Test a Perceptual Phenomenon

June 14, 2018

0.0.1 Analyze the Stroop Effect

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

The independent variable is the condition, defined as whether the words displayed are color words whose names match the colors as printed. In this case, there are two conditions, congruent and incongruent. The dependent variable is the response time to name the colors in equally-sized lists.

(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.

$$H_0 : \mu_{Congruent} \geq \mu_{Incongruent}$$

$$H_1 : \mu_{Congruent} < \mu_{Incongruent}$$

My null hypothesis states that the population mean response time to congruent words condition, notated as $\mu_{Congruent}$, is greater than or equal to the population mean response time to incongruent words condition, notated as $\mu_{Incongruent}$. My alternative hypothese asserts that the former is less than the latter.

It is important to note the assumption that we are making a statistical inference on the population from which samples are drawn.

In this case, it is appropriate to conduct a one-tailed dependent t-test for paired samples becuase our goal is to determine whether it takes longer to respond to incongruent words condition than congruent words condition. Because the sample size of 24 is smaller than 30, I've chosen the t-test instead of the z-test.

As a result, I need to make a few assumptions. First, I must assume that the samples are drawn a population with a normal distribution in order to apply the t-test. Second, becuase the population's standard deviation is unknown, the t-test is preferred due to the limited sample size and the prior assumption. Last, the assumption that sampling is random and pairs of observations are independent is necessary.

I've decided to use the paired t-test rather than the unpaired t-test since paired samples are collected on the basis of behavior expected to affect the dependent variable as defined in the prevoius question.

(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 [1]: import pandas as pd
        import numpy as np
        from scipy import stats
```

```
import matplotlib.pyplot as plt
%matplotlib inline

# Descriptive statistics section
df = pd.read_csv('stroopdata.csv')
df['diff'] = df['Congruent'] - df['Incongruent']
df.describe()
```

```
Out[1]:        Congruent  Incongruent       diff
       count  24.000000    24.000000  24.000000
       mean   14.051125    22.015917  -7.964792
       std     3.559358     4.797057   4.864827
       min     8.630000    15.687000 -21.919000
       25%    11.895250    18.716750 -10.258500
       50%    14.356500    21.017500  -7.666500
       75%    16.200750    24.051500  -3.645500
       max    22.328000    35.255000  -1.950000
```

Measure of Central Tendency: Mean of Response Times

$$\mu_{Congruent} = 14.0511$$

$$\mu_{Incongruent} = 22.0159$$
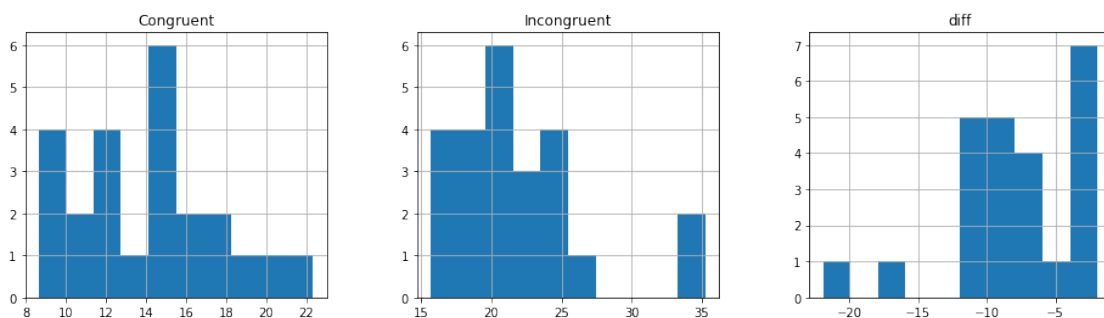
Measure of Variability: Standard Deviation of Response Times

$$\sigma_{Congruent} = 3.5594$$

$$\sigma_{Incongruent} = 4.7971$$

(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 [2]: # Histograms for 3 sets of data: congruent, incongruent, and difference
        df.hist(layout=(1, 3), figsize=(16, 4))
```
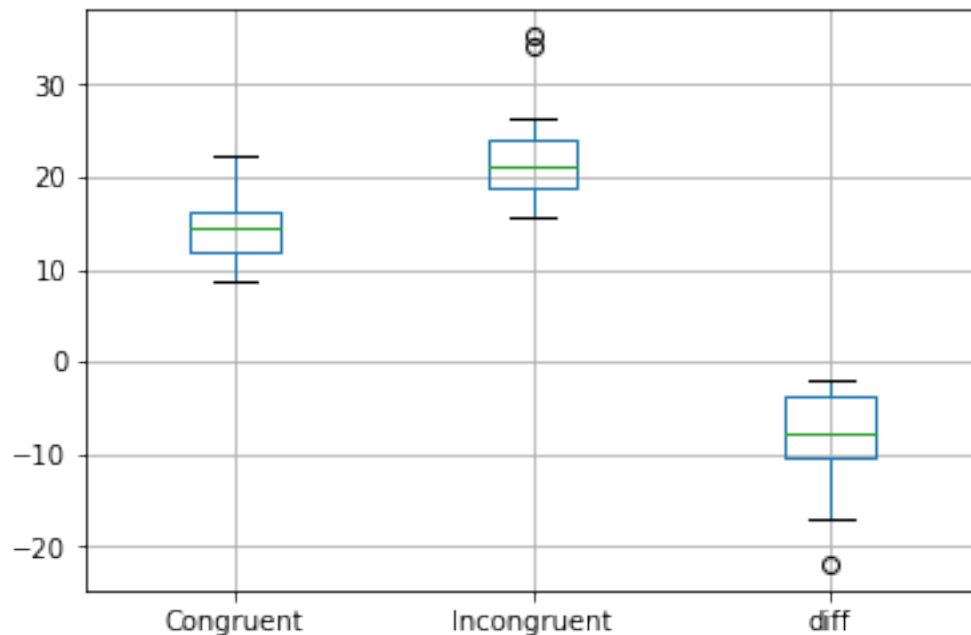
```
Out[2]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f43291c9ef0>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x7f43270f0240>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x7f43270ab780>]], dtype=object
```

The first two histograms plots show some skewness. The last histogram displays varying degrees of difference between two samples.

```
In [3]: df.boxplot()
```

```
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x7f4326e6b2b0>
```


The boxplots are more revealing in that the difference between congruent and incongruent data is clear. Even disregarding the outliers of incongruent data, I can visually establish that it takes longer to respond to incongruent words condition than to congruent words condition. This will serve as my prediction before running the statistical test.

(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.

```
In [4]: # statistical test using SciPy
        congruent = df['Congruent']
        incongruent = df['Incongruent']
        stats.ttest_rel(congruent, incongruent)
```

```
Out[4]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

3

My confidence level, or Type I error, is set at 5%, which implies a critical value of -1.7139 with a degree of freedom of 23.

My one-tailed dependent t-test for paired samples returns a test statistic of -8.0207, which is to the left of the critical value, and a p-value much smaller than the alpha. Both results point to the same conclusion that I can reject the null hypothesis. In other words, the response time to congruent words condition is shorter than that to incongruent words condition.

The results matched my expectation that incongruent words condition causes delays in brain activities since the brain needs to resolve visual conflicts. According to the data, the average delay is about 8 seconds.

(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!

I am no expert in neuroscience. After some research about this topic, I believe the Stroop effect has much to do with how the brain resolves cognitive conflicts and slows down its decision-making process given higher probability of errors.

I can think of a similar task related to the numerical Stroop effect demonstrating the relationship between numerical values and physical sizes. Let's consider scenarios in which numbers are printed on paper of large or small sizes regardless of their actual numerical values. Congruent pairs appear when sizes and values match (large/small values correspond to large/small sizes) whereas discrepancies between sizes and values results in incongruent pairs (large values correspond to small sizes or small values correspond to large sizes).

```
In [5]: from subprocess import call
        call(['python', '-m', 'nbconvert', 'Test a Perceptual Phenomenon.ipynb'])

Out[5]: 0
```

0.0.2 References

1. Wikipedia contributors. "Stroop Effect." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 16 May. 2018. Web. 14 Jun. 2018.
2. Wikipedia contributors. "Numerical Stroop Effect." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 1 May. 2018. Web. 14 Jun. 2018.