Test a Perceptual Phenomenon

Udacity Data Analyst Nanodegree - Project 1

Diego Menin – December 2015

UDACITY

Project Overview

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

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

- Dependent Variable: Time it took for the user to name all the colors
- Independent Variable: Which condition (congruent/incongruent)

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

- **Null Hypothesis (Ho)**: There is no significant difference in the average time taken to name the colors when viewing the congruent (μc) dataset vs viewing the incongruent (μi) dataset (μi - μc = 0).
- **Alternative Hypothesis (Ha)**: There is a significant difference, positive or negative, in the average time taken to name the colors when viewing the congruent dataset (μc) vs viewing the incongruent (μi) dataset (μi - μc ≠ 0).

Give that:

- we don't have population data (can't use a z-test);
- the sample size is less than 30;
- we can assume the distribution is normal;
-

I'll use a two sample t-test. It is pretty clear that the incongruent test is more complex than the congruent test so a one-tail test would suffice but just to be on the safe side, I'll run a two-tailed test.

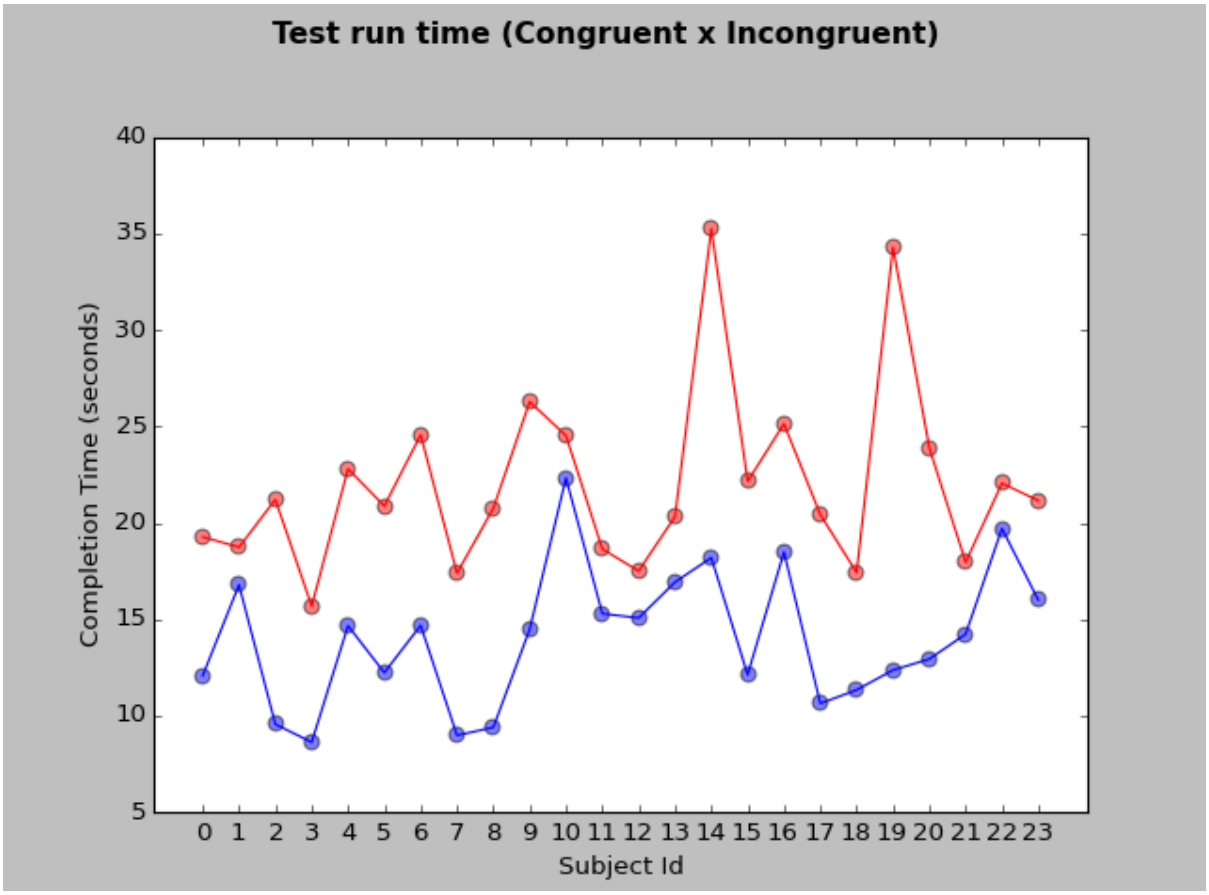
3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

	Min Value	Max Value	Mean	Sample Std
Congruent	8.63	22.32	14.05	3.55
Incongruent	15.687	35.25	22.01	4.79

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.

On the graph bellow I show a scatterplot of each test taken (called Index - x axis) and the time in seconds the subject took to finish it (y –axis). The congruent time is represented in blue and the incongruent time represented in red. I added a "line" between the points to easily see how they relate to each other.

Please note that line connecting the data points has been added only to clearly show that all the Incongruent data points are above the congruent data points, It isn't trying to imply a sequential relationship between participants, because participants are independent of each other - the performance of one participant should not imply anything specifically about the performance of the next.



Looking at the graph we can easily see that all incongruent tests took longer than the congruent ones (which is expected due to the increase in test difficulty).

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?

(See attached Jupiter notebook for more details on the calculations)

1) Find the Point estimate, the difference in the means of “c” and “I”:
   pe = 14.05 - 22.02
   pe = -7.969

2) Find the Standard deviation of the differences:
   S = 4.8648

3) Find t:
   t = pe / (S/np.sqrt(n))
   t = -8.0259

4) Find t-critical at alpha = 0.05:
   tcrit = -2.06865; 2.06865

5) Find the p-value:
   pval = stats.t.sf(np.abs(t), n-1)*2
   pval = 4.0571738513915466e-08

6) Find the Confidence Interval:
   lowerCI = -10.0242
   upperCI = -5.9157

Given that the t-value (-8) falls outside the t-critical range (-2.06 to 2.06) we can reject the null hypothesis that says that there is no significant difference in the time taken to name the colors when viewing the congruent dataset vs viewing the incongruent dataset.

Helpful links:

- http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/null-and-alternative-hypotheses/
- http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/what-is-a-hypothesis-test/
- http://www.statisticshowto.com/when-to-use-a-t-score-vs-z-score/
- http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/tests-of-means/types-of-t-tests/