# Stroop Effect

P V N S Rama Krishna

March 13, 2019

## 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 :** *congruent* word condition, *incongruent* word condition
**Dependent variable :** *time taken* to name the ink colors in equally-sized lists

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

**Null Hypothesis (** $H_0 : \mu_1 = \mu_2$**)** * : The average population of the time taken to name the ink colors in equally-sized lists for the two groups (congruent and incongruent), are equal.

**Alternative Hypothesis (** $H_0 : \mu_1 \neq \mu_2$**)** * $H_1$ : The average population of the time taken to name the ink colors in equally-sized lists for two groups (congruent and incongruent) are not equal.

**Statistical Test :** Dependent t-test (two tailed).

***Justification***:

- We need to compare the means of the two related groups to determine the statistically significant difference between these two means.
- We are assuming a normal distribution.
- There are less than 30 samples.
- It has to be two-tailed because our hypothesis is testing the 'equality of two means'.

**(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]: # Importing all the necessary packages
        import pandas as pd
        import matplotlib.pyplot as plt
        from scipy import stats

        %matplotlib inline
```

```
In [2]: # Load the data and store it as a Pandas DataFrame
        df = pd.read_csv('stroopdata.csv')

        # Print the first 5 rows of this DataFrame
        print(df.head())

        print('\n')
        print("-------")
        print("SUMMARY")
        print("-------")
        print(df.describe())   # Show a complete and quick statistical summary of the dataset
```

```
   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


-------
SUMMARY
-------
       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
```

From the above statistical description of the data, we can infer the following :-

**Central tendency** (*MEAN*)

- Congruent mean = 14.0511
- Incongruent mean = 22.0159

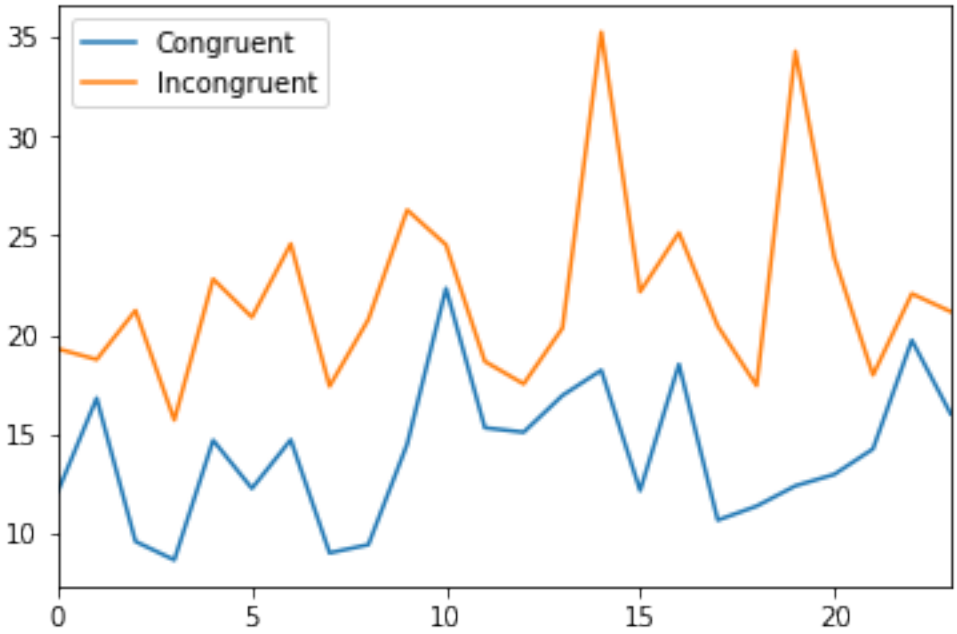
**Measure of variability** (*Standard Deviation*)

- Congruent std = 3.5594
- Incongruent std = 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.

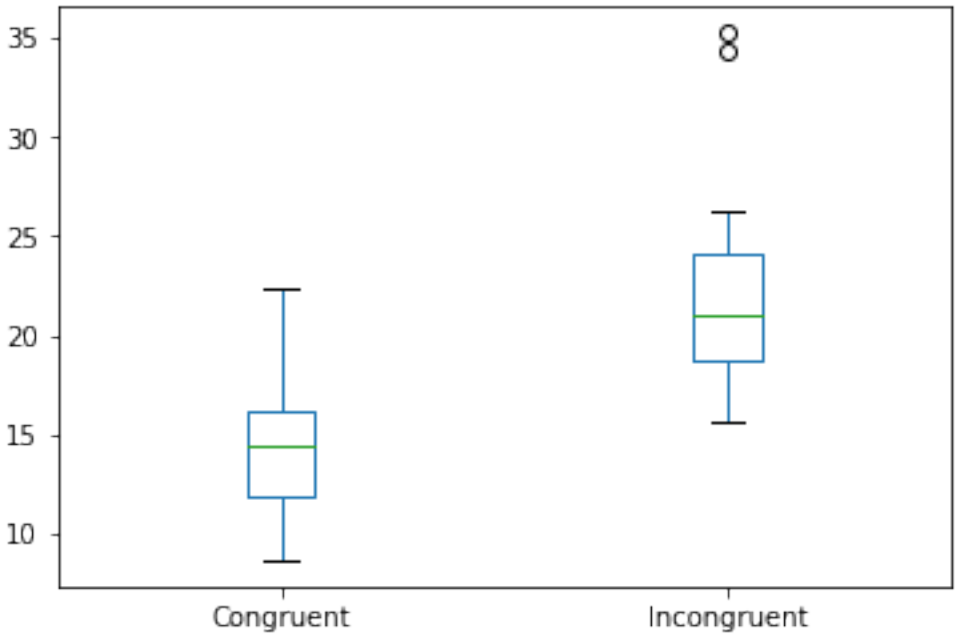
In [3]:
```
df.plot()    # Line Plot
```

Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1b10c780>



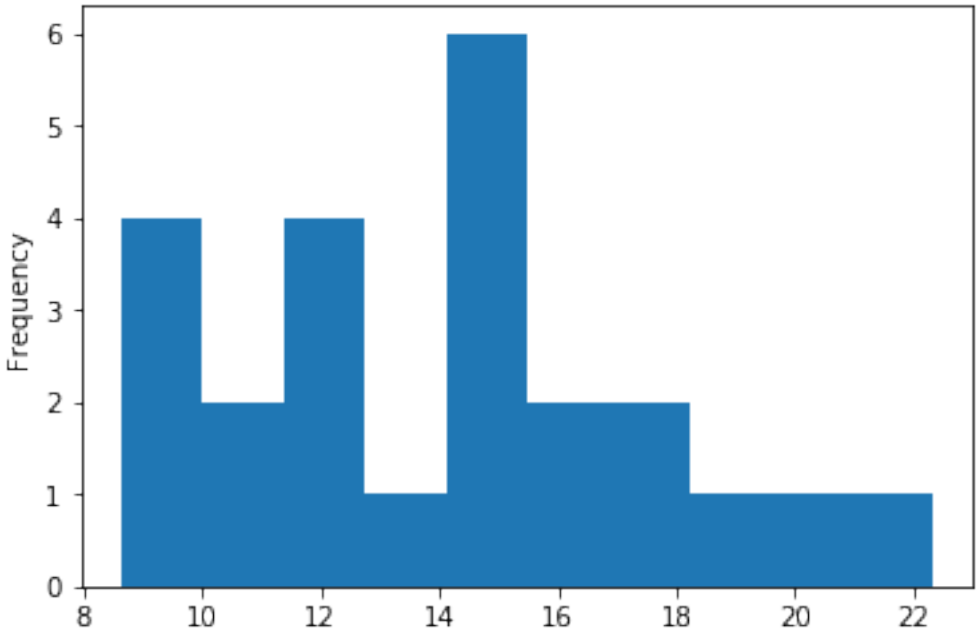

In [4]:
```
df.plot(kind='box')    # Box Plot which gives information about various quartiles and als
```

Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1b4ed550>



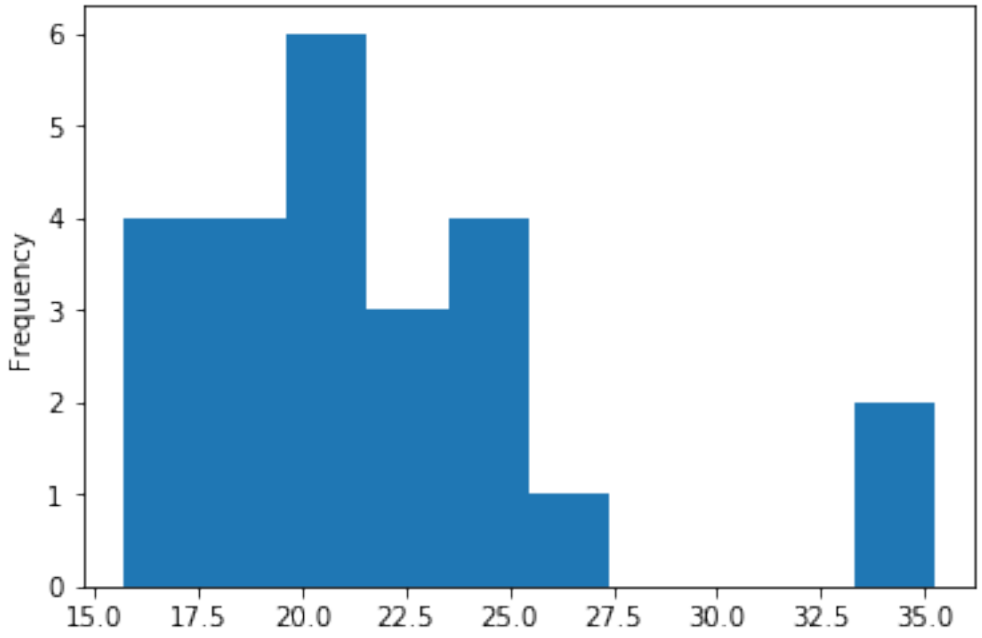
```
In [5]: df['Congruent'].plot(kind="hist")    # Histogram showing the data distribution for the 'C
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1b4ac748>
```



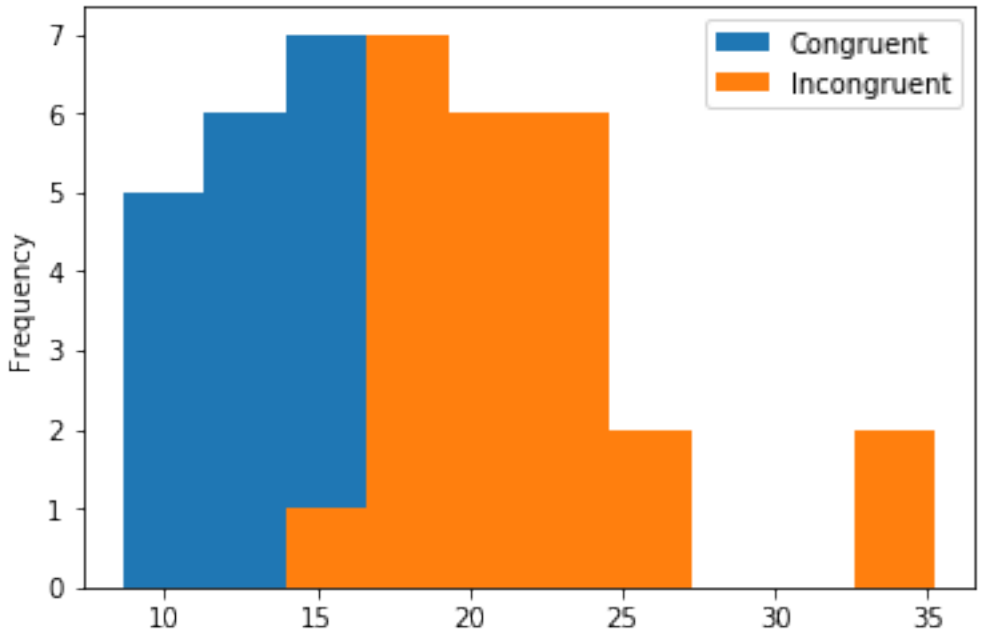

```
In [6]: df['Incongruent'].plot(kind="hist")    # Histogram showing the data distribution for the
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1b5c08d0>
```




```
In [7]: df.plot(kind="hist")    # Now, let's plot both of them side-by-side
```

```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1b6cac50>
```



By looking at the Boxplot, we can see that the average completion time of the 'Incongruent group' is *higher* (as indicated by the level of the bar). The box plot also shows that the Incongruent group has two outliers (in between 30-35). Also, the distribution for both the Congruent and Incongruent groups looks normal.

**(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 [8]: # Perform the statistical test here
        degreesOfFreedom = 23
        stats.t.ppf(0.975, degreesOfFreedom)

        # Calculates the T-test on TWO RELATED samples of scores, a and b.
        stats.ttest_rel(df['Congruent'], df['Incongruent'])
```

```
Out[8]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.103000585711178e-08)
```

From the above statistic, we can clearly see that the **p** value is *under 0.05*. We therefore reject our null hypothesis. Thus, we can conclude that the Incongruent group and the Congruent group have *different* average population time to complete the tasks.

# 1 References


- Dependent t-test
- Determining Sample Size
- Which chart or graph is right for you?
- Pandas visualization
- Scipy stats