

Input:

```
def filenames(filename):  
    df = pandas.read_csv(filename)  
    congruent = df['Congruent']  
    df['Sample Number'] = df.index+1  
    print df
```

Output:

Congruent	Incongruent	Sample Number
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	24

Fig 1.1: CSV data file in panda dataframe

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

Independent variables are Congruent and Incongruent. The dependent variables is Time (in sec)

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

Hypotheses Test

H_0 (Null Hypothesis) : There is no significant difference in population average time of congruent word and incongruent word.

H_a (Alternative Hypothesis): There is significant difference in population average time.

Statistical Test

Input: Normal Distribution Plot (Congruent)

```
def plot_normal_distribution_congruent(filename):  
    df = pandas.read_csv(filename)  
    congruent = df['Congruent']  
    congruent_mean = congruent.mean()  
    congruent_std = congruent.std()  
    range = np.arange(0, 30, 0.5)  
    plot.figure().suptitle('Congruent: ', fontsize=20, fontweight='bold')  
    plot.plot(range, mlab.normpdf(range, congruent_mean, congruent_std))
```

Output:

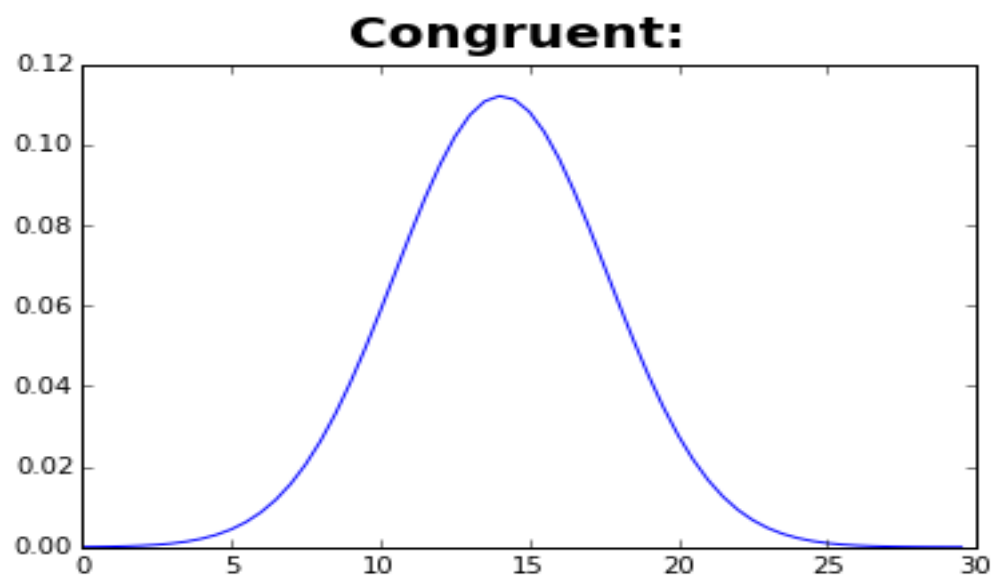


Fig 1.2: Normal Distribution curve for Congruent

Input: Normal Distribution Plot (Incongruent)

```
def plot_normal_distribution_incongruent(filename):  
    df = pandas.read_csv(filename)  
    incongruent = df['Incongruent']  
    incongruent_mean = incongruent.mean()  
    incongruent_std = incongruent.std()  
    range = np.arange(0, 45, 0.5)  
    plot.figure().suptitle('Incongruent: ', fontsize=20, fontweight='bold')  
    plot.plot(range, mlab.normpdf(range,incongruent_mean,incongruent_std))
```

Output:

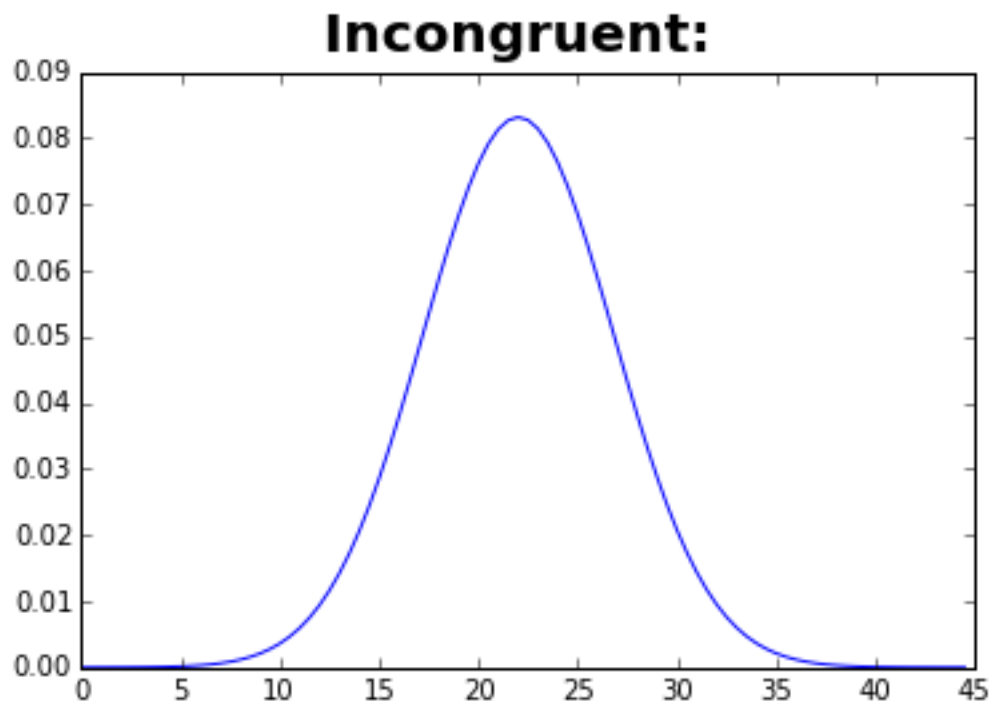


Fig 1.3: Normal Distribution curve for incongruent

As shown in the Figure 1.2 and 1.3 the data is Normally Distributed.

I used the two tailed dependent t-test because as the:-

- 1) $n < 30$
- 2) The test is done on same subject two times in different conditions. After performing the first test, subject will gain some experience and able to memorize this. He/she can use this unfair advantage in his/her second test. So, because of this logic I concluded the two tests are dependent on each other.
- 3) We don't have any population parameters, in this case the t-test is more valid.
- 4) I used the two tail test because as the one tail test prefer to test the effects of one tail only, and only consider one set of scenario. The two tail instead of this take care of all scenarios.

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

Congruent

Input :

```
congurent = df['Congruent']  
  
congurent_mean = congurent.mean()  
  
print congurent_mean
```

Output (mean): 14.051125

Input:

```
congurent = df['Congruent']  
  
congurent_std = congurent.std()  
  
print congurent_std
```

Output (standard deviation): 3.55935795765

Incongruent

Input:

```
incongurent = df['Incongruent']  
  
incongurent_mean = incongurent.mean()  
  
print incongurent_mean
```

Output (mean): 22.0159166667

Input:

```
incongurent = df['Incongruent']  
  
incongurent_std = incongurent.std()  
  
print incongurent_std
```

Output (standard deviation): 4.79705712247

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.

Input:

```
def plot_concurrent(filename):  
    df = pandas.read_csv(filename)  
    df['Sample Number'] = df.index+1  
    x = df['Sample Number']  
    y = df['Congruent']  
    colors = 'yellow'  
    area = np.pi*20  
    fig = plot.figure()  
    fig.suptitle('Congruent: ', fontsize=20, fontweight='bold')  
    ax = fig.add_subplot(110)  
    fig.subplots_adjust(top=0.85)  
    ax.set_xlabel('Sample Number')  
    ax.set_ylabel('Completion Time (in sec)')  
    plot.xticks(np.arange(min(x), max(x)+1, 1.0))  
    plot.yticks(np.arange(0, 35, 2))  
    plot.scatter(x, y, s=area, c=colors, alpha=1)  
    plot.ylim([0,35])  
    plot.xlim([0,24])  
    plot.show()
```

Output:

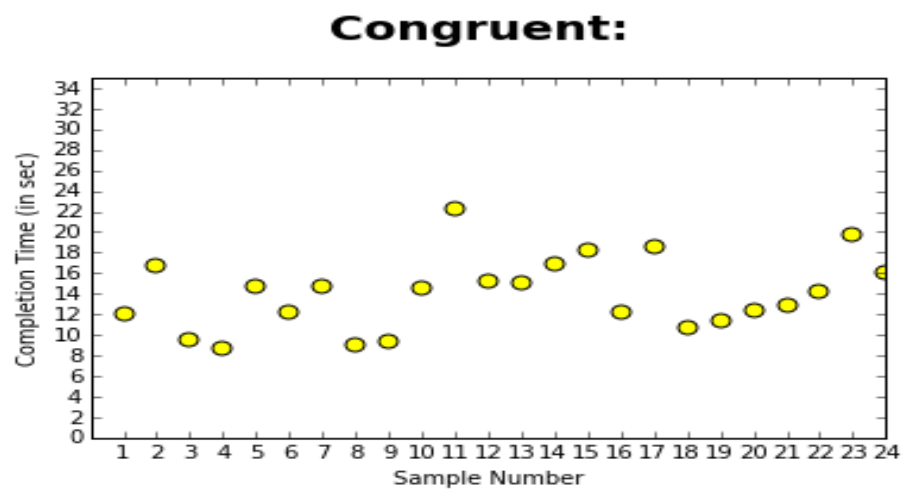


Fig 1.4: Response Time Plot of Congruent

Input:

```
def plot_incongruent(filename):  
  
    df = pandas.read_csv(filename)  
  
    df['Sample Number'] = df.index+1  
  
    x = df['Sample Number']  
  
    y = df['Incongruent']  
  
    colors = 'green'  
  
    area = np.pi*20  
  
    fig = plot.figure()  
  
    fig.suptitle('Incongruent: ', fontsize=20, fontweight='bold')  
  
    ax = fig.add_subplot(110)  
  
    fig.subplots_adjust(top=0.85)  
  
    ax.set_xlabel('Sample Number')  
  
    ax.set_ylabel('Completion Time (in sec)')  
  
    plot.xticks(np.arange(min(x), max(x)+1, 1.0))  
  
    plot.yticks(np.arange(0, 35, 2))  
  
    plot.scatter(x, y, s=area, c=colors, alpha=1)  
  
    plot.ylim([0,35])  
  
    plot.xlim([0,24])  
  
    plot.show()
```

Output:

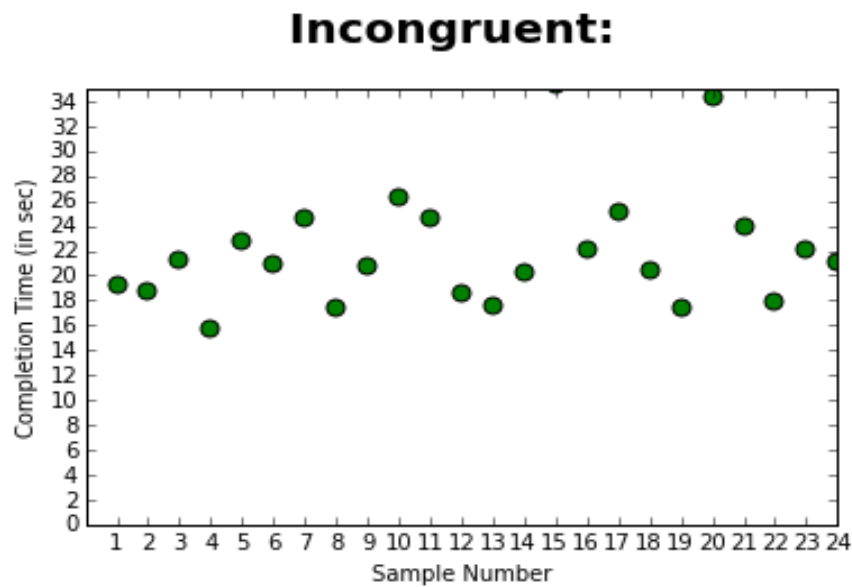


Fig 1.5: Response Time plot of Incongruent

In the Fig 1.4 the response time (in sec) is scattered from 8 to 20 sec and one sample at 22 sec. And in the Fig 1.5 the response time (in sec) is in between 15 to 22 sec and one sample at 34. As shown obviously in graph the average time of Incongruent is definitely higher than of Congruent.

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?

Confidence Level = 90%

$df = n - 1 = 24 - 1 = 23$

T-Critical Value = (+/-) 1.714

Input :

```
def ttest(filename):  
    df = pandas.read_csv(filename)  
    congruent = df['Congruent']  
    incongruent = df['Incongruent']  
    result = scipy.stats.ttest_rel(incongruent, congruent)  
    print result
```


Output: (8.020706944109957, 4.1030005857111781e-08)

Observation: T-stat = 8.020706944109957

P= 4.1030005857111781e-08

T-stat > T- Critical

Null hypothesis REJECTED. The time of response in Congruent and Incongruent is different significantly. The speed is slower in Incongruent than in Congruent as previously stated in above questions. Yes the result match with the expectations.

REFERENCES

https://en.wikipedia.org/wiki/Null_hypothesis

https://en.wikipedia.org/wiki/Alternative_hypothesis

https://en.wikipedia.org/wiki/Normal_distribution

http://matplotlib.org/api/mlab_api.html

http://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ttest_rel.html

<http://conversionxl.com/one-tailed-vs-two-tailed-tests/>