

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

Independent variable is the variable that represents the input or the reason for possible variation in the dependent variable. In turn, the dependent variable is the variable whose output or variation is being studied. In other words, the independent variable is a variable that is being manipulated in an experiment in order to observe the effect on a dependent variable.

In the Stroop task case we have 3 variables: time to name the ink color, the congruent and incongruent words. Since the task is to name the ink color, the **time it takes to name the color** is the **dependent** variable and the **word/color congruency condition** is the **independent 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.*

Hypotheses set:

H₀: The mean time it takes to name congruent and incongruent word ink color is the same, $\mu_C = \mu_I$;

H₁: It takes more time on average to name the ink color of incongruent words, $\mu_C < \mu_I$;

μ_C – population mean time; the mean time it takes to name congruent words

μ_I – population mean time; the mean time it takes to name incongruent words

The statistical test that I expect to perform is either z-test (if sample is > 30 or sample < 30 but standard error of the population is known) or t-test (if sample is < 30 and standard error of the population is unknown). In addition, at the core of both of these test is the assumption of **normal (Gaussian) distribution**. Normal distribution is useful due to the central limit, which essentially says that if you have many independent variables generated by different distributions, the aggregate of these variables will tend to normally distributed.

Type of test: **paired t-test** as each test (congruent and incongruent) is done on the same person once. Therefore, we test whether the mean of the differences between the paired observations is equal to a specified value.

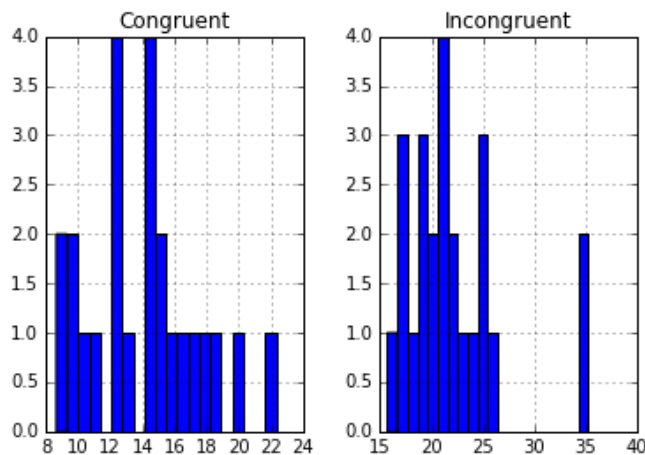
The test should be **one-tailed** as we care about deviations to one direction of the population mean. We are inferring, whether the incongruent population mean time (μ_I) is higher than the congruent population mean time (μ_C)

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

	Congruent	Incongruent
count	24	24
mean	14.051125	22.015917
std	3.559358	4.797057
min	8.63	15.687

25%	11.89525	18.71675
50%	14.3565	21.0175
75%	16.20075	24.0515
max	22.328	35.255

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.



The congruent data seems more spread out around the mean, whereas with exception of the 35s bucket, congruent data is clustered more around the mean.

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?

CI – 99%

t-statistic = -8.02070694

p-value= 4.10300059e-08

Since the p-value is ~ 0 I reject the Null hypothesis of equal means. Therefore, I reject the Null hypothesis at 99% confidence level

The results matched my expectations.

Sources:

<https://www.khanacademy.org/>