Project1 Statistics

August 2, 2017

1 Statistics: The Science of Decisions Project Instructions

1.0.1 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.

1.0.2 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.

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

Answer: > The independent variable is the congruent words and incongruent words.
> The depedent variable is the time it takes to name the ink colors.

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

Answer: > We have two word sets for test: congrent words set and incongruent words set. Both sets have the equal number of words. We need to find out that for the experimental population group, if the same person in group will have the same response time to the two tested word sets. We are using a sample in order to make inferences about the population, that is unknown, to find whether there is a difference between congruent and incongruent means in the population. We are using the sample to infer something about the population.

The appropriate set of hyposthesis is composed by Null hypothesis and Alternative hypothesis.

Null Hypothesis H_0: There is no difference in population means of response time under incongruent and congruent scenarios, which can be mathematically represented as

$$\mu_C = \mu_I.$$

The alternative H_A: Hypothesis is the response time for the same group people on congrent words set and in congruent word set are different, represented as

$$\mu_C \neq \mu_I.$$

We are proposed to use t-test instead of a z-test for example. There are a few reasons for our choice. 1. Our sample size in data sets is less than 30 smaples. 2. We don't know the distributions of the population. 3. We don't know the population standard deviation.

So it is better to choose t-test for our data set.

Now it's your chance to try out the Stroop task for yourself. Go to this link, 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 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.

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

Answer: > The measure of central tendency is the average of data sets, or called the mean of data sets. The measure of variability of data sets could be variance, standard deviation, or standard error. > We first calculate the measurement on the two experimental data sets. The mean of congrent words experimental data sets is $\overline{X_C} = 14.051125$, the standard deviation is $S_C = 3.559357958$; The mean of incongrent words experimental data sets is $\overline{X_I} = 22.01591667$, the standard deviation is $S_I = 4.797057122$; The sample size in each experiement is $n_C = n_I = 24$.

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.
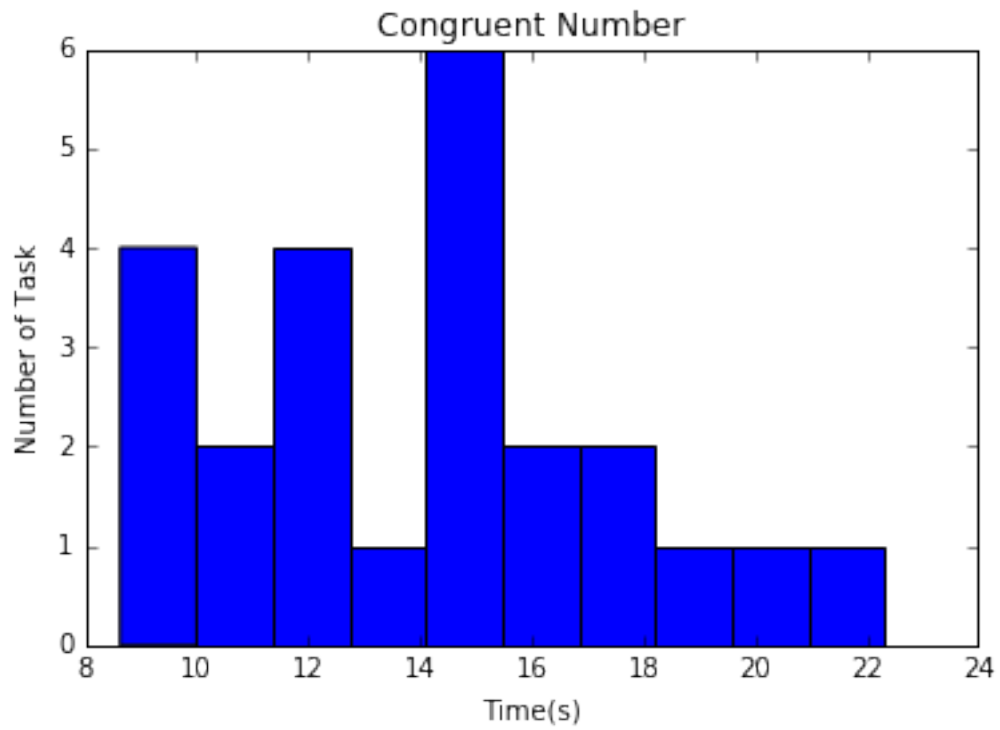
```
In [18]: import pandas as pd

         path = 'stroopdata.csv'
         df = pd.read_csv(path)

In [22]: %pylab inline

         import matplotlib.pyplot as plt

         Congruent_data = df['Congruent']
         plt.hist(Congruent_data, bins=10)
         plt.xlabel("Time(s)")
         plt.ylabel("Number of Task")
         plt.title("Congruent Number")
         plt.show()

Populating the interactive namespace from numpy and matplotlib
```
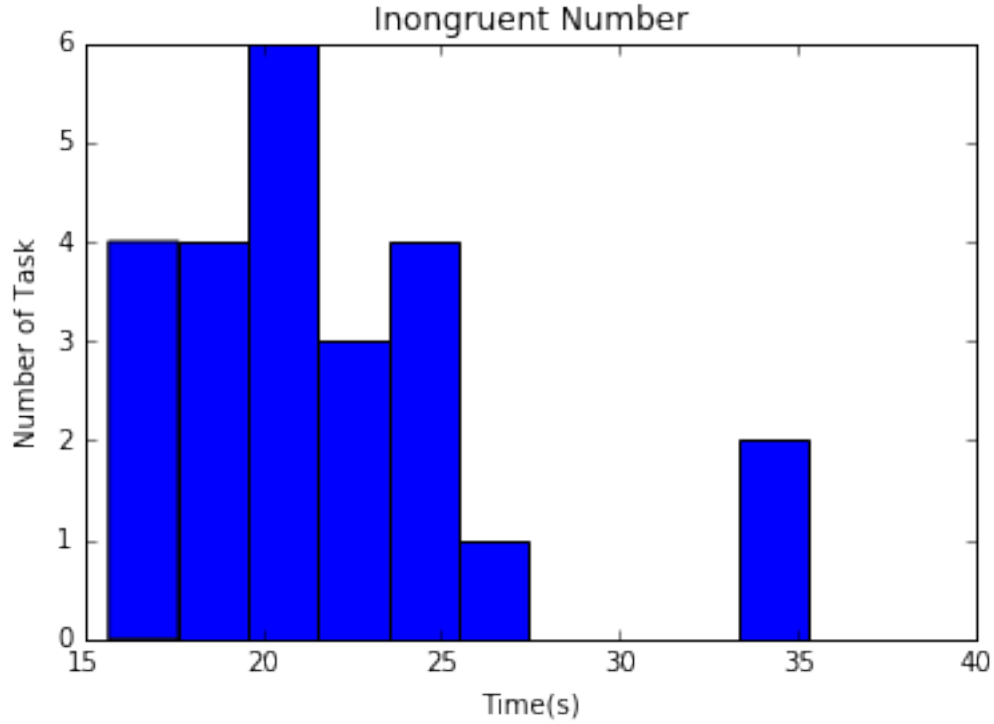
Congruent Number

```
In [24]: Incongruent_data = df['Incongruent']
         plt.hist(Incongruent_data, bins = 10)
         plt.xlabel("Time(s)")
         plt.ylabel("Number of Task")
         plt.title("Inongruent Number")
         plt.show()
```

Answer:

The shape of congrent words experimental data sets and the shape of incongrent words experimental data sets (Table1 and Table2)looks in a similar shape. Both data sets is in the shape of skewed normal distribution.

The mean, median of congrent words experimental data sets (Table1) is lower than the mean, median of incongrent words experimental data sets (Table2).

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?

Answer: > The standard deviation difference is $S_{CI} = 4.86482691$; The t-statistic value is

$$t = \frac{\overline{X_C} - \overline{X_I}}{\frac{S_{CI}}{\sqrt{n}}} = -8.02092.$$

We then set the confidence interval of our experiment is 95%, and the $\alpha = 0.05$. The freedom degree $n_c + n_I - 2 = 46$. After checking the t-table, we find the t critical statistic value with is around $t_{critical} = -2.009$.

Finally, we compare the t-statistic value with the t-critical value. Since $t = -8.02092 < t_{critical} = -2.009$, we reject the Null hypothesis H_0 and accept the alternative hypothesis H_A, that is the response time on congrent words set and congruent words set are different.

4

To conclude the experiment, we are sure that the response time on congrent words set is lower than the congruent words set in 95% confidency. This result match up my own experiment result. My own test on congruent words cost 15.745 seconds, while the incongruent words cost 43.736 seconds.

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!
Answer: