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.

Dataset

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.

Questions for Investigation

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

Independent variable: the congruency between the name and the colour of the "colour words".

Dependent variable: the time it takes to name the colours 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.

Statistical test: Two-tailed dependent samples t-test. Each participant will go through and record a time from each condition.

In this case, because the amount of time taken in each condition is measured from the same person going through both conditions, so the appropriate statistical test is a dependent samples t-test.

Two-tailed t-test is chosen because we want to know whether any of the words condition affecting the amount of time taken in naming the colours.

Justification and Assumption in choosing t-test:

- The dependent variable is measured on a continuous scale.
- The population standard deviation is unknown. As such, t-test is chosen.
- The sample size is small. In this case, the sample size is less than 30.
- Distribution of the differences in the dependent variable between the two related groups is assumed to be normally distributed.
- The samples or the data sets used to produce the difference scores are linked in the population through repeated measurement.
- The sample is obtained from random sampling of a defined population

Let:

μ_C be the population mean of the amount of time taken in naming the colours for the congruent words condition.

μ_i be the population mean of the amount of time taken in naming the colours for incongruent words condition.

H_o: $\mu_C - \mu_i = 0 \ OR \ \mu_D = 0$. The null hypothesis is that there is no statistical difference between the amount of time taken in naming colour words correctly in congruent words condition as compared to doing so in the incongruent words condition.

H_A: $\mu_C - \mu_i \neq 0 \ OR \ \mu_D \neq 0$. The alternative hypothesis is that there is statistical difference between the amount of time taken in naming colour words correctly in congruent words condition as compared to doing so in the incongruent words condition.

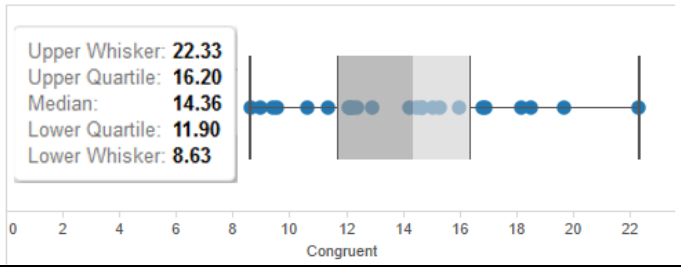
3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

	Congruent Words Condition	Incongruent Words Condition
Median	14.3565	21.0175
Mean	14.0511	22.0159
Standard Deviation	3.5594	4.7971

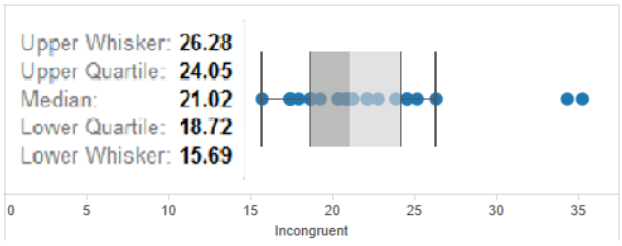
From the measure of central tendency and variability shown above, we can see that the mean and the median amount of time taken in congruent words condition is less than incongruent words condition. The standard deviation in the congruent words condition is also around 1 second less than the incongruent words condition. These results may give an early suggestion that from the sample given, it is faster to name colours in congruent words condition than in incongruent words condition.

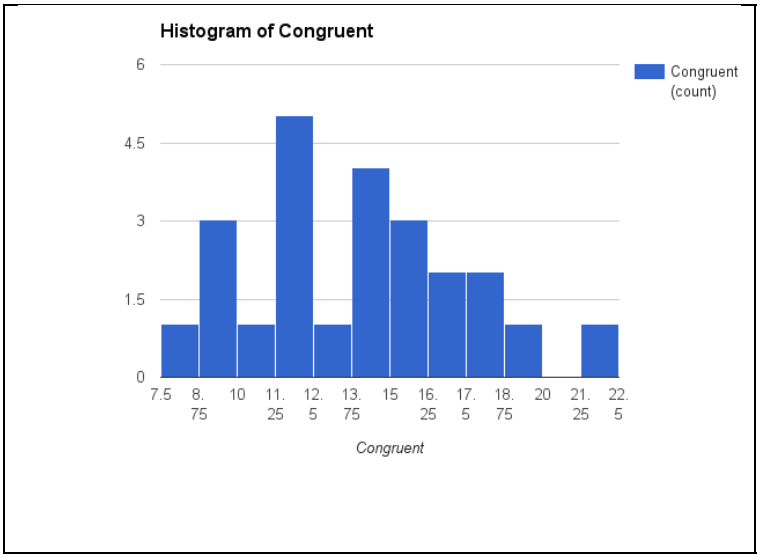
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

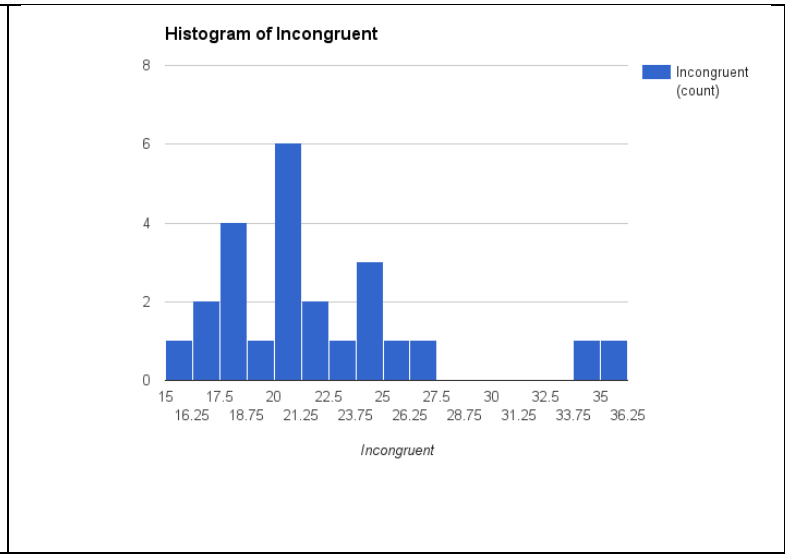
Boxplot of Congruent



Boxplot of Incongruent







From both boxplots shown above, for congruent words conditions, most of the samples have the amount of time taken that is less than that of incongruent words condition. This is shown by:

- Congruent: Q1 (11.90) to Q3 (16.20) as compared to Incongruent: Q1 (18.72) to Q3 (24.05)

From both histograms shown above, for congruent words condition, participants take less time as compared to that of incongruent words condition. This is shown by:

- The starting point of the histogram in Incongruent condition is from 15 seconds as compared to that of the congruent condition, which starts from 7.5 seconds.
- The range of time taken in the histogram for incongruent condition spans from 15 seconds to 36.25 seconds, while for congruent condition, it is only 7.5 seconds to 22.5 seconds.

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 = 0.95; two-tailed test; α = 0.025; n = 24; df = 23

M_D = -7.9648

SD of differences = 4.8648

SEM = $\frac{s}{\sqrt{n}}$

= $\frac{4.8648}{\sqrt{24}}$ = 0.9930

t-statistics = $\frac{xbar\ D - 0}{SEM}$

= $\frac{-7.9648}{0.9930}$ = - 8.0209

t-critical = ± 2.069

Based on the t-statistics and t-critical value, we can **reject** the null hypothesis at a confidence level of 0.95. This means that participants take less time to name a colour correctly when it is in a congruent words condition as compared to an incongruent words condition.

Confidence Interval = M_D ± t-critical $(\frac{s}{\sqrt{n}})$

= $(-7.9648 - 2.069\ (\frac{4.8648}{\sqrt{24}})\ ,\ -7.9648 + 2.069\ (\frac{4.8648}{\sqrt{24}})\ )$

= $(-7.9648 - 2.0545, -7.9648 + 2.0545)$

= (-10.0193, -5.9103)

Based on the confidence interval result, we can say that in average, participants in a congruent word condition will be 5 to 10 seconds faster in naming colours correctly as compared to naming the colours in an incongruent word condition.

The experiment result matched with my expectation as the initial exploratory findings have already shown that participants name colour correctly in a shorter amount of time in congruent word condition as compared to incongruent word condition.

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!

The hypotheses regarding reasons[1] for the effects observed are:

- Processing Speed. There is a delay in the brain's ability to recognize the colour of the word since the brain reads words faster than it recognizes colours.
- Selective Attention. Colour recognition as opposed to reading a word, requires more attention, the brain needs to use more attention to recognize a colour than to word encoding, so it takes a little longer.
- Automaticity. Recognizing colours is not an “automatic process”. There is hesitancy to respond. However, our brain automatically understands the meaning of words because of habitual reading.
- Parallel Distributed Processing. As the brain analyses information, different and specific pathways are developed for different tasks. Some pathways, such as reading, are stronger than others, therefore, it is the strength of the pathway and not the speed of the pathway that is important.

Similar task resulting in a similar effect:

- Comparing digits in incongruent trials (e.g., 3 5) is slower than comparing digits in congruent trials (e.g., 5 3) .

[1] https://en.wikipedia.org/wiki/Stroop_effect