Statistics: The Science of Decisions Project Instructions

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.

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

ANS : independent variable : congruent/incongruent words condition. dependent variable : the time it takes to name the ink colors in equally-sized lists.

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

Now it's your chance to try out the Stroop task for yourself. Go to thislink (https://faculty.washington.edu/chudler/java/ready.html), which has a Java-based applet for performing the Stroop task. Record the times that you received on the task (you do not need to submit your times to the site.) Now, download this dataset (https://drive.google.com/file/d/0B9Yf01UaIbUgQXpYb2NhZ29yX1U/view) which contains results from a number of participants in the task. Each row of the dataset contains the performance for one participant, with the first number their results on the congruent task and the second number their performance on the incongruent task.

2.1 What is an appropriate set of hypotheses for this task?

ANS: The hypothesis is "Incogurent words condition make people take a longer time to say the correct words than the congruent words condition."

2.2 What kind of statistical test do you expect to perform?

ANS: Let participant perform experiment with 2 conditions and measure the time. 1st condition, Say out loud the color of the ink in which the word is printed for 20 different words with a congruent words condition. 2nd condition, Say out loud the color of the ink in which the word is printed for 20 different words with an incongruent words condition. If the 2nd condition take more time then the hypothesis is correct.

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

```
In [8]:  import pandas as pd
         import os
         filename = r'~/WorkSpace/Udemy-Datascience/IntroductionToDataScience/P
         dataFrame = pd.read_csv(filename)
```

```
In [10]:  dataFrame['Congruent'].mean()
```

Out[10]: 14.051125000000004

```
In [13]:  dataFrame['Congruent'].std()
```

Out[13]: 3.559357957645195

```
In [9]:  dataFrame['Incongruent'].mean()
```

Out[9]: 22.01591666666667

```
In [14]:  dataFrame['Incongruent'].std()
```

Out[14]: 4.797057122469138

ANS:

```
    In experiment congruent condition:
    Average time it takes is 14.05.
    Unbiased standard deviation is 3.56.

    In experiment incongruent condition:
    Average time it takes of experiment incongruent condition is 2
    2.02.
    Unbiased standard deviation is 4.80.
```

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.

ANS:

```
In [15]:
# Causes plots to display within the notebook rather than in a new win
%pylab inline
import matplotlib.pyplot as plt

df_compare = pd.DataFrame({'Congruent': dataFrame['Congruent'],
                'Incongruent': dataFrame['Incongruent']})
df_compare.plot(kind='hist', alpha=0.2,bins = range(0, 37 + 3, 3))

plt.xlabel("time")
plt.ylabel("frequency")
plt.show()
```

Populating the interactive namespace from numpy and matplotlib
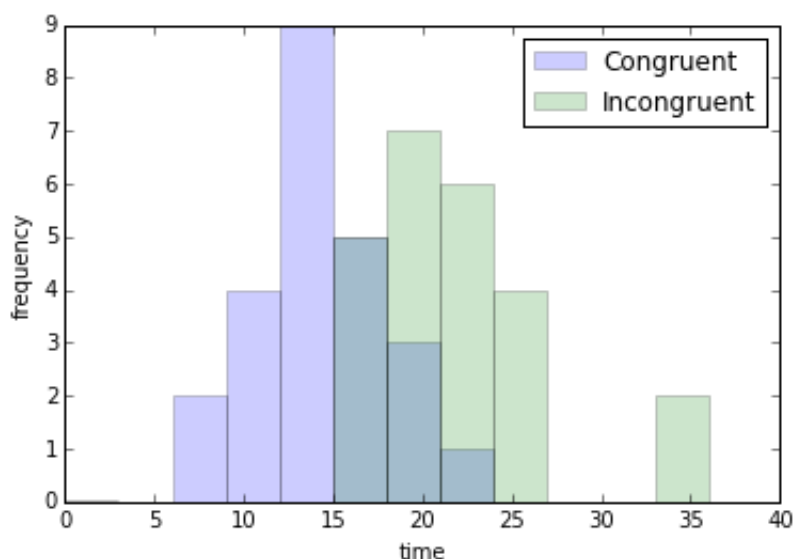
As the histrogram show the congruent condition is normal distribution but the incongruent condition is skewed right distribution.

```
In [ ]:
```

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?

ANS

```
In [24]:  import scipy
          import scipy.stats
          U,p = scipy.stats.mannwhitneyu(dataFrame['Incongruent'],dataFrame['Con
          p_critical = 0.001
          reject_result = p > p_critical

          print "U : %s " %U
          print "statistically highly significant p critical value : %s " %p_cri
          print "reject_result : %s (p = %.10f)" %(reject_result,p)
```

```
U : 37.0
statistically highly significant p critical value : 0.001
reject_result : False (p = 0.0000001201)
```

I calculated a p-value of 0.0000001201. This p-value is less than the stated highly significance level of 0.001. Therefore, we reject the null hypothesis.

We can conclude that the conditions(incongruent) statistically affect how long it take to say out loud the color of the ink in which the word is printed.

Did the results match up with your expectations? YES

(ref : http://www.statsdirect.com/help/default.htm#basics/p_values.htm (http://www.statsdirect.com/help/default.htm#basics/p_values.htm))

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 incongruent conditions is reponsible for the experiment.

We can reverse the experiment by say the words instead,the trend would be the same.