Test a Perceptual Phenomenon

Aditya Naga Sanjeevi, Yellapu
(sanjeev.any@gmail.com)

Project Description

Use descriptive statistics and a statistical test to analyse the Stroop effect, a classic result of experimental psychology. Give your readers a good intuition for the data and use statistical inference to draw a conclusion based on the results.

Background Information

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the *color of the ink* in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the *congruent words* condition, the words being displayed are color words whose names match the colors in which they are printed: for example, RED, BLUE. In the *incongruent words* condition, the words displayed are color words whose names do not match the colors in which they are printed: for example, PURPLE, ORANGE. In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

Investigation

Resources

1. http://psychclassics.yorku.ca/Stroop/
2. https://www.psytoolkit.org/lessons/stroop.html
3. https://en.wikipedia.org/wiki/Stroop_effect
4. https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg

Questions and Answers

Question 1: What is our independent variable? What is our dependent variable?

Independent Variable: Congruency condition - congruent words condition or incongruent words condition.

Dependent Variable: Time taken to identify the ink color.

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

Stroop Effect is a demonstration of interference in the reaction time of a task. It is supposed that there could be a difference in time taken to respond to certain environmental stimuli while ignoring others. In this case, the *Stroop task*, there is an incongruency in the displayed color to the named color. The hypothesis could be that there is absolutely no time difference in recognizing the color of the ink irrespective of what it is named (null hypothesis) or there is a difference of time taken to recognize the color of the ink with respective to what it is named (alternate hypothesis).

The environment: A group of individuals are first provided with congruent words and time taken to complete the test is record for each of the individual. Later the same group of individuals are provided with incongruent words and the time taken to complete the test is record for each of them. We do not know the population mean or population standard deviation.

We need a hypothesis testing tool. To verify, the selected test is T-test as it is one type of inferential statistics. It is used to determine whether there is a significant difference between the means of two groups. With all inferential statistics, we assume the dependent variable fits a normal distribution. In our case, we do not have any information about the population (mean or standard deviation), both the samples are dependent, and we assume that both distributions are normal.

The hypothesis to test:

Null Hypothesis (H0):
H0: Difference between average of congruent to average of incongruent is zero.

$\mu_c - \mu_{ic} = 0$

Alternate Hypothesis (H1):
H1: Difference between average of congruent to average of incongruent is not zero.

$\mu_c - \mu_{ic} != 0$

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

Measure of Central tendency:

```
In [8]: import math
        import pandas as pd
        import numpy as np

        data = pd.read_csv('./stroopdata.csv')
```

```
In [10]: print(data.mean())

    Congruent      14.051125
    Incongruent    22.015917
    dtype: float64
```

```
In [11]: print(data.median())

    Congruent      14.3565
    Incongruent    21.0175
    dtype: float64
```

Measure of Variability

```
In [8]: import math
        import pandas as pd
        import numpy as np

        data = pd.read_csv('./stroopdata.csv')
```

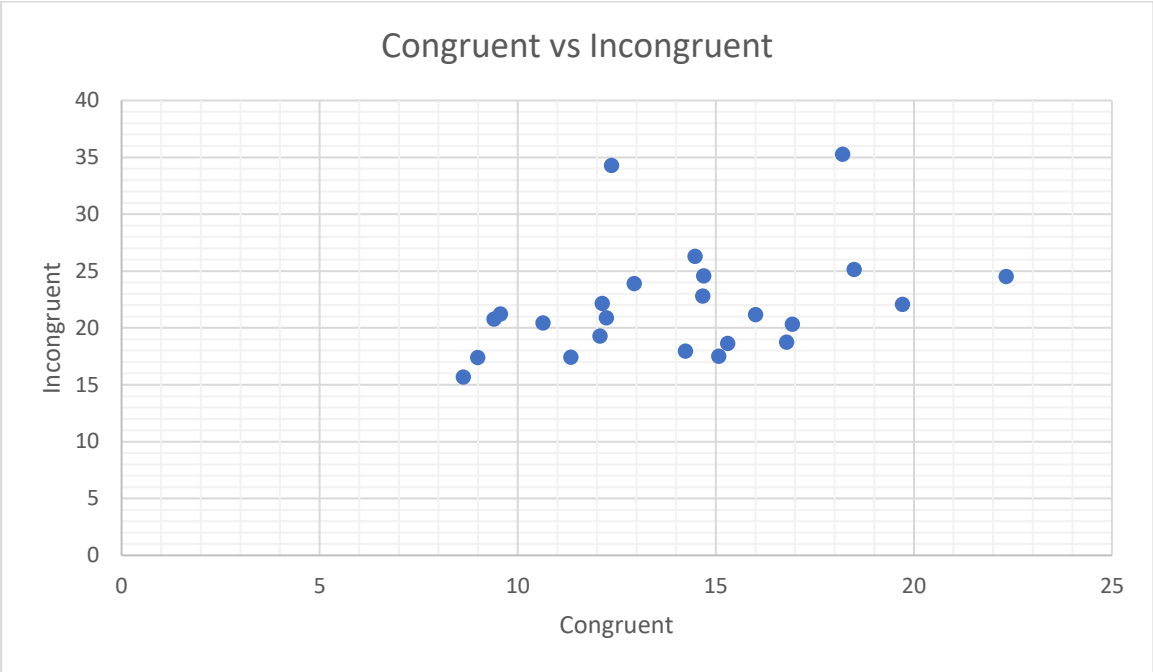
```
In [13]: print(data.var())

    Congruent      12.669029
    Incongruent    23.011757
    dtype: float64
```

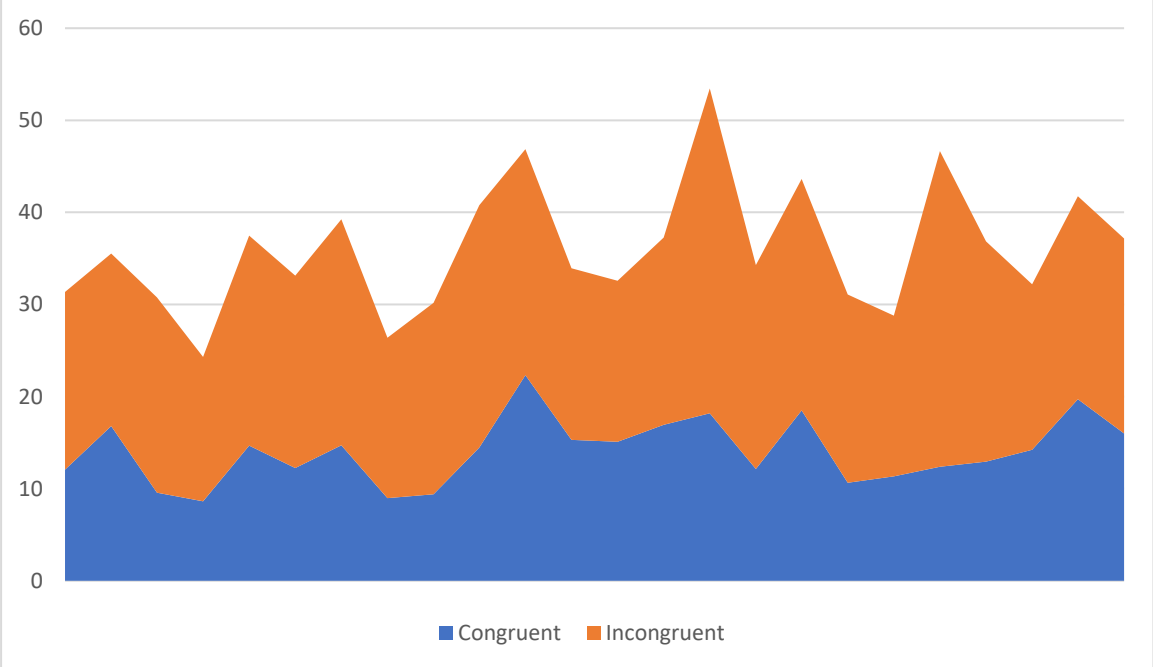
```
In [14]: print(data.std(ddof=0))

    Congruent      14.3565
    Incongruent    21.0175
    dtype: float64
```

Question 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.



While there is a slight correlation shown, the plot above shows couple outliers and that incongruent values are always higher than the congruent values.



The stacked chart above shows that incongruent values are ALWAYS greater than the congruent values.

Question 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?

In [19]:
```
# Sample size
sample_size = len(data)
print(sample_size)
```

```
24
```

In [24]:
```
congruent_mean = data['Congruent'].mean()
incongruent_mean = data['Incongruent'].mean()
print(round(congruent_mean,2))
print(round(incongruent_mean,2))
```

```
14.05
22.02
```

In [25]:
```
# PE: point estimate (in seconds)
point_estimate = incongruent_mean - congruent_mean
print(round(point_estimate,2))
```

```
7.96
```

In [30]:
```
# Sample difference
data['sample_difference'] = data['Incongruent'] - data['Congruent']

# display first and last rows
pd.set_option('display.max_rows', 2)
print(data['sample_difference'])
```

```
0     7.199
      ...
23    5.153
Name: sample_difference, Length: 24, dtype: float64
```

In [31]:
```
# Squared differences from mean
differences_from_mean = data['sample_difference'] - data['sample_difference'].mean()
data['squared_differences'] = differences_from_mean**2

# display first and last rows
print(data['squared_differences'])
```

```
0     0.586437
        ...
23    7.906172
Name: squared_differences, Length: 24, dtype: float64
```

In [33]:
```
# Sum of squared differences
sum_squared_differences = data['squared_differences'].sum()
print(round(sum_squared_differences, 2))
```

```
544.33
```

In [35]:
```
# Variance (sample in population)
variance = sum_squared_differences/(sample_size -1)
print(round(variance, 2))
```

```
23.67
```

In [40]:
```
# Sample standard deviation of differences
sample_standard_deviation = sqrt(variance)
print(round(sample_standard_deviation, 2))
```

```
4.86
```

In [42]:
```
# t-statistic = PE/(sd/√n)
t_statistic = point_estimate /(sample_standard_deviation / sqrt(sample_size))
print(round(t_statistic,2))
```

```
8.02
```

t-statistic = *8.02* (from above)

For confidence level = 95%, α= 0.05

Degrees of freedom = sample size – 1 = 23

From t-table, **t-critical** (two-tailed) = ±*1.714*

t-statistic (8.02) > t-critical (1.714)

Since t-statistic (8.02) is greater than the positive t-critical (1.17), we reject the null hypothesis. This means that the difference between the samples (congruent and incongruent) is statistically significant. The results match my expectations.

Question 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!

Viewing and recognizing of colors is one action for brain and reading a word and understanding of the meaning of the word is a different action for the brain. When brain has to function both the tasks together and there is a difference in perception between both the actions, brain takes longer time to complete the action.