Test a Perceptual Phenomenon

February 20, 2018

0.0.1 Analyzing the Stroop Effect

Perform the analysis in the space below. Remember to follow the instructions and review the project rubric before submitting. Once you've completed the analysis and write up, download this file as a PDF or HTML file and submit in the next section.

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

independent variable: Type of task , that is Incongruent or Congruent
dependent variable: Time Diffrence between completing Congruent task and Incongruent task.

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

From the given data set, n = sample size = 24 . And we don't know the the population parameters,such as Standard Deviation or Sigma,which would be required to compute Standard Error. So in the absence of sigma, we need to use Sample Standard Deviation(S.S.D.) to compute Standard Error(S.E). that is, S.E. = S.S.D./sqrt(n) .

But,sample size(n) is low(less than 30) here. Therefore distribution will be a t-distribution rather than a normal-distribution. We know,as the sample size(n) increases, t-distribution -> normal-distribution

So, we have to choose t-test over z-test.

From the Experiment,it appears it would be a "Dependent 2-tailed t-test for paired Samples". Samples are dependent ,because same subject took the test twice. Its paired sample, because each participant took both tests for two different conditions. 2-tailed,because we want to see if there is a positive or a negative difference between completing two tasks.

Lets say, Population Mean(Congruent) = Mu(C) ,Population Mean(Incongruent) = Mu(I) ;

We want to see, if there is a significant amount of time difference between completing two tasks.

So, Null Hypothese would say,there is no significant difference between two Population means.In oter words,on average people(population) would take same amount of time to complete Incongruent task and Congruent task. And Alternative Hypotheses would say, significant difference between two Population means is present there.

Lets say, H0 is Null Hypotheses & H1 is Alternative Hypotheses. And also say, Mu(D) = Mu(C)-Mu(I)

Then H0 : Mu(D) = 0 And H1 : Mu(D) != 0

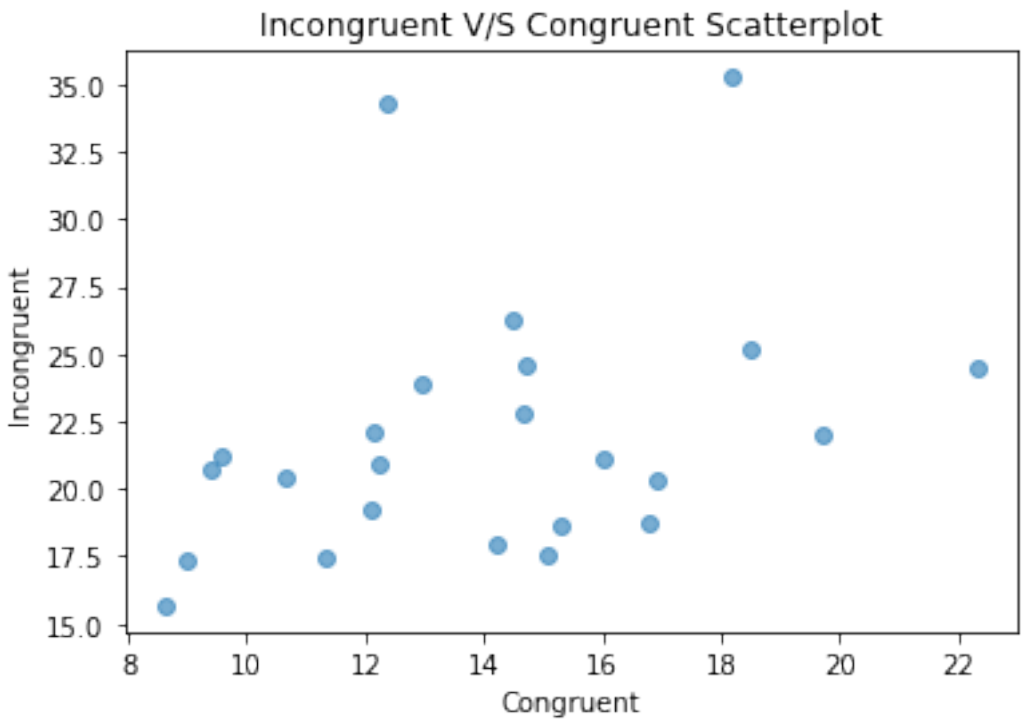
(3) Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability. The name of the data file is 'stroop-data.csv'.

```
In [4]: import pandas as pd
        import numpy as np
        data = pd.read_csv('stroopdata.csv')
        congruent = data['Congruent'].values.tolist()
        incongruent = data['Incongruent'].values.tolist()
        difference = np.subtract(congruent,incongruent)
        #'''arr = np.array([congruent,incongruent])'''

In [3]: # Mean of time difference between congruent task & incongruent task
        np.mean(difference)

Out[3]: -7.9647916666666658

In [4]: # Median of time difference between congruent task & incongruent task
        np.median(difference)

Out[4]: -7.6664999999999992

In [6]: #sample standard deviation(S.S.D.) of time difference between congruent task & incongrue
        import math
        math.sqrt(np.sum((difference-np.mean(difference))**2/(len(difference)-1)))
        #or
        np.sum((difference-np.mean(difference))**2/(len(difference)-1))**0.5

Out[6]: 4.8648269103590547
```

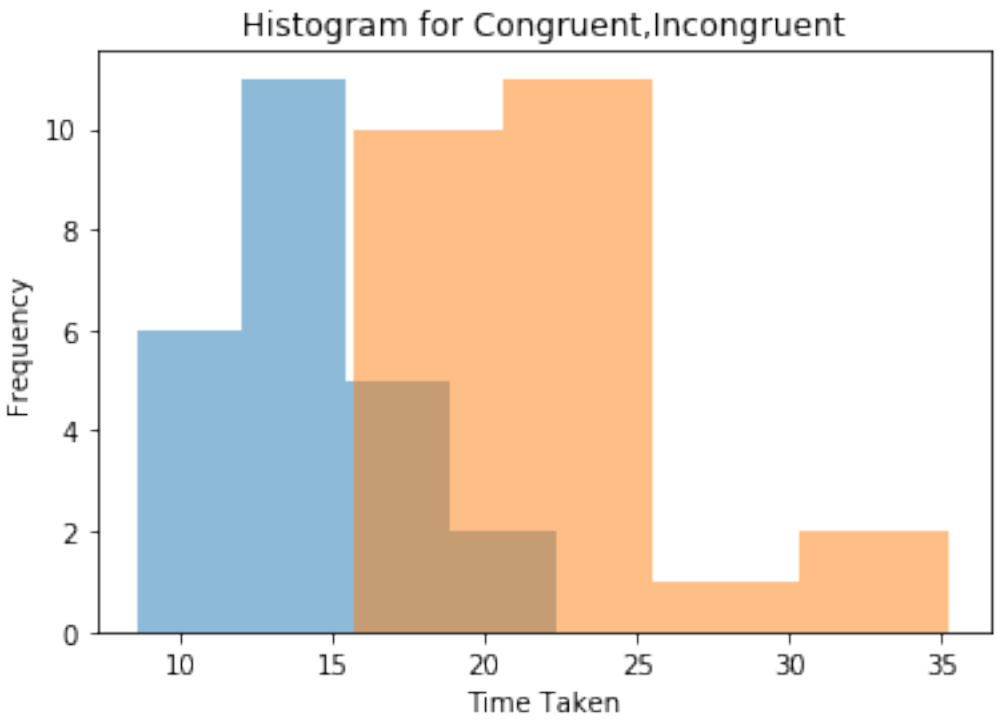
If, difference = congruent - incongruent = time difference between congruent task & incongruent task. Then, Sample Mean(difference)=-7.964791667 Sample Median(difference)=-7.6664999999999992 Stdev(difference)=Sample Standard Deviation(difference)=4.86482691

(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.

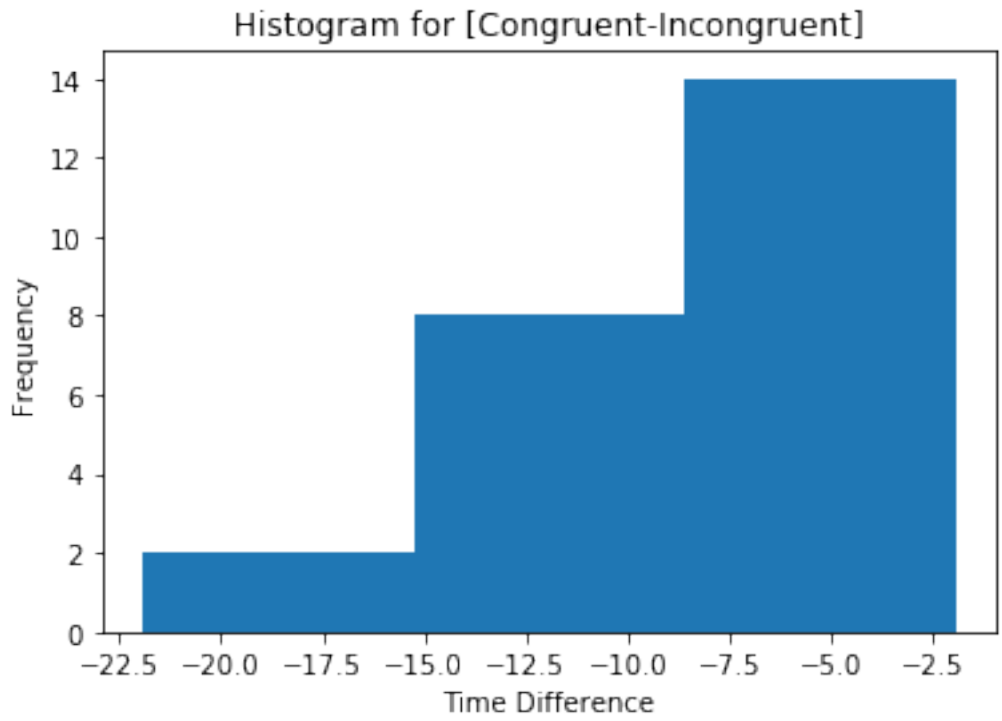
```
In [6]: import matplotlib.pyplot as plt
        import numpy as np
        #area = np.pi * (15 * np.random.rand(N))**2  # 0 to 15 point radii
        #plt.scatter(congruent, incongruent, s=area, c="RGB", alpha=0.4)
        plt.scatter(congruent, incongruent, alpha=0.6)
        '''y_pos = np.arange(len(congruent))
        plt.bar(y_pos, incongruent, alpha=0.5)
        plt.xticks([0,np.max(congruent)],congruent)'''
        plt.ylabel('Incongruent')
        plt.xlabel('Congruent')
        plt.title('Incongruent V/S Congruent Scatterplot')
        plt.show()
```



In [7]:
```
import numpy as np
#bins = np.linspace(0, 50,10)
plt.hist(congruent, bins='scott',range=(np.min(congruent),np.max(congruent)), alpha=0.5)
plt.hist(incongruent, bins='scott',range=(np.min(incongruent),np.max(incongruent)), alph
plt.title("Histogram for Congruent,Incongruent")
plt.xlabel('Time Taken')
plt.ylabel('Frequency')
plt.show()
```




```
In [12]: import numpy as np
         plt.hist(difference,3)
         plt.title("Histogram for [Congruent-Incongruent]")
         plt.xlabel('Time Difference')
         plt.ylabel('Frequency')
         plt.show()
```



From the "Incongruent V/S Congruent Scatterplot", we can say, even if two persons take the same amount of time to complete the congruent task , that does not mean they would take the same amount of time to complete the incongruent task.

From the "Histogram for Congruent,Incongruent",we can see, Participants had taken a greater amount of time to complete Incongruent task.

(5) Now, perform the statistical test and report the results. What is the 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?

```
In [56]: import scipy.stats as stats
         #stats.ttest_1samp(difference,0.0)
         stats.ttest_rel(congruent,incongruent)
```

```
Out[56]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

Assuming, Confidence Interval = CI = 95% . So, Aplha Level = a = .05 ;

Sample size = n = 24, Degree of Freedom = DF = n-1 = 23, Since test was 2-tailed, Test critical value = Tcritical = +2.069 or -2.069

Now we found, Test statistic = Tstatistic = -8.020706944

Now we can see, -8.020706944 < -2.069 ,implies Tstatistic < Tcritical . Therefore, p-value < .05 p-value=4.1030005857111781e-08

So Null Hypothesis is rejected.

Therefore, there is extremely statistically significant difference between two Population means. This means participants took significantly different amount of time to complete the incongruent task & the congruent task.

Since this is a Experimental design ,we can make causal statements.So we can say that type of the task had a causal effect on completion time of the task.

And I had expected the same result, since all participants from the sample had taken a greater amount of time to complete the incongruent task.

(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!

Literate people habituated & comfortable with reading.They are feeling more easier with reading than recognizing colors.And it tends to be more error free.And in this situation,same ink color had worked as a catalyst. So it is obvious, congruent task is faster than incongruent task. On the other hand,in incongruent task, people had to recognize the color name from its visual.Additionally, they have to avoid misleading wrong color name,which were written as a word.

If same test could be organized with a group of people those are not very well with a language like "Spanish".Let's say ,if they are in their beginner or intermediate stage of learning Spaning language. In that situation, I think that group of people would complete incongruent task faster than congruent task.

1 Reference :-

https://docs.scipy.org/doc/numpy-dev/user/quickstart.html

https://pandas.pydata.org/pandas-docs/stable/generated/pandas.read_csv.html

https://matplotlib.org/gallery/shapes_and_collections/scatter.html

https://matplotlib.org/devdocs/api/_as_gen/matplotlib.pyplot.bar.html

https://docs.scipy.org/doc/numpy/reference/generated/numpy.histogram.html

https://plot.ly/matplotlib/histograms/

https://docs.scipy.org/doc/scipy-0.15.1/reference/generated/scipy.stats.ttest_rel.html

https://brandalyzer.blog/2010/12/05/difference-between-z-test-f-test-and-t-test/