Analyzing the Stroop Effect

Perform the analysis in the space below. Remember to follow the instructions (https://docs.google.com/document/d/1-OkpZLjG_kX9J6LIQ5IltsqMzVWjh36QpnP2RYpVdPU/pub?embedded=True) and review the project rubric (https://review.udacity.com/#!/rubrics/71/view) before submitting. Once you've completed the analysis and write-up, download this file as a PDF or HTML file, upload that PDF/HTML into the workspace here (click on the orange Jupyter icon in the upper left then Upload), then use the Submit Project button at the bottom of this page. This will create a zip file containing both this .ipynb doc and the PDF/HTML doc that will be submitted for your project.

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

The independent variable is the condition, congruent words section or incongruent words section. The dependent variable is the time taken to complete each 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.

Time taken to complete congruent words section :

$$\mu_c$$

Time taken to complete incongruent words section:

$$\mu_{ic}$$

Hence we can specify our hypothesis as :

$$H_0 : \mu_{ic} - \mu_c = 0$$

$$H_1 : \mu_{ic} - \mu_c > 0$$

The reason for this is that we are trying to prove that the time taken for both cases is not the same.

The null hypothesis states that the population means of the 2 cases do not statistically differ while the alternate hypothesis states that the population means between the 2 cases are statistically different.

We are mainly trying to prove that despite having a small sample we are trying to make a conclusion on a larger population

(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 'stroopdata.csv'.

In [2]:
```
import pandas as pd
import numpy as np
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]:
```
df.describe()
```

Out[3]:

	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

In [4]:
```
#Range of Congruent
df['Congruent'].max() - df['Congruent'].min()
```

Out[4]: 13.698000000000002

In [5]:
```
#Range of Incongruent
df['Incongruent'].max() - df['Incongruent'].min()
```

Out[5]: 19.568

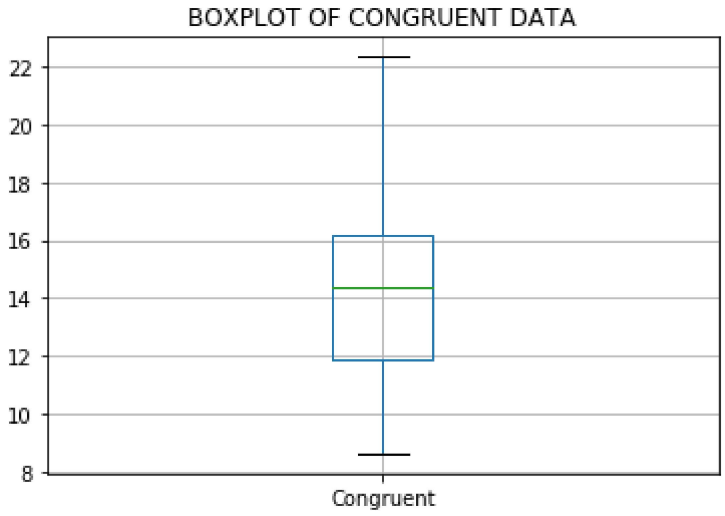
We can see from the means in both cases that the Incongruent words section has a higher average time taken.
The variability of the incongruent times are also higher as observed from the Standard Deviations.

The Incongruent words also have a higher range than the Incongruent words

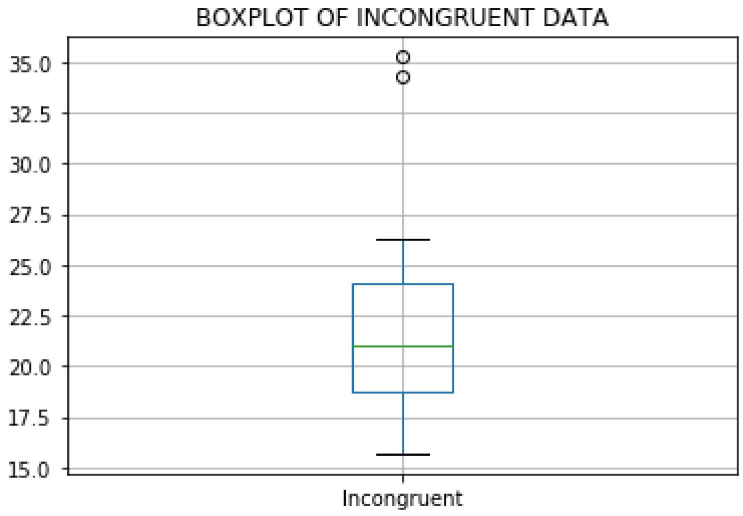
(4) Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

In [50]:
```
#Import packages and histograms
import matplotlib.pyplot as plt
% matplotlib inline
```

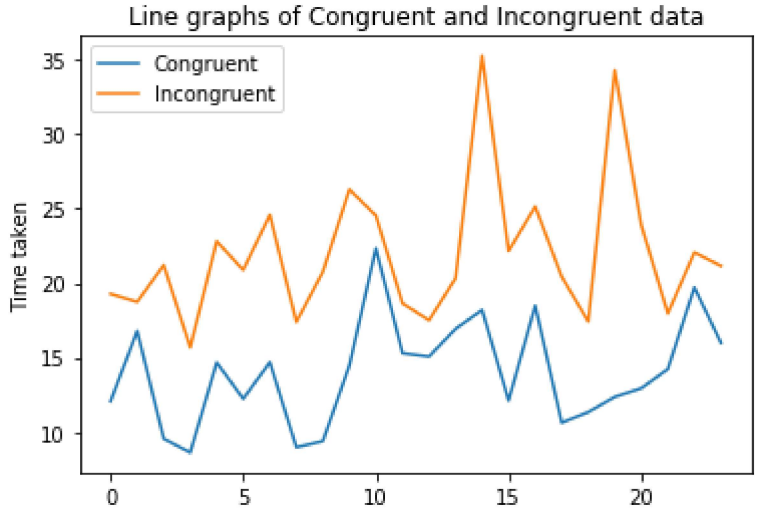
In [40]:
```
#Box plot of Congruent section
df.boxplot('Congruent');
plt.title('BOXPLOT OF CONGRUENT DATA');
```



In [39]:
```
#Box plot of Incongruent section
df.boxplot('Incongruent');
plt.title('BOXPLOT OF INCONGRUENT DATA');
```



In [49]:
```
plt.plot(df['Congruent']);
plt.plot(df['Incongruent']);
plt.ylabel('Time taken');
plt.title('Line graphs of Congruent and Incongruent data');
plt.legend();
```



From the boxplots we can observe that the Incongruent data has 2 outliers at 35.0.
**From the line graphs we can observe that for all cases the time taken for Incongruent words is higher than that for Congruent words. **

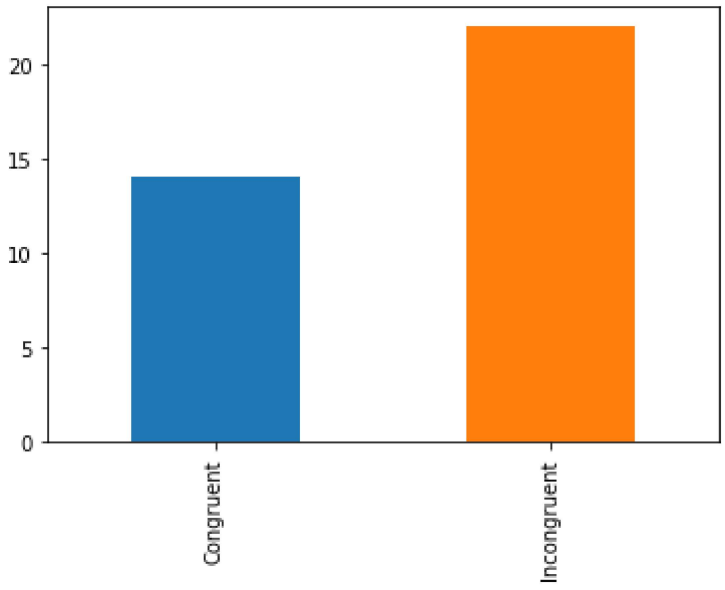
(5) Now, perform the statistical test and report your results. What is your confidence level or Type I error associated with your test? What is your conclusion regarding the hypotheses you set up? Did the results match up with your expectations? Hint: Think about what is being measured on each individual,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

Since our sample size is very small here with a meager 24 samples, I have decided to use a t-test as our statistical test.
Also, we are trying to compare between 2 population means and the Incongruent data is dependent on the Congruent data, hence we need to use a paired t-test.

In [78]:
```
import scipy.stats as stats
cong = df['Congruent'].values.tolist()
incong = df['Incongruent'].values.tolist()

df.mean().plot('bar');
stats.ttest_rel(df['Incongruent'],df['Congruent'])
```

Out[78]: Ttest_relResult(statistic=8.020706944109957, pvalue=4.103000585711178e-08)



We are attempting to prove that the two samples t-statistics do statistically differ at an alpha of 0.05. We have obtained a p-value of 0.000000004103.

$$p < \alpha$$

Therefore we can reject the null hypothesis.

Hence even for a large population, the difference between the 2 population means is greater than 0. This means that generally:

Mean time taken for Incongruent section > Mean time taken for Congruent section

This did line up with my expectations since evertime i tried to take the test it felt unnatural while doing the Incongruent section and took relatively longer to give my responses

(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!

I think our brain has different processes that are undergone when trying to interpret color versus when trying to interpret text. So in the case of congruent words, any one of the two processes would give the correct answer hence takes less time. Whereas for incongruent words, the processes might seem to clash with each other because we also have to decide on which we need to give more priority to , the color or the text? And since the brain reads words faster than it recognizes colors, it leads to an increase in reaction time.
A similar task can be when asked to perform actions with the non-dominant hand, e.g. cooking or writing. It would take longer for right-handed people to work with their left-hand and vice versa