Test a Perceptual Phenomenon

June 2, 2019

0.0.1 Analyzing the Stroop Effect

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

0.0.2 Independent Variable Definition

An **independent variable** is defines as the variable that is changed or controlled in a scientific experiment. It represents the cause or reason for an outcome.

Independent variables are the variables that the experimenter changes to test their dependent variable. A change in the independent variable directly causes a change in the dependent variable. The effect on the dependent variable is measured and recorded.

0.0.3 Dependent Variable Definition

A **dependent variable** is the variable being tested in a scientific experiment.

The **dependent variable** is 'dependent' on the independent variable. As the experimenter changes the independent variable, the change in the dependent variable is observed and recorded. When you take data in an experiment, the dependent variable is the one being measured.

In this experiment The **dependent variable** is the **response time** and the **independent variable** is the **congruency condition**

(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.

Null hypothesis (H_0): Congruency does not influence the time to complete the test.

Alternative hypothesis (H_1): Congruency influences the time to complete the test, so congruent tasks take shorter time to complete than the incongruent tasks.

$$H_0 : \mu_d = 0 \; H_1 : \mu_d 0$$

Where :

$$\mu_d = \mu_i - \mu_c$$

μ_i is the population mean of incongruent values.
μ_c is the population mean of congruent values.
Statistical test is **Dependent t-test** for :

- There is no data about the population, so the population standard deviations is unknown

- The distribution should be approximately normally distributed.

- The Normal distribution is symmetric around the center(mean)

- There is only one sample that has been tested twice (repeated measures) as the same subjects were tested for congruent and incongruent words.

- Dependent variable should be measured on a continuous scale.

- The dependent t-test requires the sample data to be numeric and continuous, as it is based on

(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]:
```
# performing the analysis here
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
% matplotlib inline
```

In [2]:
```
df = pd.read_csv('stroopdata.csv')
df.head()
```

Out[2]:

	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

In [3]:
```
#measuring of central tendency(mean)
df.Congruent.mean(),df.Incongruent.mean()
```

Out[3]: (14.051124999999999, 22.015916666666666)

In [4]:
```
#calculating the median
df.Congruent.median(),df.Incongruent.median()
```

Out[4]: (14.3565, 21.017499999999998)

In [5]:
```
#measuring of variability(variance)
df.Congruent.var(), df.Incongruent.var()
```

Out[5]: (12.669029070652176, 23.011757036231884)

In [6]:
```
#calculating the standard diviation
df.Congruent.std(), df.Incongruent.std()
```

Out[6]: (3.5593579576451955, 4.7970571224691376)

0.0.4 Measure of central tendency

```
-The Congruent mean = 14.05
-The Incongruent mean = 22.02
-The Congruent median = 14.36
-The Incongruent median = 22.02
```

0.0.5 Measure of variability

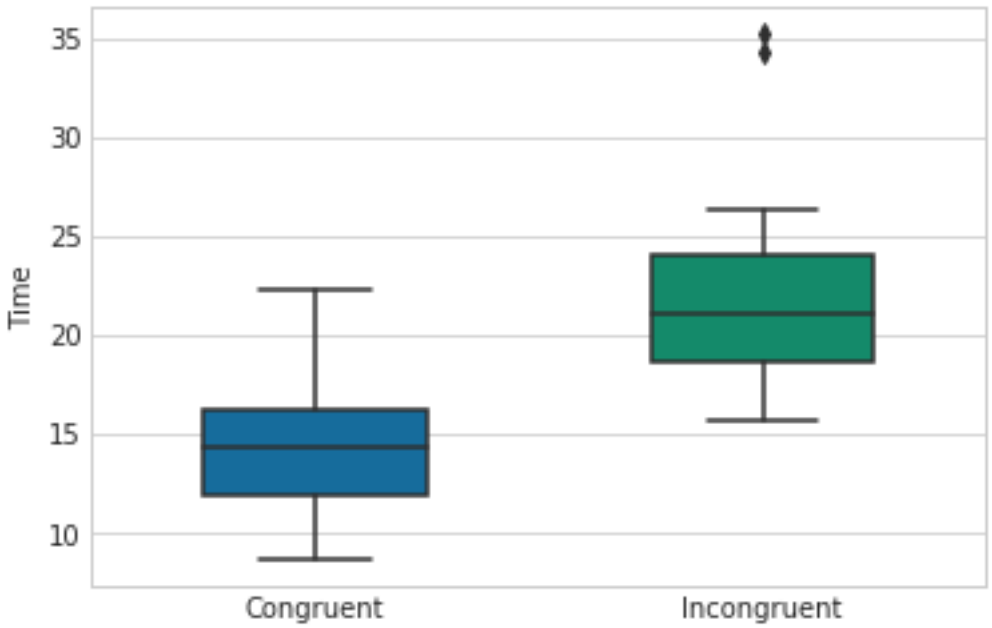
```
-The Congruent variance = 12.67
-The Incongruent variance = 23.01
-The Congruent median = 3.56
-The Incongruent median = 4.80
```

```
In [7]: #more informations
        df.describe()
```

```
Out[7]:        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
```

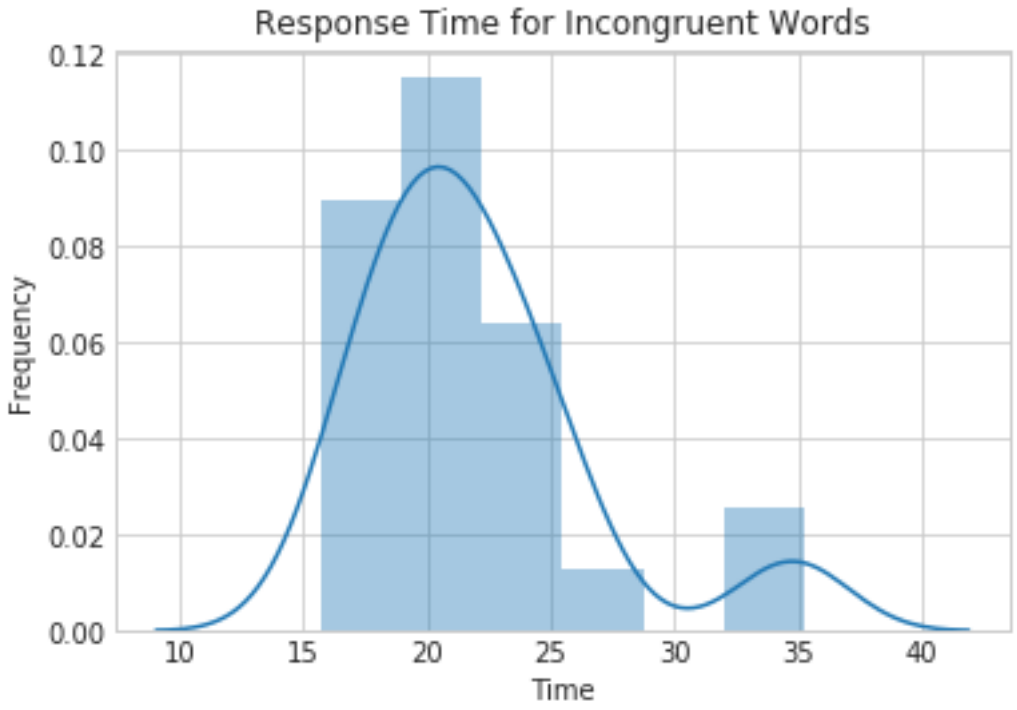
(4) Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

```
In [8]: #plot a box
        sns.set_style("whitegrid")
        sns.boxplot(data=df[['Congruent', 'Incongruent']], orient="v",  width=0.5, palette="colo
        plt.ylabel("Time");
```

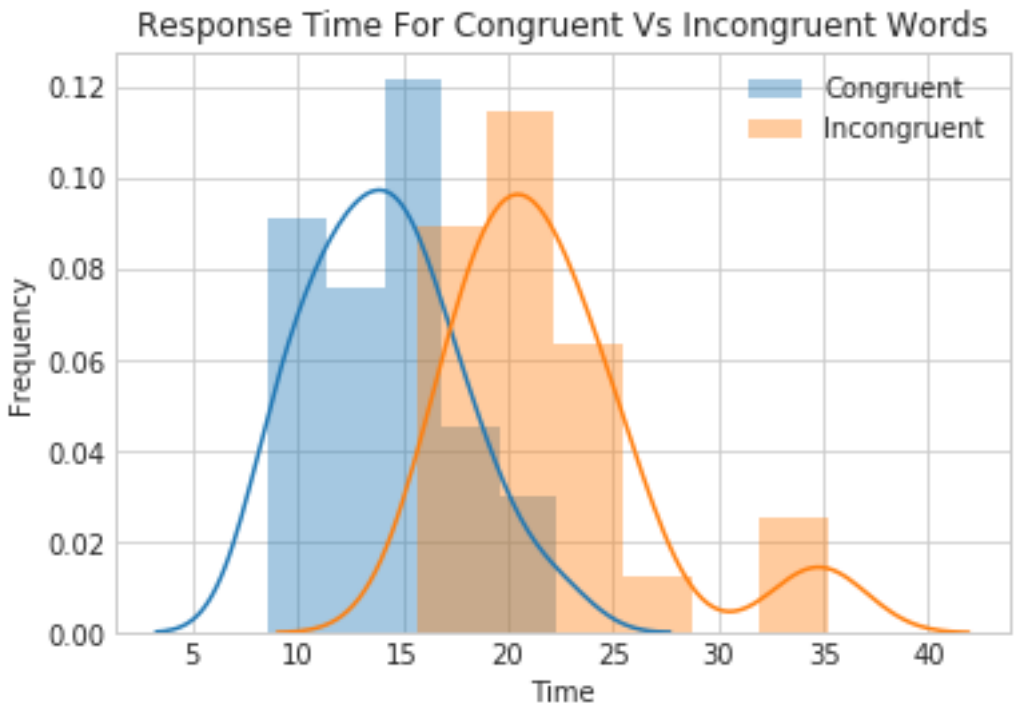


In [9]:
```
#plot a graph for congruent dataset
sns.distplot(df['Congruent'])
plt.xlabel("Time");
plt.ylabel("Frequency");
plt.title("Response Time for Congruent Words");
```

In [10]:
```
#plot a graph for incongruent dataset
sns.distplot(df['Incongruent'])
plt.xlabel("Time");
plt.ylabel("Frequency");
plt.title("Response Time for Incongruent Words");
```



In [11]:
```
#comparing both the datasets
sns.distplot(df['Congruent'],label = "Congruent")
sns.distplot(df['Incongruent'],label = "Incongruent")
plt.xlabel("Time");
plt.ylabel("Frequency");
plt.title("Response Time For Congruent Vs Incongruent Words");
plt.legend();
```



Answer

-The box plot clearly displays the difference between the median of two datasets.
-In the box plot the distribution of time taken to name the color for congruent words are betwee
-There are two outliers in the distribution of incongruent words.
-Both the distributions are look like the normal distribution.
-The mean is different for both the distributions.

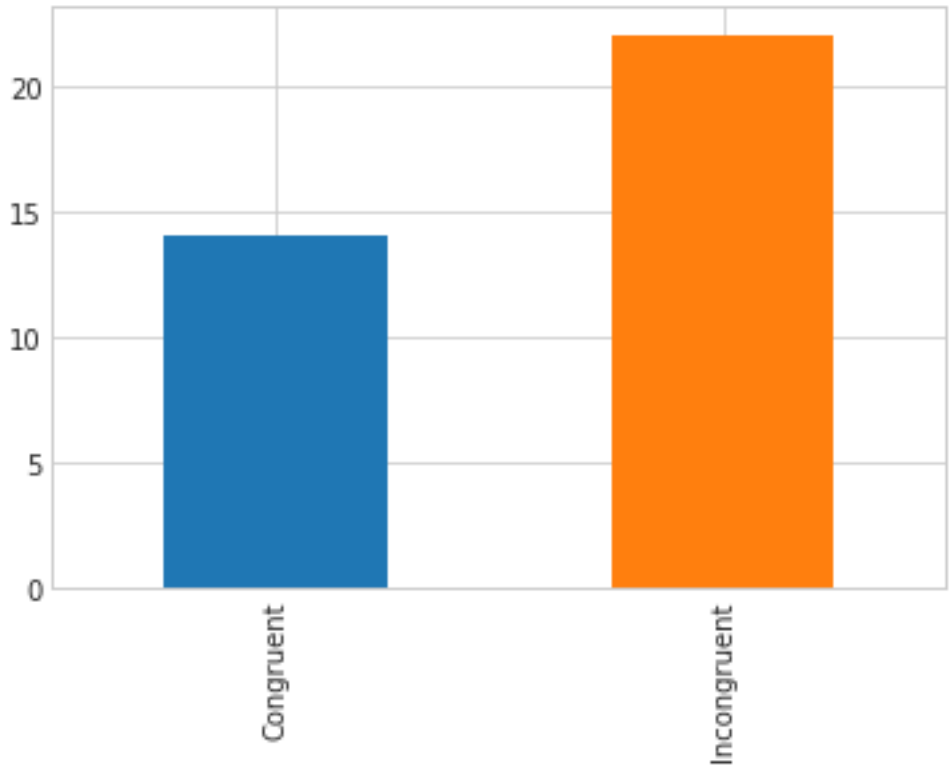
(5) Now, perform the statistical test and report your results. What is your confidence level or Type I error associated with your test? What is your conclusion regarding the hypotheses you set up? Did the results match up with your expectations? **Hint:** Think about what is being measured on each individual, and what statistic best captures how an individual reacts in each environment.

In [12]:
```
# Perform the statistical test here
# Set columns to lists to use in ttest function
con = df['Congruent'].values.tolist()
inc = df['Incongruent'].values.tolist()

# Plot the means
df.mean().plot('bar')

# Perform ttest
stats.ttest_rel(con, inc)
```

Out[12]:
```
Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```



Answer :

Null hypothesis is rejected because **pvalue** is **less than = 0.05**, the time to name colours is significantly different between congruent and incongruent tasks. People do not name colours at the same speed when the word's meaning and its colour match, as when they do not match.

(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!

The interference between what the words say and the color of the words seem to confuse the brain.

There are **two theories** that may explain the Stroop effect:

```
-Speed of Processing Theory: the interference occurs because words are read faster than colors a
-Selective Attention Theory: the interference occurs because naming colors requires more attenti
```

Alternative tasks to try:

```
-Use non-color words such as "dog" or "house."
-Use emotional words such as "sad" or "happy" or "depressed" or "angry."
```

0.1 References

Stroop Effect

Stroop Effect Wiki
Student's_t-test
Paired t-tests
T-tests
scipy.stats.ttest_rel
thoughtco

In [ ]: