DATA ANALYST NANODEGREE

PART 2 – STATISTICS

PROJECT: THE STROOP EFFECT

Florence Hervé

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.

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?
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 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.
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.
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?
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!

ANSWERS

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

Dependent variable (what is tested, measured): the time taken, in seconds, for the user to read the colour list.

Independent variable (what affects the result, differ from one sample to another): the type of task asked from the user (congruent – the words on the page are aligned with their colour; or incongruent – the words and the colours are not aligned).

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

a. *Hypotheses*

The research question we are trying to answer here is whether the time taken by a user to read the list is impacted by the type of task in the experiment – congruent or incongruent. By performing hypotheses testing using the results from our dataset, we will try to draw a conclusion that goes beyond our samples – that is, whether the difference we may notice in our samples extend to the whole population.

We will perform a t-test (please see section b. for justification); that means we will perform a **test of means** on our samples X in order to prove our hypothesis about the population U we do not know about.

Our hypotheses would be as follow:

H0: the true (population) mean difference between the two conditions **is** zero
H1: the true (population) mean difference between the two conditions **is not** zero

In mathematical terms, we can rephrase our hypotheses as following:

$$\begin{cases} H0: \mu_{congruent} - \mu_{incongruent} = 0 \\ H1: \mu_{congruent} - \mu_{incongruent} \neq 0 \end{cases}$$

With $\mu_{congruent}$ and $\mu_{incongruent}$ being the congruent and incongruent population means, respectively.

Under H0, all observable differences between the two sets of results are explained by random variation. The alternative hypothesis H1 states otherwise.
Because we are considering all directions of the difference, if any, we will be performing a **two-tailed test** here.

b. Statistical test

As seen above, we are considering a comparison of means here. There are two main types of tests of means: z-tests and t-tests. Choosing one or the other category depends on the characteristics of our experiment:

- Our sample size is 24 (which is < 30)
- We do not know the population parameters (mean and standard deviation)

Those two characteristics call for a **t-test**.

Now, we need to decide which type of t-test to perform. Potential t-test types include 1-sample t-tests, 2-sample t-tests, or paired t-test. Here, our experiment is performed on the same sample of users, similarly to a "before-after" experiment, and our datasets are not independent. This is a case of **repeated measures** data (meaning that the dependent variable is measured more than once for each subjects).

Therefore, we need to use a **dependent t-test** (or paired t-test). We will compare the means of the two related samples and evaluate if there is a statistically significant difference in between those means. As noted in the section above, we will perform a **two-tailed** test here, and choose a level of significance $\alpha = 0.05$

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

The dataset contains 24 observations for each test (n=24). Based on this, we can compute some measures of central tendency and measures of spread:

	Measures of central tendency	
	Congruent	*Incongruent*
Mean	14,05	22,02
*Mode**	12	21
Median	14,36	21,02

**rounded up seconds*

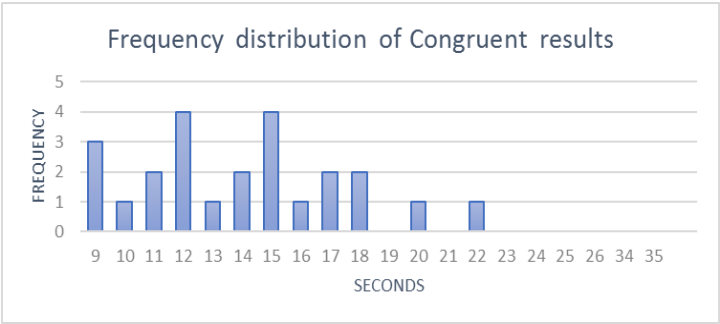
	Measures of spread	
	Congruent	*Incongruent*
Variance	12,14	22,05
Standard Deviation	3,48	4,70
Standard Error	0,71	0,96
Bessels' correction:		
Variance	12,67	23,01
Standard Deviation	3,56	4,80
Standard Error	0,74	1,00

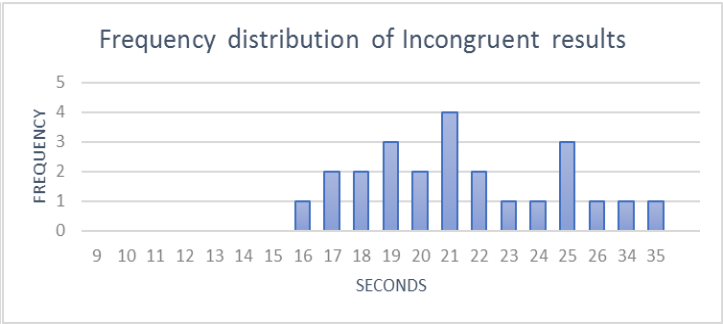
Note: I am using a French version of Excel, so decimals are shown with a comma here.

Calculations for the descriptive statistics above are provided in the associated Excel file.

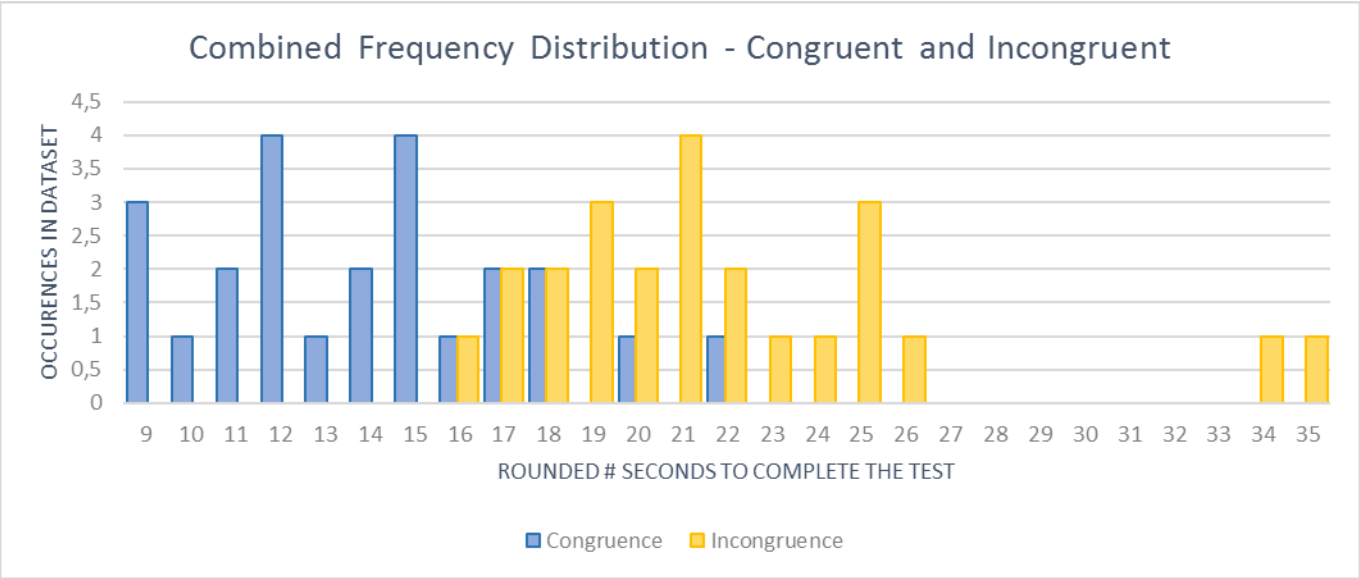
4. **Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.**

To visualize our data, we first round the number of seconds in our datasets to whole numbers (9, 10, etc). From this, we see that the minimum time to complete either test was 9 seconds (it happened several times during the congruent test) and the maximum was 35 seconds (it happened once during the incongruent test). From this, we can plot our frequency distributions for the congruent and incongruent tests, first separately and then combined:





Our dataset is quite small (24 observations each), therefore it is difficult to properly determine any trend in our results. However, it seems like both congruent and incongruent tests results observe a bell-shaped (normal) distribution. It is particularly striking in the combined histogram, where the incongruent tests results seem to have a wider range (from 16 to 35 seconds), compared to the congruent sample (with a range of 9 to 22 seconds) but where the normal distribution is somewhat more visible.



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?**

Here we conduct a paired t-test for dependent samples. Our experiment is a **within subject design** with a **'two conditions'** scenario: the control experiment is the congruent test, and our "treatment" corresponds to the incongruent test.

From our descriptive statistics section above, we know that:
$\bar{x}_{congruent}$= 14.05
$\bar{x}_{incongruent}$= 22.02

So our point estimate for $\mu_{congruent} - \mu_{incongruent}$ = -7.96, which is equal to the average of the differences in between our samples.
We then proceed to calculate a standard deviation of 23.67.
Finally, the sample standard deviation of the differences is 4.86 (intermediary calculations in Excel provided); and **our t-statistic is equal to -8.02**.

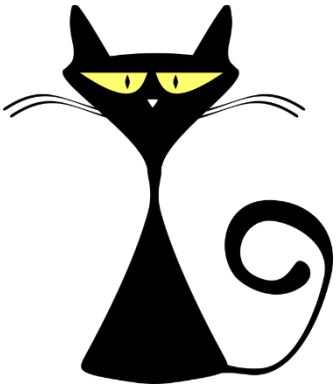
We take an $\alpha = 0.05$, and we have 23 (n-1) degrees of freedom here. Our t-critical values are therefore -2.069 and 2.069.
With a t-statistic of -8.02, we are well above the threshold of our critical values, therefore **we can reject H0**, which is coherent with our own observations when completing the test**.**
Our confidence interval is calculated to be [-10.02, -5.91]. Therefore, we can conclude that on average, **users will take almost 10 to 6 seconds less to perform a congruent test than an incongruent test.**

T-test		
Results		
Sample size	24	
Degrees of freedom	23	
Average difference	-7,96	
Standard deviation	23,67	** divide by n-1*
s	4,86	
t-statistic	**-8,02**	
Cohen's d	-1,64	
t critical value	2,069	
Confidence Interval	-10,02	*min*
	-5,91	*max*

6. **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!**

To me, it seems that it simply confuses the brain to have to ignore what the words read on a page to focus more on the colour. Doing that test the other way around (meaning having to read the words aloud while ignoring in which colour they are written) would be far easier. Another variation of the test that might have the same effect on the users could be to have them name what they see on an image (everyday objects or animals for instance) while ignoring what is written below the image. It can even be more confusing if somehow those two things are related…

Example:



DOG

FAIRY

RESOURCES

1. **On Statistics Theory:**

http://www.theanalysisfactor.com/longitudinal-repeated-measures/

https://statistics.laerd.com/statistical-guides/dependent-t-test-statistical-guide.php

https://statistics.laerd.com/statistical-guides/descriptive-inferential-statistics.php

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/tests-of-means/types-of-t-tests/

http://www.statisticshowto.com/when-to-use-a-t-score-vs-z-score/

http://www.statisticssolutions.com/manova-analysis-paired-sample-t-test/

2. **On the Stroop effect:**

https://imotions.com/blog/the-stroop-effect/