Test a Perceptual Phenomenon

Course Name: Machine Learning Foundation - Udacity

Project Name: Test a Perceptual Phenomenon

Student Name: Sugam Khetrapal

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.

Questions for Investigation

As a general note, be sure to keep a record of any resources that you use or refer to in the creation of your project. You will need to report your sources as part of the project submission.

Q1. What is our independent variable? What is our dependent variable?

Independent Variable: Congruent words condition, and an incongruent words condition

Dependent Variable: Time taken to name the ink colors

Q2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices

μC: Population Mean of Response Time under Congruent Condition

μI: Population Mean of Response Time under Incongruent Condition

Null Hypothesis (H_0): There is no difference in the population means of response time under Congruent and Incogruent conditions, i.e. H_0: μC = μI

Alternative Hypothesis (H_1): Population Mean of the response time under Incongruent Condition will be significantly larger than the response time under Congruent Condition, i.e. H_1: μC < μI

Since we have reasons to assume the directionality beforehand, this would be a one-tailed test.

Directional Hypothesis (http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/directional-and-nondirectional-hypotheses/ (http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/directional-and-nondirectional-hypotheses/)): A directional alternative hypothesis states that the null hypothesis is wrong, and also specifies whether the true value of the parameter is greater than or less than the reference value specified in null hypothesis.

```
Justification:

Objective: We need to compare the means of two “related groups” to understand
and determine whether the difference between the two means is statistically
significant or not.

Assumption: Normal Distributions and comparison is taking place between
dependent data samples

Sample Size: 24 observations only

Population Standard Deviation is not available
```

Q3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

In [1]:
```
import pandas as pd

df=pd.read_csv('stroopdata.csv')
print(df.head(10))
```

```
   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
```

In [2]:
```
print("Congruent (Mean): ", df['Congruent'].mean())
print("Congruent (Median): ", df['Congruent'].median())
print("Congruent (Mode): ", df['Congruent'].mode())
print("Congruent (Quantile): ", df['Congruent'].quantile())

print("Congruent (Variance): ", df['Congruent'].var())
print("Congruent (Standard Deviation):", df['Congruent'].std())

print("Congruent (Skewness):", df['Congruent'].skew())
print("Congruent (Kurtosis):", df['Congruent'].kurt())
```

```
Congruent (Mean):  14.051125000000004
Congruent (Median):  14.3565
Congruent (Mode):  0      8.630
1      8.987
2      9.401
3      9.564
4     10.639
5     11.344
6     12.079
7     12.130
8     12.238
9     12.369
10    12.944
11    14.233
12    14.480
13    14.669
14    14.692
15    15.073
16    15.298
17    16.004
18    16.791
19    16.929
20    18.200
21    18.495
22    19.710
23    22.328
dtype: float64
Congruent (Quantile):  14.3565
Congruent (Variance):  12.669029070652174
Congruent (Standard Deviation): 3.559357957645195
Congruent (Skewness): 0.41689987447903953
Congruent (Kurtosis): -0.20522482332339598
```

In [3]:
```
print("Incongruent (Mean): ", df['Incongruent'].mean())
print("Incongruent (Median): ", df['Incongruent'].median())
print("Incongruent (Mode): ", df['Incongruent'].mode())
print("Incongruent (Quantile): ", df['Incongruent'].quantile())

print("Incongruent (Variance): ", df['Incongruent'].var())
print("Incongruent (Standard Deviation):", df['Incongruent'].std())

print("Incongruent (Skewness):", df['Incongruent'].skew())
print("Incongruent (Kurtosis):", df['Incongruent'].kurt())
```

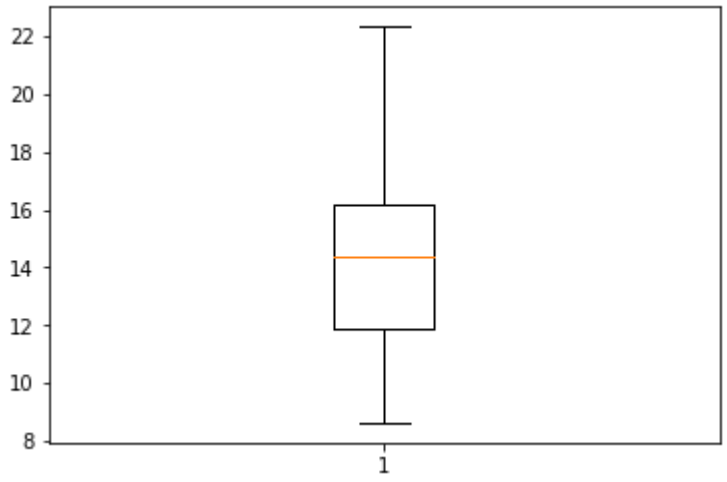
```
Incongruent (Mean):  22.01591666666667
Incongruent (Median):  21.0175
Incongruent (Mode):  0     15.687
1     17.394
2     17.425
3     17.510
4     17.960
5     18.644
6     18.741
7     19.278
8     20.330
9     20.429
10    20.762
11    20.878
12    21.157
13    21.214
14    22.058
15    22.158
16    22.803
17    23.894
18    24.524
19    24.572
20    25.139
21    26.282
22    34.288
23    35.255
dtype: float64
Incongruent (Quantile):  21.0175
Incongruent (Variance):  23.011757036231884
Incongruent (Standard Deviation): 4.797057122469138
Incongruent (Skewness): 1.547590025915552
Incongruent (Kurtosis): 2.6889001984359964
```

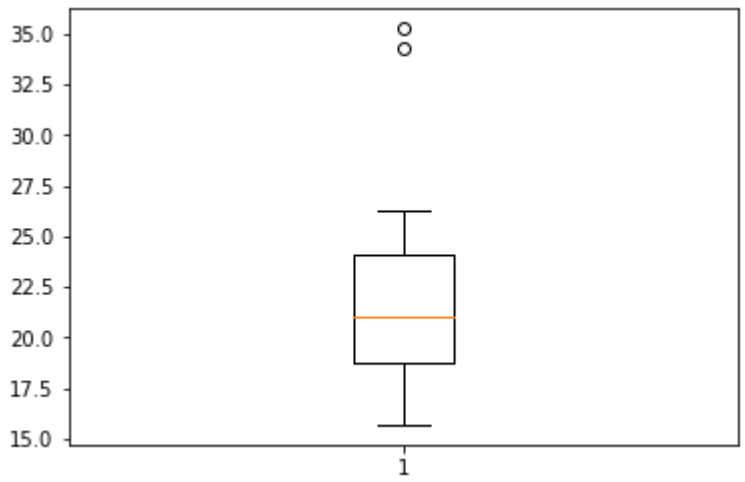
Q4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

In [4]:
```
import matplotlib.pyplot as plt
import numpy as np

plt.plot(df.index, df['Congruent'], df.index, df['Incongruent'], linewidth=2.0)
plt.title('Time taken to identify ink color of Congruent words (blue) vs Incongru
plt.xlabel('Records');
plt.ylabel('Time (seconds)');
print(plt.show())
```

```
<matplotlib.figure.Figure at 0x2250de9c198>

None
```

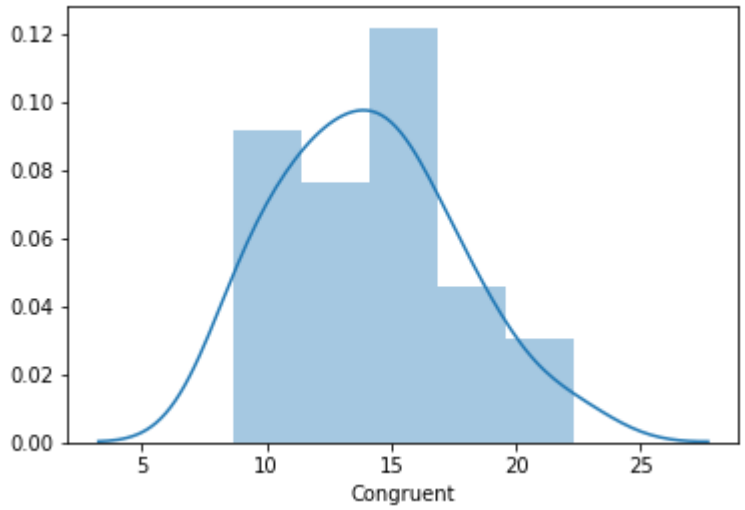
In [5]:
```
plt.figure()
plt.boxplot(df["Congruent"])
plt.show()
plt.figure()
plt.boxplot(df["Incongruent"])
plt.show()
```





In [6]:
```
import seaborn as sns
from scipy.stats import t

sns.distplot(df['Congruent'])
```

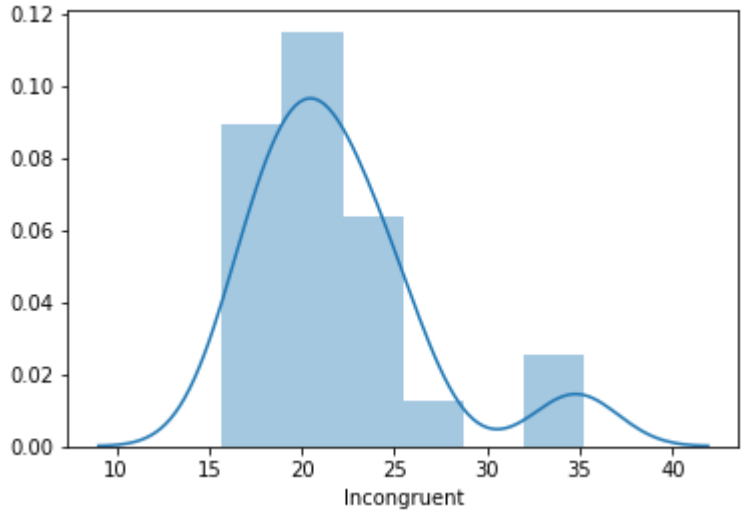
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x22510633748>



(Congruent) Normal distribution is depicted in the above graph.

In [7]:
```
sns.distplot(df['Incongruent'])
```

Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x2251064b400>



(Incongruent) Although the above graph is also Normally distributed, it is skewed towards the Right (i.e. Positively Skewed).

Q5. 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 [8]:
```
print(df['Congruent'].size)
print(df['Incongruent'].size)
```

```
24
24
```

In [9]:
```
#Critical value for 95% confidence level and 23 d.f. (size-1, i.e. 24-1)
t.ppf(.95, 23)
```

Out[9]:
```
1.7138715277470473
```

Difference between Incongruent (Mean) & Congruent (Mean), i.e. 22.01591666666667 - 14.051125000000004 = 7.96479166666667

In [10]:
```
df['DifferenceSD'] = df['Congruent'] - df['Incongruent']
print("Standard Deviation of the Difference: ", df['DifferenceSD'].std(axis=0))
```

```
Standard Deviation of the Difference:  4.864826910359056
```

t-statistic:

Difference of Means / (Standard Deviation of the Difference / Square Root of number of Observations)

In [11]:
```
import math

(22.01591666666667 - 14.051125000000004) / (4.864826910359056 / math.sqrt(24))
```

Out[11]:
```
8.020706944109955
```

Result:

T-statistic (8.020706944109955) > Critical value (1.7138715277470473)

Hence, Null Hypothesis (that there is no difference between the 2 times) can be rejected.

The result is as per my expectation because it took me less time to complete words with Congruent words condition than it took me to complete words with Incongruent words condition.

References:

https://en.wikipedia.org/wiki/Stroop_effect (https://en.wikipedia.org/wiki/Stroop_effect)

https://imotions.com/blog/the-stroop-effect/ (https://imotions.com/blog/the-stroop-effect/)

https://www.youtube.com/watch?v=EGpzftQf8oI (https://www.youtube.com/watch?v=EGpzftQf8oI)

https://en.wikipedia.org/wiki/Stroop_effect#Warped_words (https://en.wikipedia.org/wiki/Stroop_effect#Warped_words)

Q6. 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!

1. On a daily basis, we are accustomed to reading words and not focusing on the ink of the color of the words. In this task, we had to say aloud the color of the ink of the word instead of the word, which was a change from what we have been used to.
2. Our exposure to reading text printed in black color makes us accustomed to recognizing words and not colors. Hence, when confronted with printed text in any color apart from black, our brain spends additional effort in making sure that we focus on the color of the ink instead of the word itself.

As per Wikipedia, "the warped words Stroop effect produces the same findings similar to the original Stroop effect. Much like the Stroop task, the printed word's color is different from the ink color of the word; however, the words are printed in such a way that it is more difficult to read (typically curved-shaped).[36] The idea here is the way the words are printed slows down both the brain's reaction and processing time, making it harder to complete the task."

In [ ]: