

DESCRIPTIVE STATISTICS project

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

Independent variable is **text color and meaning of the text is same or different**

Dependent variable is **time taken for identify the color of text**

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

Set of Hypotheses:

H_0 : There is no difference in time duration between the congruent and incongruent tests

H_1 : There are some increases in time duration between the congruent and incongruent tests.

There are two means we need to find

μ_1 = mean of time spent for a group of congruent test

μ_2 = mean of time spent for a group of incongruent test

For H_0 (null hypotheses)

$(\mu_1 - \mu_2) \geq 0$ [the time duration difference of the congruent and incongruent tests would equal 0 or be greater than 0, i.e the incongruent times taken would be less than or equal to the congruent times taken for the population]

For H_1 : (alternative hypotheses)

$(\mu_1 - \mu_2) < 0$ [the time duration difference of the congruent and incongruent tests would be less than 0, i.e the incongruent times taken would be greater than the congruent test times taken for the population]

Statistical test:

Here we are going to find statistics test for two groups and doing hypotheses test. So for that we need to perform **t-test**, because we don't know mean of total population and standard deviation.

Type of T-Test

Paired t test: The participants were measured using two time points (congruent and incongruent).

Dependent t-test route: This test is with the same group of people that was tested on the congruent words then tested on the incongruent words.

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

Central tendency:

From the data set of the Stroop experiment

Determining the Means of the congruent (as X_c) and the incongruent (as X_{ic})

$$X_c = 14.05 \quad X_{ic} = 22.02$$

Determining the Median of the congruent (as C_{med}) and the incongruent group (as IC_{med})

$$C_{med} = 14.36 \quad I_{med} = 21.02$$

No Mode available as per the dataset, because the dataset contains unique time.

Variability:

Formula of sample standard deviation

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

When applying this formula step by step for *congruent*

1. Mean is = 14.051125

2. $\sum (x_i - \bar{x})^2 = 301.6250729$

3. $\frac{\sum (x_i - \bar{x})^2}{n-1} = 12.66902907$

4. Standard deviation of sample statistics(*congruent*) is SD= 3.559

When applying this formula step by step for *incongruent*

1. Mean is = 22.01591667

2. $\sum (x_i - \bar{x})^2 = 550.5485907$

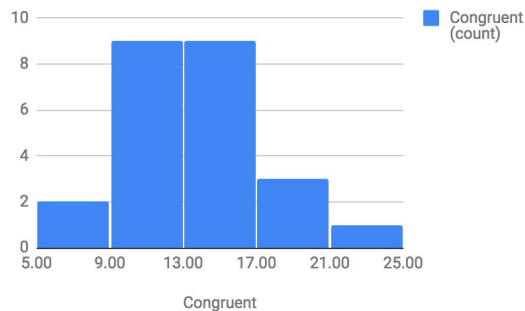
3. $\frac{\sum (x_i - \bar{x})^2}{n-1} = 23.01175704$

4. Standard deviation of sample statistics(*incongruent*) is SD= 4.797

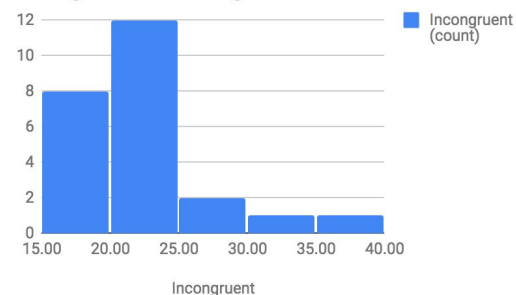
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

Separate Histogram chart for congruent and incongruent chart.

Histogram of Congruent



Histogram of Incongruent

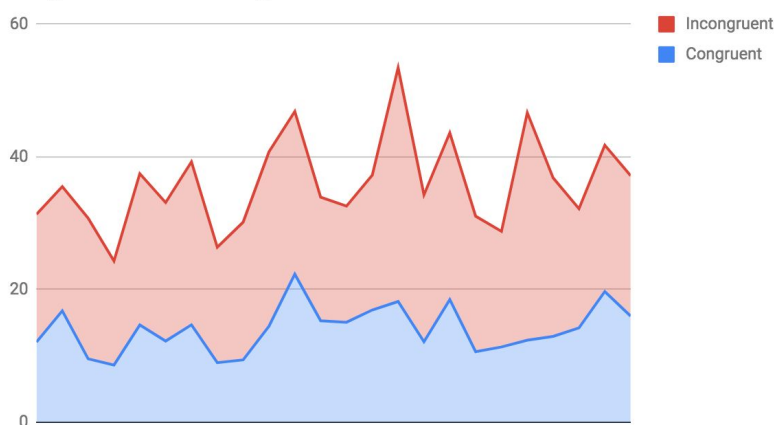


Here we can see the congruent is *mid right skewed distribution*. Majority of the time taken within 21 seconds. The congruent histogram if ignoring the tail beyond 20-21 seconds resembles closely to a normal distribution curve

And the incongruent is *Positively skewed (right skewed) distribution*. It shows most of the time with 15 to 25 seconds.

Combination:

Congruent and Incongruent



In the data we see that there is a significant increase in duration of time taken to finish the test from the congruent test vs the incongruent test.

When considering all data, *the time taken for incongruent test will take longer to finish than the congruent test.*

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?

We already found mean for congruent and incongruent

$$X_c = 14.05 \quad X_{ic} = 22.02$$

We can calculate point estimate : $X_c - X_{ic} \Rightarrow 14.05 - 22.02 = -7.97$

A point estimate is handy to know as it gives us our “Best guess” on the difference of the population between the two tests (population difference estimate of congruent versus incongruent times).

We need to find **standard deviation of differences (s)**

1. Mean is $\bar{x} = -7.964791667$

2. $\sum (x_i - \bar{x})^2$ (Difference of Congruent and Incongruent – mean of differences)²
= 544.3304

3. Variance of difference = 23.667

4. $\frac{\sum (x_i - \bar{x})^2}{n-1} = 4.865$

5. Standard deviation of Differences **SD= 4.86482691**

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Now we can find t-statistics

$$t = \frac{-7.97}{4.865 / \sqrt{24}}$$

$$\mathbf{t = -8.026}$$

Now we have the t-Statistic. We can compare it to the t-critical value to determine if the null hypothesis can be accepted or rejected.

We know our sample has 24 participants then $n=24$. The degrees of freedom is 23

Since we are aiming to find with 99% Confidence level for a one tail our $\alpha = 0.005$

So $df = 23$, and $\alpha = 0.005$ using that find value in t-table => **+ - 2.807**

Our t-statistics is **-8.026**.

So we **reject the hypothesis** due to statistics and critical area.

The t-test confirms that incongruent tests would normally take longer than congruent tests.

So, reading the text color of the different printed name is **take bit more time** compared to same color and printed name.