Context

As Washington University Stroop Test website mentions, *"Stroop Effect is named after J. Ridley Stroop who discovered this strange phenomenon in the 1930s that states how **the words themselves have a strong influence over your ability to say the color**. The interference between the different information (**what the words say and the color of the words**) your brain receives causes a problem.*

There are two theories that may explain the Stroop effect interference:

1. *Speed of Processing Theory: **words are read faster than colors are named**.*
2. *Selective Attention Theory: **naming colors requires more attention than reading words.**"*

Question 1: Identify variables in the experiment.

Independent:

- **condition**: categorical variable with two possible values, *congruent* or *incongruent*. First part of the experiment is applied showing congruent list of *words match the color*. Second part of the experiment shows a list of incongruent words, that is *color words not matching the color ink*.

Dependent:

- **completion time:** continuous variable in seconds. The time each individual takes to complete the experiment step depending on condition.

Question 2a: Establish hypotheses.

By looking at the theories, we expect **people require more time (one-tail test) to complete the test when saying incongruent words**, in other words, people tend to get confused when saying color words that don't match their color ink, hence, taking on average, more time to complete the test.

Null Hypothesis[1]: there is *no significance difference* between population mean μ_c (time to complete Stroop test with congruent words) and the population mean μ_i (time to complete Stroop test with *incongruent words as intervention*) and not by chance.

$$H_0 : \mu_c = \mu_i$$

Alternative hypothesis: population mean μ_c takes *significative* less time to complete the test compared to population mean μ_i after a condition (time to complete Stroop test with *incongruent words as intervention*) is applied and not by change.

$$H_a : \mu_c < \mu_i$$

[1] https://classroom.udacity.com/nanodegrees/nd002/parts/accef51e-2114-4bd9-8ba3-e982e697b76e/modules/22472716-0909-458e-b2ed-2242650a9626/lessons/4578095863/concepts/1470193290923

Question 2b: Establish a statistical test

Since we cannot assume congruent test represents a population, we have few samples (<30) and we don't know population standard deviation σ , so I'd not use a z-score test. Instead we just have two samples, first same taken after a test is applied and a second sample after an intervention is applied to the same subjects. We have conditions, the results are dependent to a condition. Additionally it excludes the need for a independent variables t-test because we are applied the same experiment to the same people and the sample size is not big enough.

- Longitudinal test doesn't make sense because we are not comparing a variables with the same sample in different points in time only, we are including an intervention and there is a short period of time between tests.
- **Repeated measures design pretest-posttest makes sense because we are measuring a variable for the same sample before and after a treatment (change in color/word) is applied. Also we control the subjects taking both tests and the period of time.**

Therefore, we will apply a dependent/paired t-test to see if there is a significant effect in time (seconds) due to the treatment (incongruent color and words) with the same people taking the treatment.

1. We have one paired sample, one with completion time for tests using congruent condition and a second including completion time for tests using an intervention condition with incongruent colors.
2. We then calculate **n**umber of observations, **mean**, point estimate (**diff**) and standard deviations (**sd**) for each sample and standard error:
    a. $diff = \mu_q - \mu_a$

    b. $S = \sqrt{\frac{\sum_{i=1}^{n} (\mu_i - \mu)^2}{n-1}}$

    c. $StdErr = \frac{S}{\sqrt{n}}$
3. Then, t is calculated with the following formula:
    a. $t = \frac{diff}{StdErr}$

4. Assumption: Alpha level would be 0.05, one tailed and we define t-critical values[2] with 23 degrees of freedom:
    a. critical region = -1.714

5. We decide whether to reject or fail to reject[3] the null hypothesis based on t statistics and critical region.
6. We define a causal effect statement since we applied an experimental design.

[2] https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg

[3] https://classroom.udacity.com/nanodegrees/nd002/parts/accef51e-2114-4bd9-8ba3-e982e697b76e/modules/22472716-0909-458e-b2ed-2242650a9626/lessons/4577134255/concepts/1364883890923

Question 3: Report descriptive statistics.

Congruent fivenum

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.63	11.90	14.36	14.05	16.20	22.33

variance: 12.66903

Incongruent fivenum

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.69	18.72	21.02	22.02	24.05	35.26

variance: 23.01176 (almost double compared to first sample)

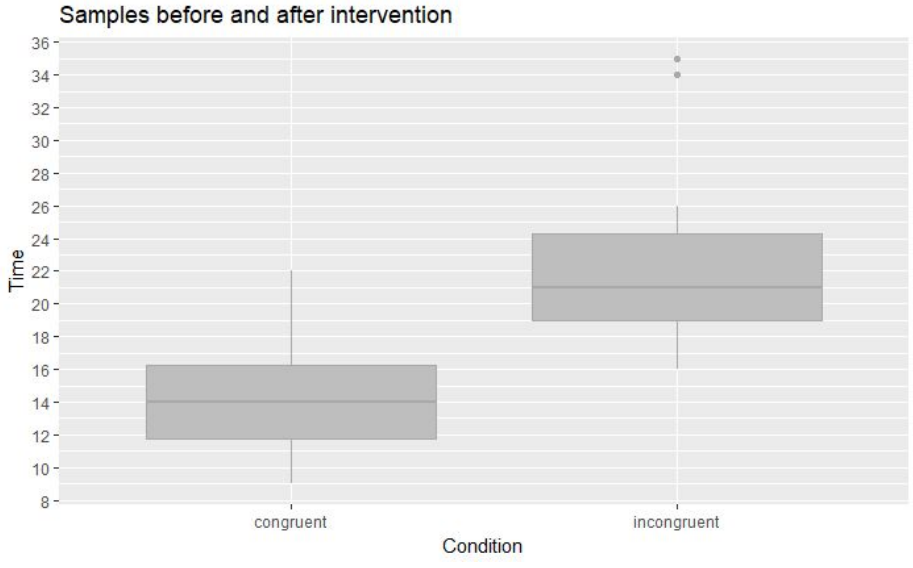
As described in the previous question we have calculated some key descriptive statistics:

- n = **24**.
- μ_q = **14.0511** seconds (mean with no intervention)
- μ_a = **22.0159** seconds (mean after intervention)
- diff= **-7.9648** seconds (mean difference, point estimate)
- degrees of freedom = **23**
- S = **4.864827** (calculated in attached xls file and with R sd function -uses Bessel correction-)
- Standard Error = **0.9930**
- t = **-8.0207 (calculated in spreadsheet) also resulting from R test:**

Paired t-test

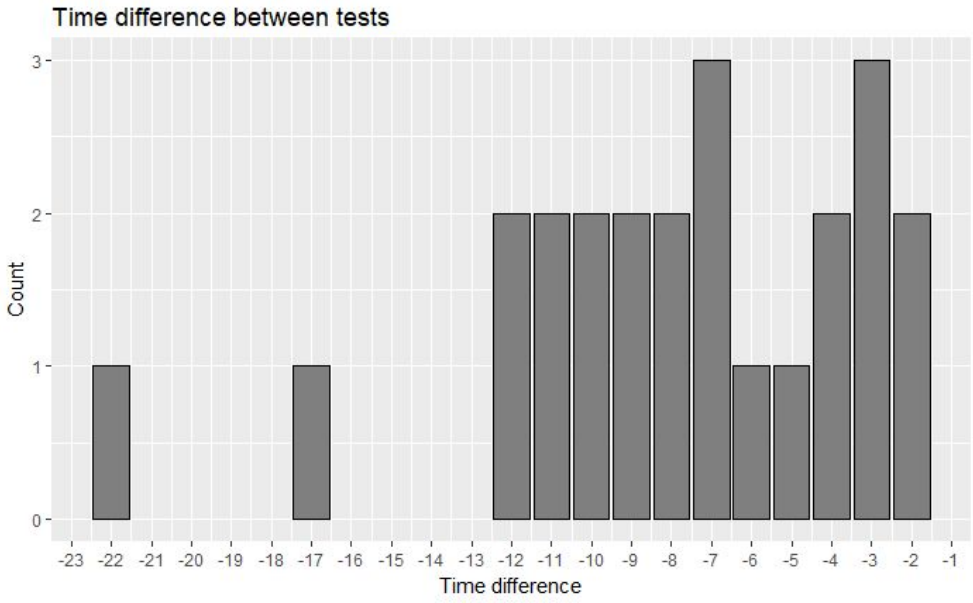
data: pf$Congruent and pf$Incongruent
t = -8.0207, df = 23, p-value = 1
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval: (-9.666715 , Inf)
sample estimates:
mean of the differences -7.964792

Question 4: Plot the data.



This plot shows key differences (fivenum) between sample1 data sample2, before and after intervention. Mean after intervention almost double first test ,14 seconds to 22 seconds. Test 2 seems to have two outliers (but we don't have enough information to remove them from analysis)

Additionally, time difference between tests doesn't show a normal distributions. Time scoring in test 1 are considerably lower to results in tests 2 and not necessarily Uq - Ua = 0.



Question 5: Perform the statistical test and interpret your results.

- Mean Difference = **-7.965**
- Effect size measure, Cohen's D = **-1.637219949**
- Alpha = **0.05**
- T critical, left tail: **-1.714** using t-table.
- T Value = **-8.021**
- Confidence interval = Mean diff +- T critical value * Std Error = **(-9.667,-6.263)**
    - So, users would take between 6.3 and 9.7 additional seconds to complete the seconds test with incongruent color/words.

Since t value -8.021 is within critical region (-1.714 or less) with 95% confidence level we reject the null hypothesis.

Based on this experiment is clear to me that intervention really affect user behaviour on second test, there is a big difference in seconds. One thing I notice is that test has two possible outliers. By removing those two values and testing again I got the same conclusion, t-value is still lower and within critical region.