Test a Perceptual Phenomenon

January 8, 2019

0.0.1 Analyzing the Stroop Effect

(1) What is the independent variable? What is the dependent variable?

The independent variable is a list of words, with each word displayed in a color of ink. The dependent variable is the time it takes to name the ink colors.

(2) What is an appropriate set of hypotheses for this task? Specify your null and alternative hypotheses, and clearly define any notation used. Justify your choices.

The set of hypotheses for this task in the mathematical statement is:

H0: 1 = 2
H1: 1 2,

1 - the time it takes to name the ink colors in a list with congruent words condition, 2 - the time it takes to name the ink colors in a list with incongruent words condition.

Another word: Null hypothesis(H0): The time it takes to name the ink colors in a list with congruent words condition and in a list with incongruent words condition is equal. Alternative hypothesis(H1): The time it takes to name the ink colors in a list with congruent words condition and in a list with incongruent words condition is not equal.

A statistical test.

I selected the paired sample t-test for my hypothesis because: we can compare the means of two groups to determine the statistically significant difference between two means, we assume that data is normally distributed, and the sample size is small - 24.

(3) Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability. The name of the data file is 'stroop-data.csv'.

```
In [29]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import scipy.stats as stats
         %matplotlib inline

         df = pd.read_csv('stroopdata.csv')
         df.head()
```

```
Out[29]:    Congruent  Incongruent
          0     12.079       19.278
          1     16.791       18.741
          2      9.564       21.214
          3      8.630       15.687
          4     14.669       22.803

In [30]: df.describe()

Out[30]:         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

In [31]: # calculate Means
          df.mean()

Out[31]: Congruent      14.051125
          Incongruent    22.015917
          dtype: float64
```

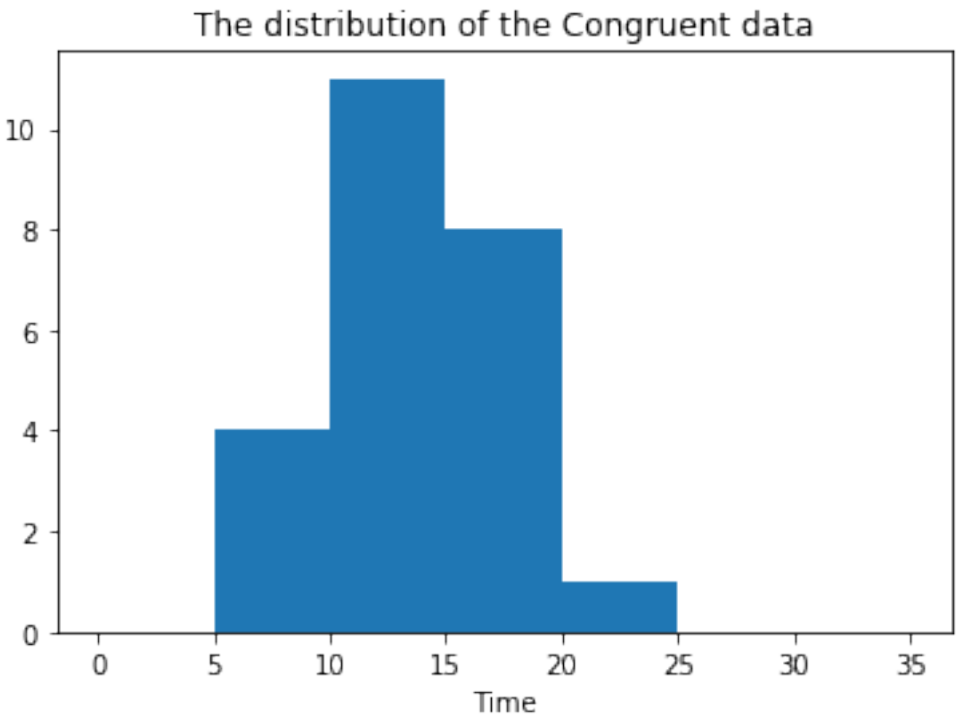
The average time for Congruent word is 14.05. The average time for Incongruent is 22.02. Therefore, the average time in the incongruent list is longer than the average congruent list.

```
In [34]: # calculate Standard deviations
          df.std()

Out[34]: Congruent      3.559358
          Incongruent    4.797057
          dtype: float64
```

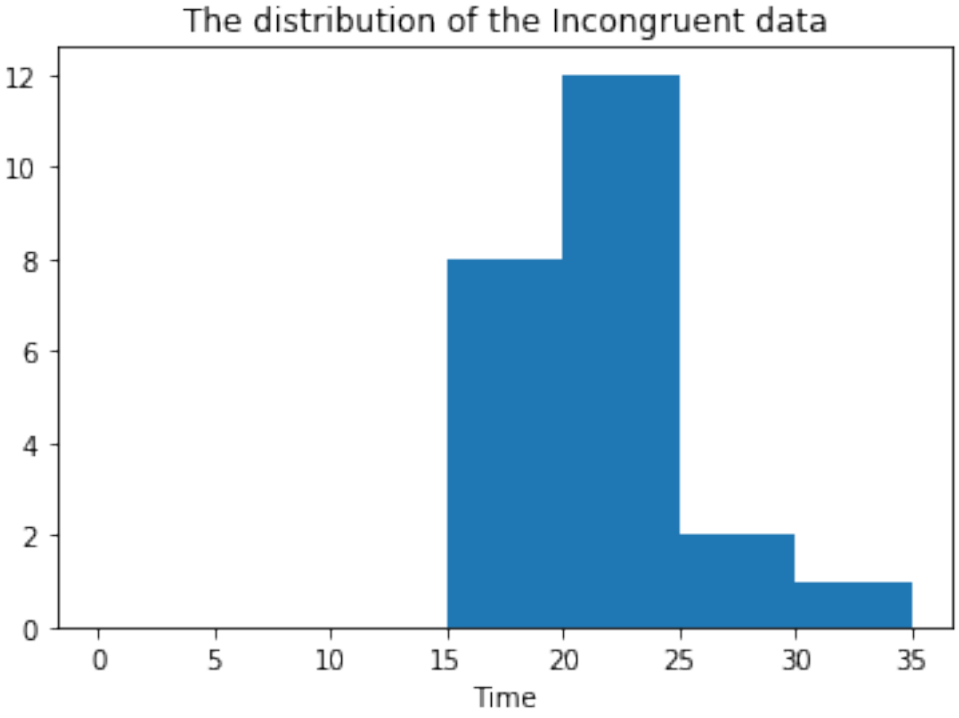
The standard deviation for Congruent data is 3.56, the standard deviation for Incongruent data is 4.80. Therefore, the spread of Incongruent numbers is more than Congruent numbers.

(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 [35]: # Build the visualizations
          bins = [0,5,10,15,20,25,30,35]

          plt.hist(df.Congruent, bins)
          plt.title('The distribution of the Congruent data')
          plt.xlabel('Time');
```



In [36]:
```
plt.hist(df.Incongruent, bins)
plt.title('The distribution of the Incongruent data')
plt.xlabel('Time');
```



As we can see on the plot, the value incongruent data is generally higher than the congruent data.

(5) Now, perform the statistical test and report your results. What is your confidence level or Type I error associated with your test? What is your conclusion regarding the hypotheses you set up? Did the results match up with your expectations? **Hint:** Think about what is being measured on each individual, and what statistic best captures how an individual reacts in each environment.

We will use paired sample t-test.

```
In [26]: # Perform the statistical test here
         result = stats.ttest_rel(df['Incongruent'], df['Congruent'], axis=0)
         (t_stat, p_value) = (result)

In [28]: print('t-statistic:', t_stat)
         print('p-value:', p_value)

t-statistic: 8.02070694411
p-value: 4.10300058571e-08
```

For the analysis, we use the Significance level of 0.05 (with 23 degrees of freedom).

We calculated, t-stat for test data = 8.02, and P-value for t-stat = 4.10e-08. As we can see, p-value 4.10e-08 is smaller than the Significance level 0.05. Therefore, we reject the null hypothesis test that the time it takes to name the ink colors in a list with congruent words condition and in a list with incongruent words condition is equal. The result match up with my expectation.