Test a Perceptual Phenomenon

February 1, 2019

0.0.1 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, 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 whether the test is congruent or incgruent. The dependent variable is the time taken to say the colors of all the words

(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.

For time in seconds: Null hypothesis: There is no difference in population means of response time under incongruent and congruent scenarios H0: muC = muI where mu is the population mean

Alternative hypothesis: There is a significant difference in population means of response time under incongruent and congruent scenarios H1: muC muI

We choose this becuase our default (null hypothesis) is that we assume there's no significant difference between population means between Congruent and Incongruent and the alternative is that there's a significant difference in means between these two conditions.

(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 [16]: # Perform the analysis here
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
         import math
         from scipy.stats import t
         %matplotlib inline
```

```
In [16]: df = pd.read_csv('stroopdata.csv')
         df.head()
```

Out[16]:

	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 [17]: df.describe()
```

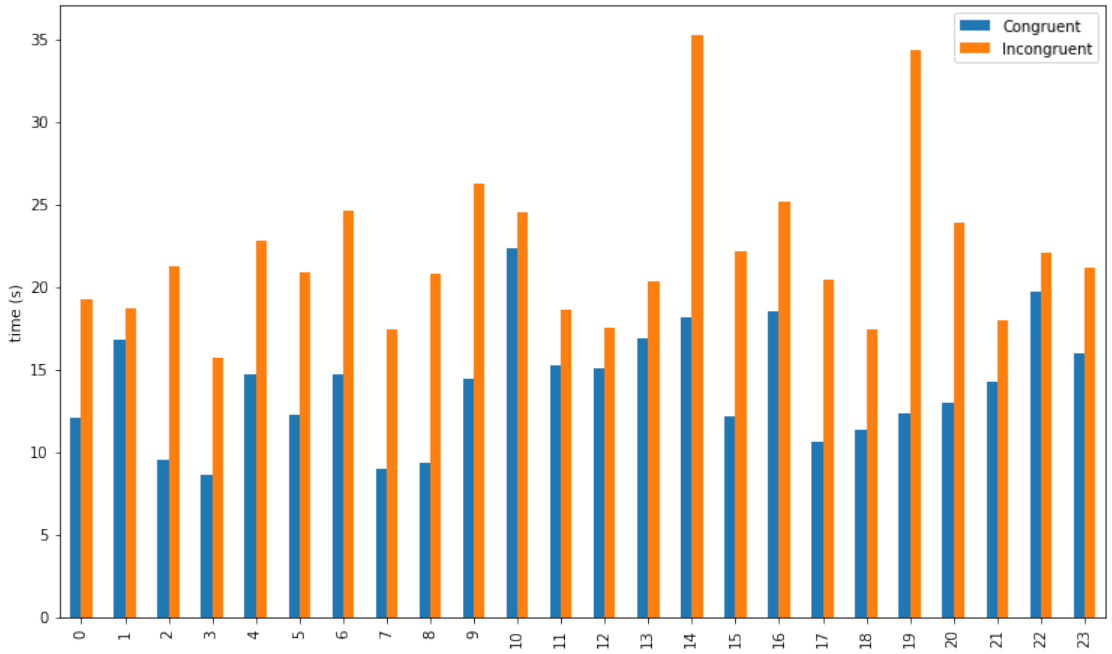
Out[17]:

	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

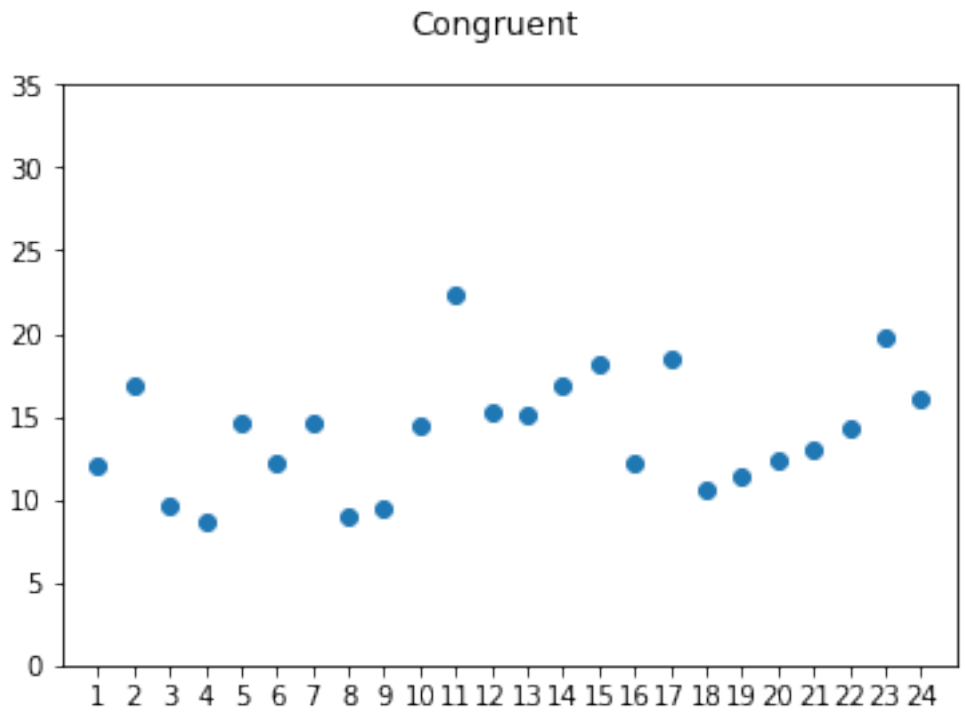
We take the mean as a measure of centeral tendency which is 14.0511 for Congruent and 22.0159 for Incongruent. As for the measure of variability we take the standard deviation which is 3.5593 for Congruent and for Incongruent it is 4.7971

(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 [18]: df.plot.bar(figsize = (12.5,7.5));
         plt.ylabel('time ''(s)');
```

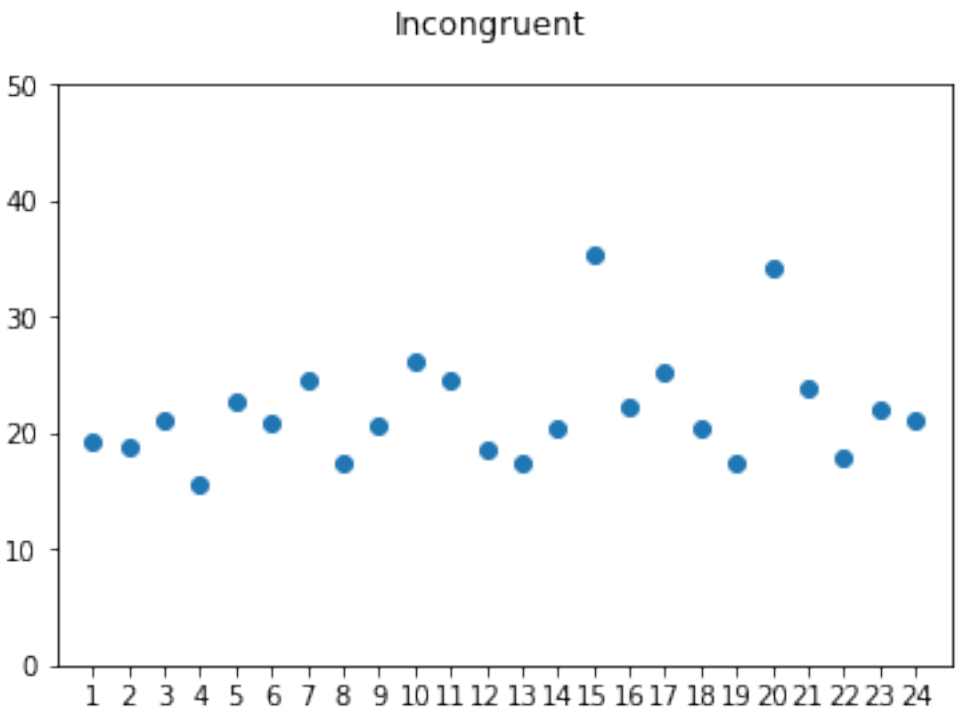


In [31]:
```
N = 24
x = df.index+1
y = df['Congruent']


fig = plt.figure()
fig.suptitle('Congruent')

ax.set_ylabel('time ''(s)')

plt.xticks(np.arange(min(x), max(x)+1, 1.0))
plt.scatter(x, y)
plt.ylim([0,35])
plt.xlim([0,25])
plt.show()
```



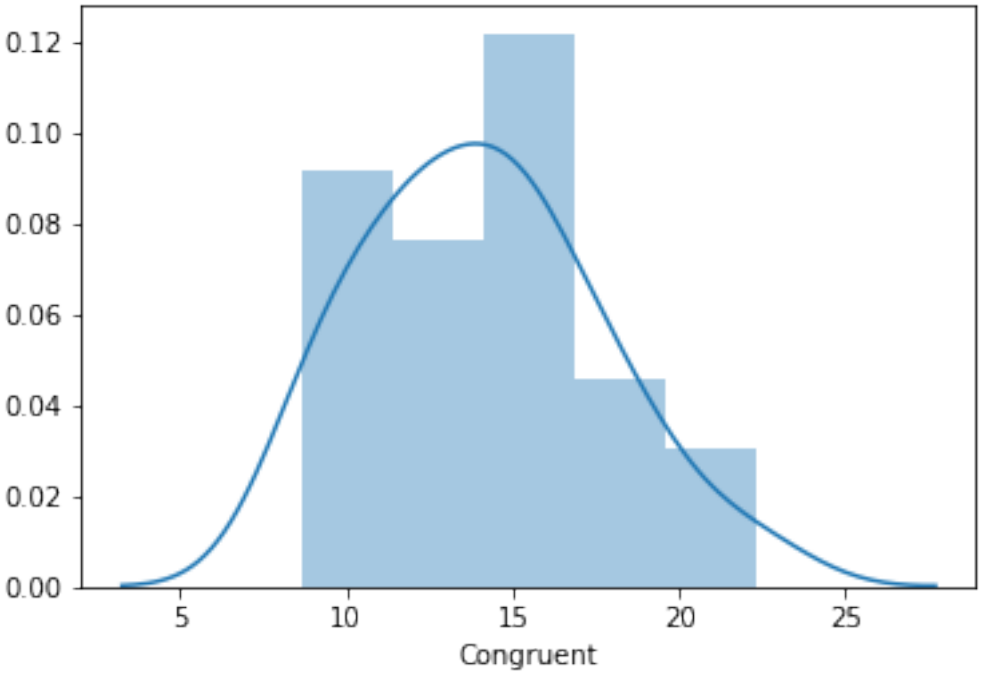
In [33]:
```
N = 24
x = df.index+1
y = df['Incongruent']
```

```
fig = plt.figure()
fig.suptitle('Incongruent')

ax.set_ylabel('time ''(s)')

plt.xticks(np.arange(min(x), max(x)+1, 1.0))
plt.scatter(x, y)
plt.ylim([0,50])
plt.xlim([0,25])
plt.show()
```

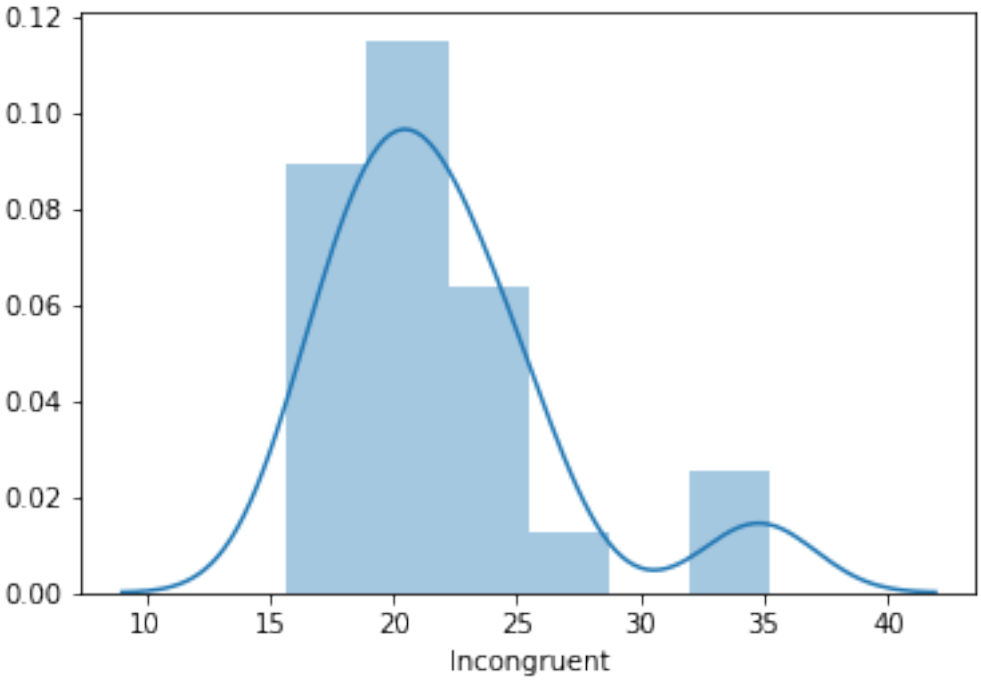


We notice that Incongruent time is always more than congruent time. The biggest difference between an incongruent and congruent is around 25 seconds.

```
In [20]: sns.distplot(df['Congruent']);
```



In [21]:
```
sns.distplot(df['Incongruent']);
```



From the last 2 plots, we can say that they are normally distributed and the peak for the 2 graphs is similar to the mean which we calculated earlier in Question 3.

(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.

Because of the following conditions, I choose to use t-test instead of the other testing methods like z-test: - we're comparing between 2 samples. - Population standard deviation is not given. - the sample size is n<30.

The confidence level will be 95%

```
In [22]: # Perform the statistical test here
         # Since they're normally distibuted, I'll perform Two Sample T-test
         pincongruent = df['Incongruent'].mean()
         pcongruent = df['Congruent'].mean()

         n = len(df)

         diff =  pcongruent - pincongruent

         # in order to find the standard deviation of this sample data set, we create a new colu

         df['diff'] =  df['Congruent'] - df['Incongruent']
         df['diff']

         #df2.plot()
```

```
Out[22]: 0     -7.199
         1     -1.950
         2    -11.650
         3     -7.057
         4     -8.134
         5     -8.640
         6     -9.880
         7     -8.407
         8    -11.361
         9    -11.802
         10    -2.196
         11    -3.346
         12    -2.437
         13    -3.401
         14   -17.055
         15   -10.028
         16    -6.644
         17    -9.790
         18    -6.081
```

```
19   -21.919
20   -10.950
21    -3.727
22    -2.348
23    -5.153
Name: diff, dtype: float64
```

```
In [23]: # calculating the standard deviation of the difference between Congruent and Incongruen
         df['diff'].std()
```

```
Out[23]: 4.8648269103590556
```

```
In [24]: # now we calculate the standard error of mean
         df['diff'].std() / math.sqrt(24)
```

```
Out[24]: 0.99302863477834058
```

```
In [25]: # now we obtain the t-value
         diff / (df['diff'].std() / math.sqrt(24))
```

```
Out[25]: -8.020706944109957
```

```
In [26]: # now we calculate t-critical value at 95% confidence level
         tcv = t.ppf(.975,23)
         tcv
         #since it's two sided, the confidence level will be on both sides
         # 23 is the number representing our degree of freedom which is n-1
```

```
Out[26]: 2.0686576104190406
```

```
In [27]: diff
```

```
Out[27]: -7.9647916666666667
```

```
In [28]: print(diff - tcv, diff + tcv)
```

```
-10.0334492771 -5.89613405625
```

t-value: -8.020706944109957 t-crirical value: (2.0686576104190406, - 2.0686576104190406) the mean of the difference between Congruent and Incongruent: -7.9647916666666667 Confidance Interval(CI): (-10.0334492771, -5.89613405625)

From the above statistics we can see that t-value is less than t-critical value. Therefore, we reject the null hypothesis

Refrences: https://www.isixsigma.com/tools-templates/hypothesis-testing/making-sense-two-sample-t-test/ https://study.com/academy/lesson/z-test-t-test-similarities-differences.html https://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf