Stroop Effect statistical analysis

Background

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.

Investigation

In this case the independent variable is word condition (congruent or incongruent).
Dependent variable is time it takes to name the ink color of a word (μ_c for congruent condition and μ_{in} for incongruent condition).
Null hypothesis: it takes the same or even less time to name the ink color of the word when when the color of the ink doesn't match the name of the color, in other words it takes the same or less time to name words in incongruent condition.
Alternative hypothesis: it takes longer time to name the ink color of the word when the color of the ink doesn't match the name of the color, in other words when the word has incongruent condition.

$H_0: \ \mu_c \geq \mu_{in}$

$H_A: \ \mu_c < \mu_{in}$,

where μ_{in} - average time to name the ink color of words in incongruent condition, μ_c - average time to name the ink color of words in congruent condition.
As soon as we don't know the real population parameters such as mean value or standard deviation of a population and we only have samples it was chosen t-test to find how different two samples are from each other. In this case there are two dependent samples since each participant performed both conditions. Out of all measures of center I will use the mean. As soon as the sample sizes are not very big (only 25) to estimate population standard deviation I'll use sample standard deviation with Bessel's correction. A sampling distribution of sample means in this case is normal but it depends on a sample (population parameters are unknown), that's why I'll use t-distribution which is more prone to error, in other words more spread out and thicker in the tails than normal distribution.
As soon as the main assumption is that the time spent to name ink color of words in incongruent condition is longer than to name ink color of words in congruent condition it is more appropriate

to use one-tailed test in positive direction. It is also possible to use two-tailed test to say is the time significantly different or not for this two samples. But in this case it is expected that the time will increase so one-tailed test is more convenient. I'll perform one-tailed test in positive direction with α = .05, this will allow to say if results are statistically significant.

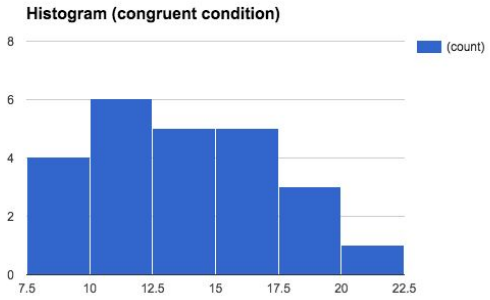
Descriptive statistics

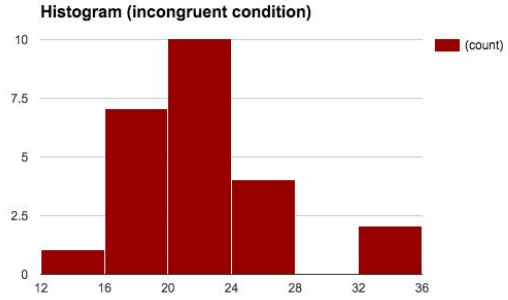
The sample size n = 24 for each condition.

Congruent	Incongruent
12.079	19.278
16.791	18.741
9.564	21.214
8.63	15.687
14.669	22.803
12.238	20.878
14.692	24.572
8.987	17.394
9.401	20.762
14.48	26.282
22.328	24.524
15.298	18.644
15.073	17.51
16.929	20.33
18.2	35.255
12.13	22.158
18.495	25.139
10.639	20.429
11.344	17.425
12.369	34.288
12.944	23.894
14.233	17.96
19.71	22.058
16.004	21.157

$\overline{x}_c = 14.051$, $SD_c = 3.559$

$\overline{x}_{in} = 22.016$, $SD_{in} = 4.797$





Both graphs are positively skewed. For congruent condition the graph is much smoother.

Statistical test

n = 24 (sample size) => df = 23 (degrees of freedom)
t(23) = 1.714, p < .05, one-tailed
SE = 0.993
t-statistical = 8.021
95% CI on the mean difference: (5.91, 10.019)
Effect size measures: d = 1.637219949, $r^2 = 0.737$

Result: this results are statistically significant. Reject H_0.
Conclusion: after statistical investigation it is possible to conclude that time spent to name the ink color of a word in incongruent condition (the color of the ink doesn't match the name of the color) is significantly longer than for words in congruent condition. This result corresponds the original hypothesis and our expectations.

Additional research:

There are several theory explanations for this effect:

- Processing speed: our brain reads the word faster than recognises the color because of the brain's ability lag to recognise the color.
- Selective attention: there is theory that color recognition is more difficult task for brain, so it will take a longer time to do that.
- Automaticity: this is most common theory in "Stroop Effect" and it says that as soon as color recognition is not "automatic process" there will be fluctuation in responses. Because of the habitual reading ("automatic process") the brain faster understands the word meaning. Probably if proceed this test with people who can't read the word meaning (words in unknown language for person) the results for congruent and incongruent conditions will be similar.
- Parallel distributed processing: this theory says that in our brain some analysis pathways are stronger than another. In this case the pathway of reading is stronger than color recognition.

All these theories explain “Stroop Effect” and all results that was get after samples investigation.

There are several variations of Stroop Test that provide similar results. One of them is Numerical Stroop Test that demonstrates the relationships between numerical values and physical sizes. The congruent condition is when size and value of a pair of numbers correspond, incongruent - when they don’t. This experiment needs to be processed separately for more statistical detail, but there is a conclusion that people recognize the values slower in incongruent number condition.

Sources

1. Stroop Effect https://en.wikipedia.org/wiki/Stroop_effect
2. Numerical Stroop Effect https://en.wikipedia.org/wiki/Numerical_Stroop_effect
3. Example data set https://drive.google.com/file/d/0B9Yf01UaIbUgQXpYb2NhZ29yX1U
4. Statistical calculator http://www.graphpad.com/quickcalcs
5. All data calculations https://docs.google.com/spreadsheets/d/1rFq0yBXlK0qFSYYT1pnvQVCCMU3ElSt02Sux6wXk290