Test a Perceptual Phenomenon

Jin Lim

Udacity Data Analyst Nanodegree

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

   - Independent: the congruent words condition.

   - Dependent: time it takes for participant to name the ink colors in each word set.

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

   a) Hypotheses

   - The Stroop effect states that when the ink color and word's meaning are incongruent, it takes us longer to name the ink color and we are more prone to errors. Thus, the null hypothesis should be there is no difference in time taken between congruent and incongruent conditions.

   - Let μC = the population's mean time to identify ink colors under congruent conditions. Let μI = the population's mean time to identify ink colors under incongruent conditions.

   - $H_0: \mu C - \mu I = 0$

   - $H_1: \mu C - \mu I \neq 0$

   b) What statistical test?

   - A Z Test seems like a good candidate, but our population $n = 24 < 30$ and we do not know the standard deviations and variances.

- Thus we use a T Test. We should use a two-tailed test as our hypotheses state that there is/isn't a difference between population mean times and does not make an explicit judgment between less or more time. We cannot use an unpaired or independent T Test because the participants are not divided between "congruent" and "incongruent" conditions.
- **Two-tailed Paired-Samples T-test** should be used.

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

```
In [3]:  import pandas as pd
         import matplotlib.pyplot as plt
         import numpy as np
         import math
```

```
In [6]:  path = "C:\\Users\\--\\Desktop\\stroopdata.csv"
         stroop = pd.read_csv(path)
         stroop.describe()
```

Out[6]:

	Congruent	Incongruent
count	24.000000	24.000000
mean	14.051125	22.015917
std	3.559358	4.797057
min	8.630000	15.687000
25%	11.895250	18.716750
50%	14.356500	21.017500
75%	16.200750	24.051500
max	22.328000	35.255000

- Central Tendency. Congruent mean: 14.05. Incongruent mean: 22.02.
- Variability. Congruent sd: 3.56. Incongruent sd: 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.

```
In [41]: stroop.plot(kind = 'box',
                vert = False,
                color = dict(boxes = 'RoyalBlue',
                             whiskers = 'Crimson',
                             medians = 'Black',
                             caps = 'Black')
                ).set_xlabel('Response Time (in seconds)')
         plt.show()
```
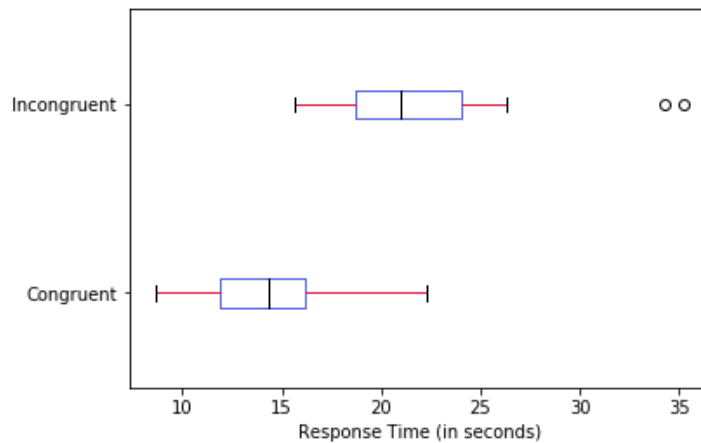

- You can visually tell that response times under incongruent conditions are much longer than under congruent conditions. Further testing is required to determine if this is statistically significant.

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?

- df = 23

- $\alpha = 0.05$ at a 95% confidence interval

- $t_{critical} = \pm 2.0687$

- $\mu C - \mu I = 14.051 - 22.016 = -7.965$

- S = [In Excel] STDEV.S(Congruent – Incongruent Columns) = 4.865

- $t = (\mu C - \mu I)/(S/\sqrt{n}) = -8.021 > \pm 2.0687$.

- **Thus, we reject the null hypothesis that there is no difference in population mean times between congruent and incongruent conditions.**

- When confronted with incongruent conditions, participants took significantly longer to complete the task. These results match with my expectations, as you would expect the cognitive challenge of overcoming a habitual response to a word to slow down times.