Test a Perceptual Phenomenon

January 12, 2019

0.0.1 Analyzing the Stroop Effect

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

The independet variable is the variable that we contol to see its effect on the dependent variable, meaning that the value of the dependent variable changes according to the change applied to the independent variable. In the stroop test the independent variable is the word condition and the dependent variable is the time taken to recognize the color.

(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.

- H0: *congruent = incongruent*

- H1: *congruent != incongruent*

- Null Hypotheses: the avarage time taken to say out loud congruent words is the same as the avarage time taken to say out loud incongruent words.

- Alternative Hypotheses: the avarage time taken to say out loud congruent words diffes from the avarage time taken to say out loud incongruent words.

since the population size is less than 30 it's prreferable to use t-test.

(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'.

```python
In [15]: # Perform the analysis here
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         from scipy import stats

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

```
Out[15]:        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
```

- The mean time to read congruent words is: 14.051125 and for incongruent words is: 22.015917.

- The standard deviation of congruent words is: 3.559358 and for incongruent words is: 4.797057

(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 [10]: *# Build the visualizations here*

```python
plt.hist(df.Congruent, alpha=0.5, label='Congruent')
plt.hist(df.Incongruent, alpha=0.5, label='Incongruent')
plt.xlabel("Time to read")
plt.ylabel("repition")
plt.legend(loc='upper right')
plt.show()
```
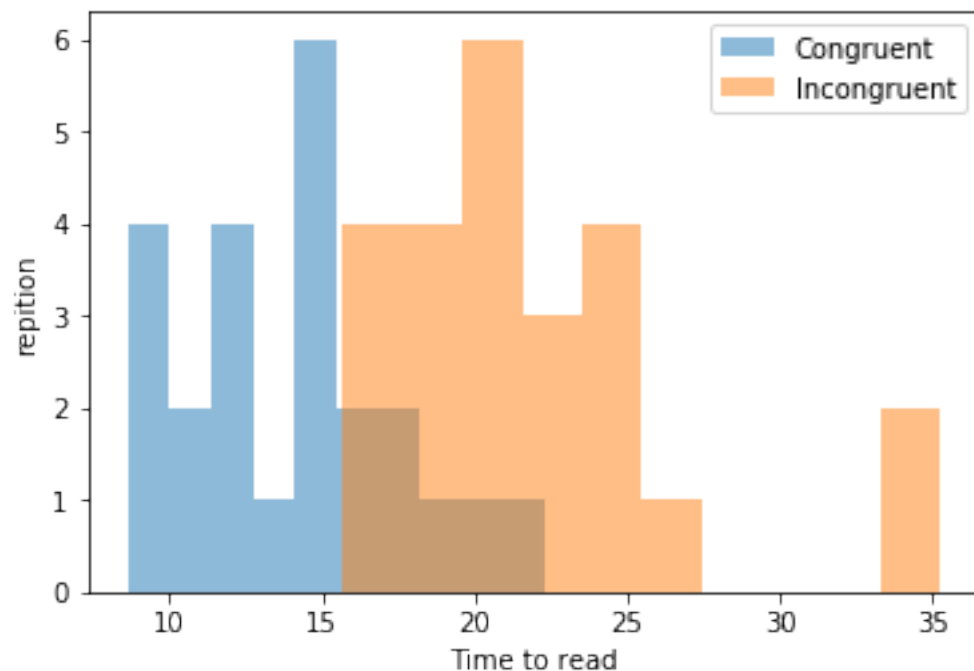

from the histogram above we see that time taken to read congruent words starts approximately from 5 and reaches 22, 15 sec is repeated 6 times across the dataset. Moreover, the time taken to

2

read incongruent words starts approximately from 15.5 and reaches 35, 21 sec is repeated 6 times across the dataset. The histogram also show that the it's most likely to read congruent words in less time than to read incongruent words.

(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.

notes: - Type 1 error (false positive): we say that the times differ but they are the same. - Type 2 error (false negative): we say that the times are the same but they differ.

```
In [18]: df.count()
```

```
Out[18]: Congruent      24
         Incongruent    24
         dtype: int64
```

in order to decide wether to accept or reject the null hypothses we have to compare the critical t-value with the calculated t statistic

Since the sample size is 24, the degree of freedom = 23 (n-1), with 95% confidence interval, we can calculat the critical t-value:

```
In [28]: stats.t.ppf(.95, 23)
```

```
Out[28]: 1.7138715277470473
```

calculating t-statistic:

```
In [42]: N = 24 #size of the dataset
         mean_diff = df.Incongruent.mean() - df.Congruent.mean()
         diff = df.Incongruent - df.Congruent
         t = (mean_diff)/(diff.std()/np.sqrt(N))
         t
```

```
Out[42]: 8.020706944109957
```

Since the t-statistic is greater than the critical t-value, the test proves that there is a statistically significant difference between the mean time requirred to read congruent words and to read incongruent words. Therefore, we reject the null hypothesis.

0.1 references:

- https://stackoverflow.com/questions/6871201/plot-two-histograms-at-the-same-time-with-matplotlib
- http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-hypothesis-tests-confidence-intervals-and-confidence-levels
- https://stackoverflow.com/questions/19339305/python-function-to-get-the-t-statistic
- https://towardsdatascience.com/inferential-statistics-series-t-test-using-numpy-2718f8f9bf2f