Test a perceptual phenomenon: the « Stoop effect »

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.

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

Independent variable is the word condition (congruent or incongruent).
Dependent variable is the time it takes to name the ink colors of the words displayed.

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

In order to compare results of the « Stroop effect » test and figure out whether test conditions has a significant impact on the time to carry out the test, it is proposed to carry out a statistical test.

The paired sample t-test, sometimes called the dependent sample t-test, is a statistical procedure used to determine whether the mean difference between two sets of observations is zero. In a paired sample t-test, each subject or entity is measured twice, resulting in pairs of observations. Common applications of the paired sample t-test include case-control studies or repeated-measures designs. Suppose you are interested in evaluating the effectiveness of a company training program. One approach you might consider would be to measure the performance of a sample of employees before and after completing the program, and analyze the differences using a paired sample t-test.

Type:
Each test (congruent words and incongruent words conditions) has been performed by the same person. However results for the tests are limited in size and not representative of the entire population. Therefore the statistical test to compare the mean results of each test should be a **paired sample t-test**.

Direction:
We don't know if a test condition has a favorable or unfavorable effect on the time to carry out the test. Therefore we chose to perform a **two tailed test**.

Null and alternative hypothesis:
The null hypothesis is that test condition (congruent or incongruent) has no impact on the test results (time to complete the test). In that case the true mean difference is equal to zero

$$H_0 : \mu_{delta} = 0$$

The alternative hypothesis is that the test condition has an impact on the test results. In that case the true mean difference is not equal to zero

$$H_1 : \mu_{delta} \neq 0$$

In above equations
      - μ_{delta} stands for the true population mean of the difference between the incongruent and the congruent test
   - H_0 stands for the null hypothesis
      - H_A stands for the alternative hypothesis

Threshold:
In order to reject the null hypothesis, a **significance level is set at 5%**. That means, the null hypothesis will be rejected if the probability that there is a difference between the test results is higher than 95%.

3 & 4: Descriptive statistics and visualisation of the distribution of the sample data

Following figure provides a description of the test results for congruent words and incongruent words conditions.
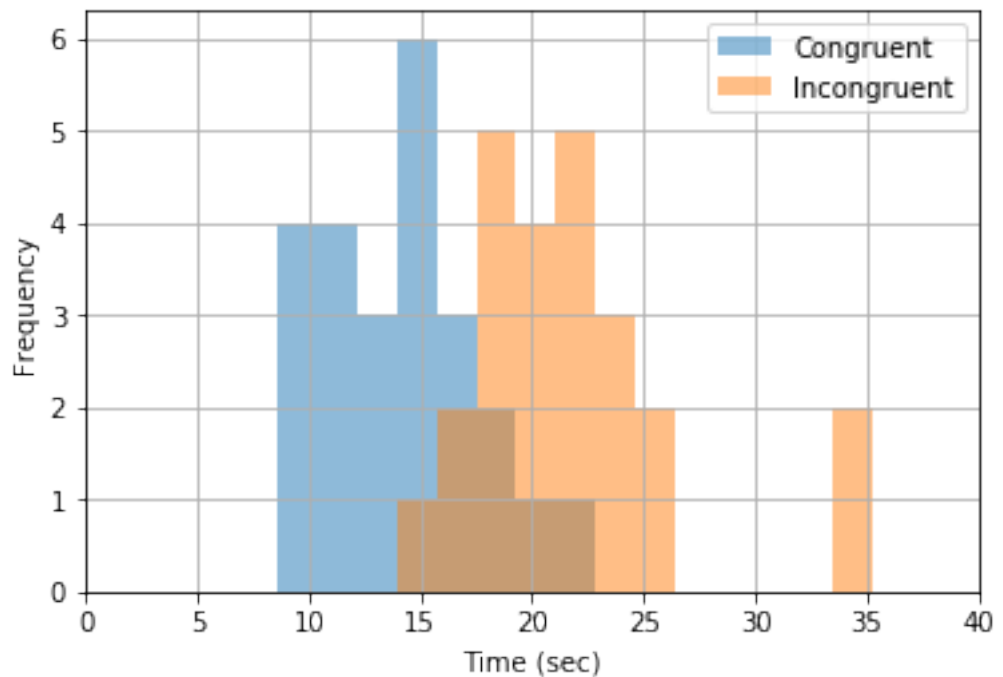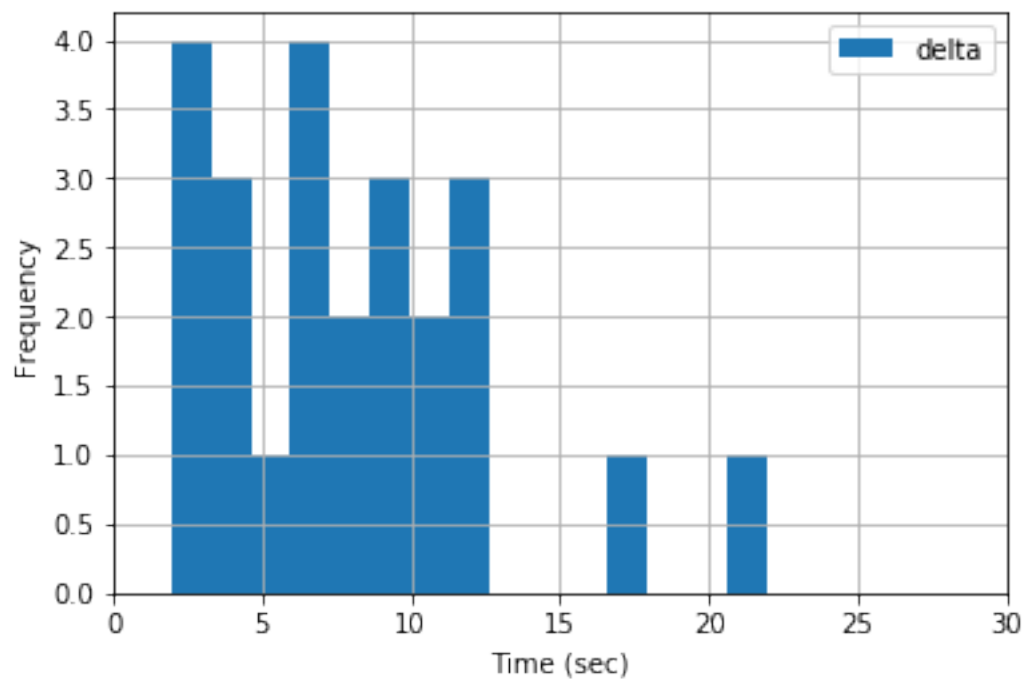

	Congruent	Incongruent
Number of samples	24.0	24.0
Min (sec)	8.63	15.69
Max (sec)	22.33	35.26
Mean (sec)	14.05	22.02
Median (sec)	14.36	21.02
Sample std (sec)	3.56	4.8

The visualisation of the distribution of the sample data (histogram), shows that both test results show a normal distribution with some overlap.

The descriptive statistics analysis shows that:
     - in average it takes 8 more secondes to carry out the incongruent test compared to the congruent test
     - sample standard deviation for both tests is around 4 secondes

Following figure provides a description of the difference in test results (incongruent - congruent).

	delta Incongruent - Congruent
Number of samples	24.0
Min (sec)	1.95
Max (sec)	21.92
Mean (sec)	7.96
Median (sec)	7.67
Sample std (sec)	4.86

The visualisation of the distribution of the sample data (histogram), shows that the delta has a positively skewed distribution.

The descriptive statistics analysis shows that:
- the difference is always positive and range from 2 to 22 secondes
- in average the difference is 8 secondes
- sample standard deviation of the difference is around 5 secondes

An inferential statistical analysis is required to conclude whether results of both tests are significantly different.

5. Statistical test

A two tailed, paired sample t-test is performed on the dataset described above.

Python code output:

```
alpha: 0.05
degree of freedom: 23
t_statistic: 8.021
t_critical: 2.069
hypothesis testing result: reject H0
r2: 74%
margin of error: 2.05
delta mean: 7.96
confidence interval for delta mean: (5.91,10.02)
```

Result of the paired sample t-test is
$t(23)=8.021$, $p<.05$, two tailed
We reject the null hypothesis and conclude there is a significant difference between the two sample results.
The confidence interval at 95% for the delta between incongruent and congruent test is (5.91 to 10.02 secondes).
74% of the difference between the two test results can be explained by the test conditions.

Results are in line with my expectation and with my experience of the test. In fact it took me longer to tell the colors when they were not matching the words.

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!

My understanding of the phenomena is that human brain have different perception capabilities and when those are in conflict it takes more time to process. In the present case, the two perception capabilities are word reading and color identification. When both informations are aligned (congruent words condition), brain processing time is quick while when informations are different (incongruent words condition) it takes more time.

I've made some tests at home with my family and it did not work as well due to the fact the words were written in English. When I used the French version of the test (see figure below), the time difference between the congruent and incongruent conditions was much more important. I believe it is due to the fact that my daughters (9 and 7 years old) do not read English yet while they just learnt to read French. I would like to do the same test with my younger son (4 years old). My intuition is that there should be no difference between the English and the French test as well as between the congruent and the incongruent conditions because he does not know how to read yet. Of course a proper statistical test should be set up if I wanted to prove it.

The website from the faculty of Washington (http://faculty.washington.edu/chudler/words.html) proposes other « Stroop effect » test with direction, numbers and animals instead of colors. I've tested and found them easier than the color test. It is probably due to the fact that the brain capabilities in conflict (direction reading vs positioning, digit reading vs digits counting, word reading vs animal recognition) are of more even strength.