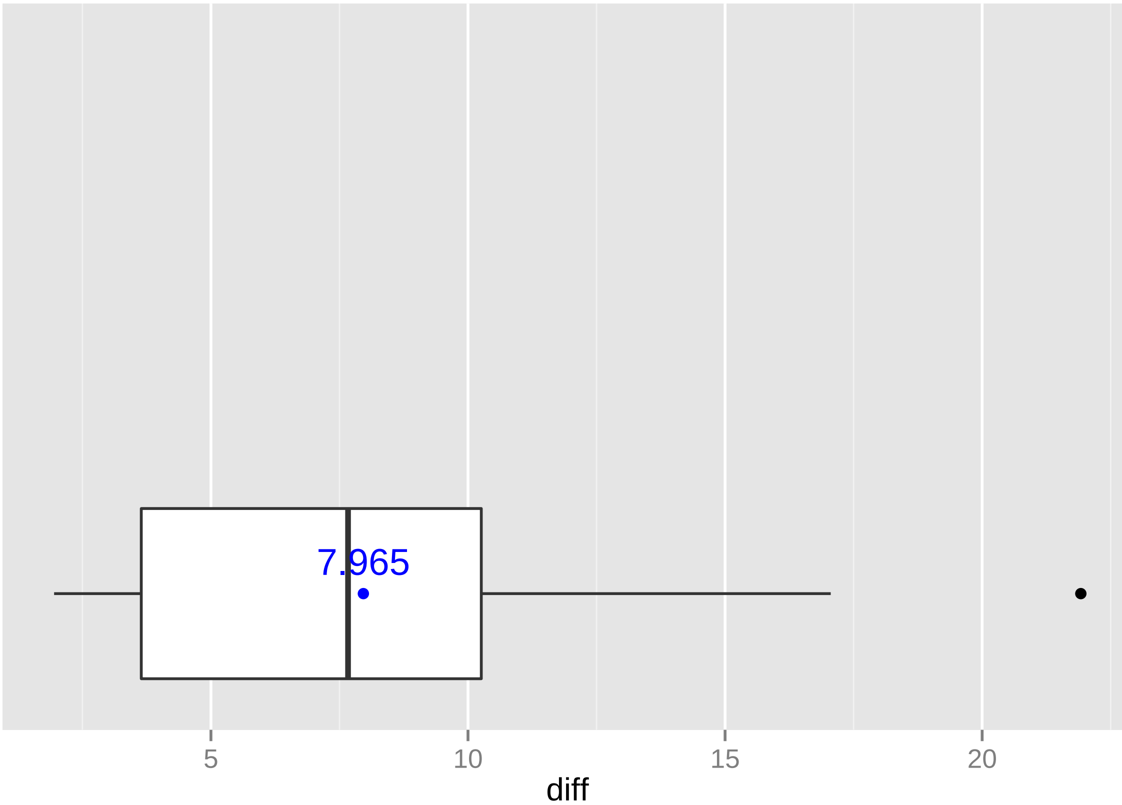
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.950   3.646   7.666   7.965  10.260  21.920
```

```
## Variance:  23.66654
```

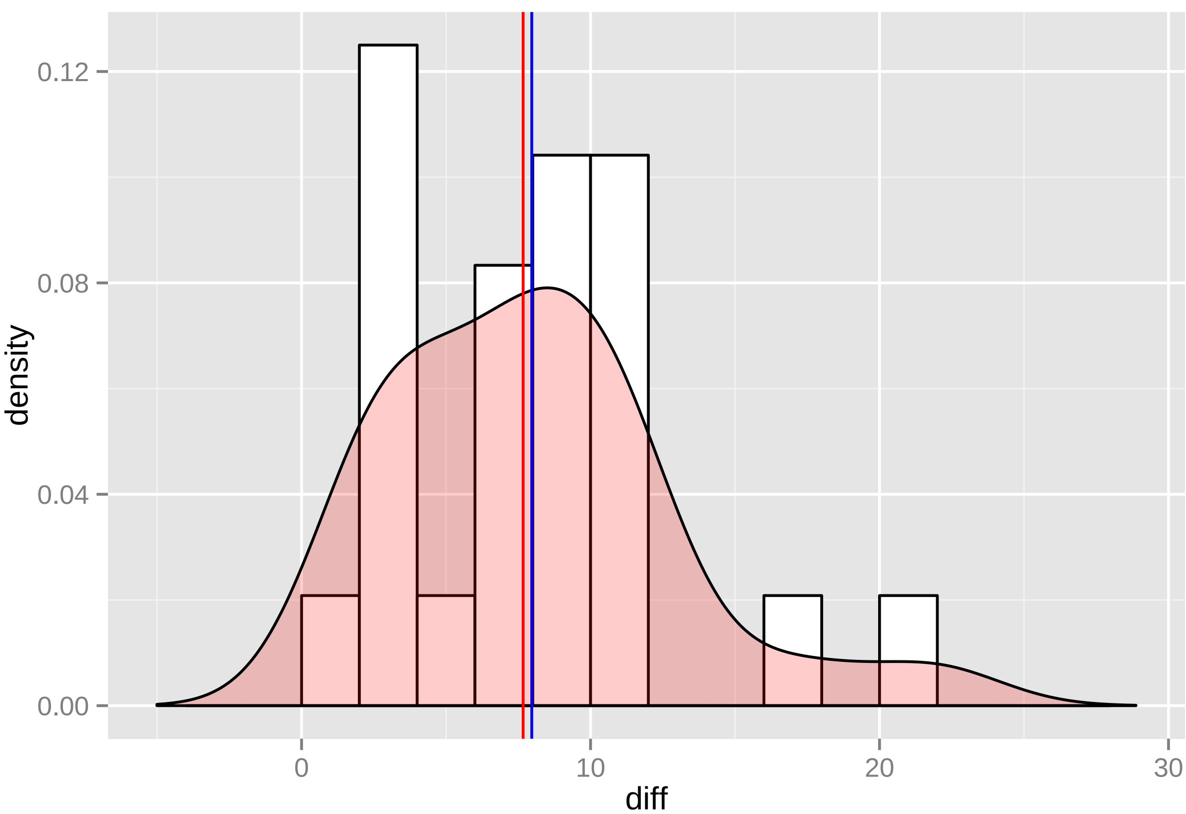
```
## Standard deviation:  4.864827
```

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.

According to the previous measures the boxplot looks like:



The interquartile range (IQR) are defined by the 1st and 3rd quartiles (3.646 and 10.26) where resides the 50% of the data. The median (7.666) is smaller than the mean (7.965)



We can see that the data is a bit right skewed (mean in blue > median in red) but the sample is enough big (n = 24) so we can proceed using $t - distribution$ for this application.

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?

We compute the standard error associated with $\bar{x}_{diff}(7.9648)$ using the standard deviation of the differences ($s_{diff} = 4.8648$) and the number of differences ($n_{diff} = 24$):

$$SE_{\bar{x}_{diff}} = \frac{s_{diff}}{\sqrt{n_{diff}}} = \frac{4.8648}{\sqrt{24}} = 0.993$$

Now, we compute the test statistic to see how far away is from the $t - critical$ with an alpha level of 0.05 ($\alpha = 0.05$). So we compute the T-score of $\bar{x}_{diff}$ under the null condition that the actual mean difference is 0:

$$T = \frac{\bar{x}_{diff} - 0}{SE_{\bar{x}_{diff}}} = \frac{7.9648}{0.993} = 8.0209$$

$$t - critical_{df=23} = 1.714$$

The difference between both values are very big and using a statistical sofware the p-value obtained for our T-score is very low (less than 0.0001). By conventional criteria, this difference is considered to be extremely statistically significant. Because the p-value is less than 0.05, we reject the null hypothesis. We have found convincing evidence that it takes more time reading incongruent words, on average, than congruent ones.

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!

As we saw befere, the difference in reading times is statistically significant and is explained by the automation of reading, where the mind automatically determines the semantic meaning of the word but not the color itself.

Another similar task is the $FlankerorEriksenTask$. It is a choice reaction time where there is dimensional overlap between the irrelevant stimulus and the relevant stimulus. In a typical Eriksen task, subjects are shown a string of letters on a screen, and are told to press a left key or a right key depending on what letter appears in the center of the screen (the target letter). The surrounding flanker letters are irrelevant, but can be either consistent ('HHH') or inconsistent ('SHS') with the target. Responses are faster and more accurate for consistent stimuli than for inconsistent stimuli.

Appendix A

Sources

The list of sources used to complete this exploratory analysis are:

- The stroop data set provided by the course.
- Video lectures of the course.
- OpenIntro Statistics book (https://www.openintro.org/stat/?stat_book=os).
- GraphPad online calculator (http://www.graphpad.com/quickcalcs/) to compute the p-value.

- Wikipedia to look for more information about the Stroop effect (https://en.wikipedia.org/wiki/Stroop_effect).
- Flanker task research (http://dimensional-overlap.com/the-flanker-task/).
- ggplot2 cookbook for visualizations in R (http://www.cookbook-r.com/Graphs/Plotting_distributions_(ggplot2)/).
- RStudio with Knit to create the final PDF file.