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.

Analysis

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

Answer: The two main variables in an experiment are the independent and dependent variable. An independent variable is the variable that is changed or controlled in a scientific experiment to test the effects on the dependent variable.If the independent variable is changed, then an effect is seen in the dependent variable.

The given data set stroopdata.csv contains both an independent and dependent variable. In this experiment, the dependent variable is the response time for each participant to name the ink color and the independent variable is the congruency condition including congruent words condition and incongruent words condition.

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.

Answer:

- I choose the two tailed t-test for my given data set. The dependent t-test compares the mean of two paired groups to see if there are statistically significant differences between these means. The same subjects were tested for congruent and incongruent words. By using the same subject to test two different condition, we eliminate the individual differences that occur between subjects.

Reason Of Selecting Two Tailed t-test

- We don't have any imformation about the population, we have just 24 sample datasets.
- The ability to compare the means of the dataset for the pre and post test validates the benefit of this test selection
- We don't have direction of the test so we choose two tailed test.

Assumptions:

- Normality: The distribution of the differences in the dependent variable between the two related groups should be approximately normally distributed. Normal distributions are symmetric around the center(mean) and follow a 'bell- shaped' distribution.
- The observations are independent of one another.
- 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 the normal distribution.
- There should be no significant outliers in the differences between the two related groups. Outliers are simply single data points within your data that do not follow the usual pattern.

Hypothesis

Null Hypothesis: A null hypothesis is a statistical hypothesis in which there is no significant difference exist between the set of variables. It is the original or default statement, with no effect, often represented by H0 (H-zero). In this experiment the null hypothesis is that there is no

significant difference exist between the time taken to say the congruent words and incongruent words, means the Mean (xbar) of the Reaction time for the congruent sample will be equal to the mean(xbar) of the reaction time for the incongruent sample.

Alternative Hypothesis: A alternative hypothesis is a statistical hypothesis in which there is no significant difference exist between the set of variables. Often denoted by H1 (H-one).In this experiment the alternative hypothesis is that there is significant difference exist between the time taken to say the congruent words and incongruent words, means the Mean (xbar) of the Reaction time for the congruent sample will not be equal to the mean(xbar) of the reaction time for the incongruent sample.

H_0 (Null Hypothesis) : (Population Mean)$_{\text{incongruent}}$ = (Population Mean)$_{\text{congruent}}$ OR $\mu_i = \mu_c$

H_1 (Alternative Hypothesis) : (Population Mean)$_{\text{incongruent}}$ ≠ (Population Mean)$_{\text{congruent}}$ OR $\mu_i \neq \mu_c$

- Where μ_i = Mean time taken to name the ink color for incongruent words.
- μ_c = Mean time taken to name the ink color for congruent words.

- If we get a significant result, we can reject the null hypothesis and accept the alternative hypothesis that there are statistically significant differences between the mean time taken to name ink color between two test conditions.

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

In [2]:
```
import math
import seaborn as sns
import numpy as np
import pandas as pd
from scipy.stats import t as pt
import matplotlib.pyplot as plt
%matplotlib inline
```

In [3]:
```
#Take all the data from the csv file and print
data = pd.read_csv("stroopdata.csv")
data
```

Out[3]:

	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
5	12.238	20.878
6	14.692	24.572
7	8.987	17.394
8	9.401	20.762
9	14.480	26.282
10	22.328	24.524
11	15.298	18.644
12	15.073	17.510
13	16.929	20.330
14	18.200	35.255
15	12.130	22.158
16	18.495	25.139
17	10.639	20.429
18	11.344	17.425
19	12.369	34.288
20	12.944	23.894
21	14.233	17.960
22	19.710	22.058
23	16.004	21.157

In [4]:
```
#Find the mean,Median and standard deviation for congruent and incongruent
#Find the mean
c_mean = data['Congruent'].mean()
i_mean = data['Incongruent'].mean()

#Find the median
c_median = data['Congruent'].median()
i_median = data['Incongruent'].median()

#Find the standard deviation
c_std = data['Congruent'].std()
i_std = data['Incongruent'].std()

#print mean,median and standard deviation in a table
ls = [[int(24),int(24)],[c_mean,i_mean],[c_median,i_median],[c_std,i_std]]
detail = pd.DataFrame(ls,index=['Sample Size','Mean','Median','Standard Deviation
detail
```

Out[4]:

	Congruent	Incongruent
Sample Size	24.000000	24.000000
Mean	14.051125	22.015917
Median	14.356500	21.017500
Standard Deviation	3.559358	4.797057

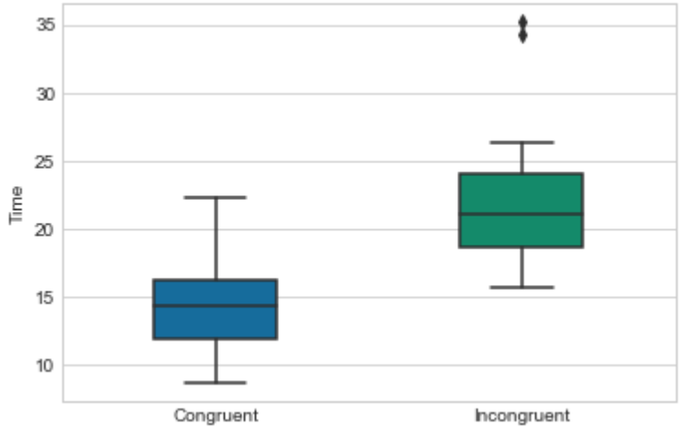
In [5]:
```
#More Detail About the dataset
data.describe()
```

Out[5]:

	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

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?
Answer:

In [6]:
```
#Make a boxplot for both datasets
sns.set_style("whitegrid")
sns.boxplot(data=data[['Congruent', 'Incongruent']], orient="v",width=0.4, palett
plt.ylabel("Time");
```



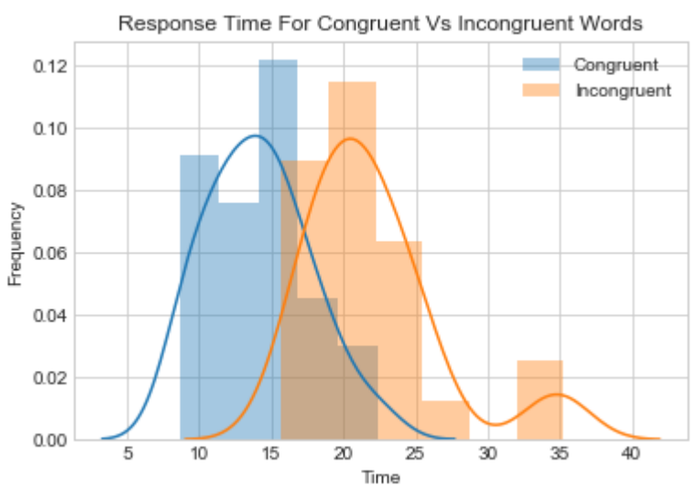
Outcome:

- The box plot clearly displays the difference between the median of two datasets.
- As we can see in the box plot the distribution of time taken to name the color for congruent words are between 8 to 23 and the distribution of time taken to name the color for incongruent words are between 16 to 36. And we can see two outliers in the distribution of incongruent words.

In [7]:
```
#Plot a Graph for congruent dataset
sns.distplot(data['Congruent'])
plt.xlabel("Time");
plt.ylabel("Frequency");
plt.title("Response Time for Congruent Words");
```

In [8]:
```
#Plot a Graph for congruent dataset
sns.distplot(data['Incongruent'])
plt.xlabel("Time");
plt.ylabel("Frequency");
plt.title("Response Time for Incongruent Words");
```

In [9]:
```
#Compare Both the datasets and make a graph
sns.distplot(data['Congruent'],label = "Congruent")
sns.distplot(data['Incongruent'],label = "Incongruent")
plt.xlabel("Time");
plt.ylabel("Frequency");
plt.title("Response Time For Congruent Vs Incongruent Words");
plt.legend();
```



Both the distribution are looks like the normal distribution and we can see that the mean is different for both the distributions.

Questions: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?

Answer:

- The Test is two tailed t- test so find the critical value of t at 95% confidance level.
- Sample Size(n) = 24
- Degree of freedom(df) = n - 1 = 23

In [10]:
```
#find t-critical value for 95% confidance interval and 23 degree of freedom for t

print("t-critical value for two tailed test is: ",round(pt.ppf(0.975,23),4))
```

```
t-critical value for two tailed test is:  2.0687
```

In [11]:
```
#find the diffenence of each data
data['difference'] = data['Congruent'] - data['Incongruent']
data
```

Out[11]:

	Congruent	Incongruent	difference
0	12.079	19.278	-7.199
1	16.791	18.741	-1.950
2	9.564	21.214	-11.650
3	8.630	15.687	-7.057
4	14.669	22.803	-8.134
5	12.238	20.878	-8.640
6	14.692	24.572	-9.880
7	8.987	17.394	-8.407
8	9.401	20.762	-11.361
9	14.480	26.282	-11.802
10	22.328	24.524	-2.196
11	15.298	18.644	-3.346
12	15.073	17.510	-2.437
13	16.929	20.330	-3.401
14	18.200	35.255	-17.055
15	12.130	22.158	-10.028
16	18.495	25.139	-6.644
17	10.639	20.429	-9.790
18	11.344	17.425	-6.081
19	12.369	34.288	-21.919
20	12.944	23.894	-10.950
21	14.233	17.960	-3.727
22	19.710	22.058	-2.348
23	16.004	21.157	-5.153

In [12]:
```
#sd and mean of the differenced dataset
s_std = data['difference'].std()
print("Standard Deviation of the diffenenced dataset: ", round(s_std,4))
s_mean = c_mean - i_mean
print("Mean of Difference: ", round(s_mean,4))
```

```
Standard Deviation of the diffenenced dataset:  4.8648
Mean of Difference:  -7.9648
```

- For two tailed t-test at 95% confidence level
- Sample size(n) = 24
- Degree Of Freedom(df) = 23
- t - critical ($t_{critical}$) = ∓2.069
- Standerd Deviation(SD) = 4.8648
- Mean of difference data ($\bar{x}$) = -7.9648
- Confidance Interval(CI) at 95% level = (-10.019028, -5.910555)

In [13]:
```
# Calculate the t-value
t_value = s_mean/(s_std/math.sqrt(24))
print("t-Value is: ",t_value)
```

```
t-Value is:  -8.020706944109955
```

- The calculated t-statistic is -8.0207 for the difference in colour recognition time means of the congruent and incongruent words data. And -8.0207 is much lower than the t-critical value. Since the t-statistic is in the critical region, so the null hypothesis is rejected.
- **Null hypothesis rejected** At α = 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. The result confirms my expectations.
- According to the confidence intervals, we can say that the true difference between the congruent and incongruent group mean time is between -10.019028 and -5.910555.

Question:6 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?
Answer:

The brain has an image association between the shape of the word and the colour. When there is a mismatch, additional time is necessary for the prefrontal cortex to process the information and decide on its meaning.The words themselves have a strong influence over your ability to say the color. The interference between the different information (what the words say and the color of the words) your brain receives causes a problem. There are two theories that may explain the Stroop effect:

- Speed of Processing Theory: the interference occurs because words are read faster than colors are named.
- Selective Attention Theory: the interference occurs because naming colors requires more attention than reading words

Similar effect to the this effect can be reserve stroop effect.In this test the participant's task is to say out loud the word that is printed not its color.