P1: Test a perceptual Phenomenon

Shreyas Ramnath
6th June 2017
Data Analyst Nanodegree

Questions

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

*The **color of the word**, whether it is same as that of the color denoted by the word or not is the **independent variable** and the **time taken** by the participants is the **dependent variable** as it depends on the color of the word.*

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

***H_0** - The null hypothesis is the mean response time for incongruent words is lesser than or equal to the response time for congruent words.*
***H_A** - The alternative hypothesis would be that the mean response time for incongruent words is greater than that of congruent words*

> ***$H_0: \mu_I \leq \mu_C$** (lower tailed test)*
>
> ***$H_A: \mu_I > \mu_C$***

Where μC and μI represent, respectively the congruent and incongruent population means.Hence, it would be a good idea to use a one-tailed,dependent sample t-test since the standard deviation of the entire population is unknown and the sample size is small (<30). Since the distribution deviates from the normal distribution, a z-test is not recommended. A t-test for dependent means has the following assumptions

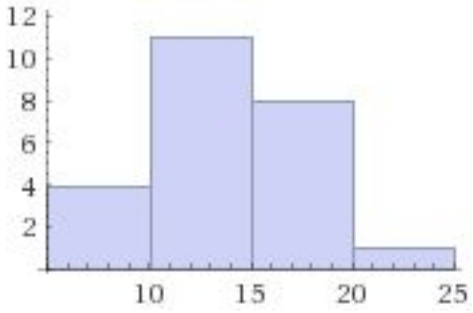
• *Interval or ratio scale of measurement (approximately interval)*
• *Random sampling from a defined population*
• *Samples or sets of data used to produce the difference scores are linked in the population through repeated measurement, natural association, or matching*
• *Scores are normally distributed in the population; difference scores are normally distributed*

A t-test is logical as we need to see that the word being incongruent does not reduce response time hence a one-tailed dependent t-test we help in analyzing the effect in the negative (lower-tailed) direction. The t-test is dependent as we have comparison of means of two related groups and analyze whether they are significantly different statistically and that the same participants are present in both groups.

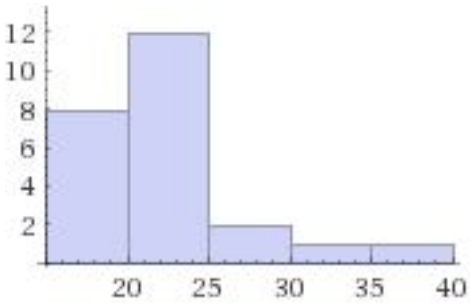
3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

Congruent words	***Incongruent words***
Length of data : *24 items*	***Length of data :*** *24 items*
Mean : *14.05*	***Mean :*** *22.02*
Median : *14.36*	***Median :*** *21.02*
Variance : *12.666*	***Variance :*** *23.011*
Standard deviation : *3.559*	***Standard deviation :*** *4.797*

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.



Congruent Words



Incongruent Words

X axis - Response Time in Seconds
Y axis - Frequencies

The maximum response time taken for a congruent word is 25s and 40s for an incongruent word. Also the frequency for the congruent words is peaking between 10-15s and for incongruent words between 20-25s.

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?

Mean difference *= -7.97*
Degrees of freedom *= 24-1 = 23*
Standard deviation *= 4.86*
Standard Error *= 4.86/√24 = .99*
$\propto$ *= .01*
T-statistic *= -7.97/.99 = -8.05*
From the t-table at df 23 and $t_{.99}$ *we get a value of*
T-critical *= -2.5 in the negative direction one tail t-test*

The t-statistic is in the critical region, hence alternative hypothesis is selected thus rejecting the null and $\propto$ *= .01 gives sufficient ground to conclude that incongruent words take more response time to be color recognized than congruent words. Yes the results did match up to my 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!

People are basically more adept at reading words than naming colors, hence the latter has less interference. The human brain is wired to read words effortlessly known as "automaticity hypothesis" but naming a color is not the normal instinct hence portraying itself as an interference in the stroop effect. Hence this basic instinct must be turned off before naming a color which is not always successful.

Counting the number of words rather than reading them out , produces a similar effect as that of the stroop effect. Human beings again are wired to read the words presented than to say the frequency of the word.

References

[1] https://www.thestudentroom.co.uk/showthread.php?t=301293
[2] http://www.statisticshowto.com/when-to-use-a-t-score-vs-z-score/
[3] https://en.wikipedia.org/wiki/Stroop_effect
[4] https://www.rit.edu/cla/gssp400/sbackground.html
[5] https://faculty.washington.edu/chudler/words.html
[6] https://faculty.washington.edu/chudler/java/readyn.html
[7] http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

[8] http://www.psychology.emory.edu/clinical/bliwise/Tutorials/TOM/meanstests/assump.htm
[9] http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/basics/null-and-alternative-hypotheses/