

Q. 1

Independent Variable: Type of word: Congruent, Incongruent words

Dependent Variable: Time it takes to name the ink colors (on th congruent and the incongruent tasks).

Q. 2

$u(\text{congruent})$ = Mean of the time taken for speaking Congruent words,

$u(\text{Incongruent})$ = Mean of the time taken for speaking Incongruent words

Set of Hypothesis:

H₀ (Null hypothesis):

- $u(\text{congruent}) - u(\text{incongruent}) = 0$
- There is no difference in the time taken to name the color of the ink (in which the word is printed, congruent and Incongruent words)

H_a (Alternate Hypothesis):

- $u(\text{congruent}) - u(\text{incongruent}) \neq 0$
- The time taken to name the ink colors of congruent words differ from that of the Incongruent words.

Kind of statistical test: -

Two tail t-test Paired t-test

1. The paired t-test calculates the difference within each before-and-after pair of measurements, determines the mean of these changes, and reports whether this mean of the differences is statistically significant.
 2. The sample are ****dependent**** i.e., we have dependent samples and that's why using paired t-test is recommended.
 3. Because we do not know the population standard deviation that's why we use t-test. So, we will use the sample standard deviation.
 4. I chose two-tailed test because I do not know (or am not sure) that whether it will take more time to speak incongruent words or not, for every participant. So, we can only guess that maybe both the times will differ as per our Alternate Hypothesis.
- Assumption made by the t-test:
 - Population from which the sample is from is Normal; Scores are normally distributed in the population
 - Samples or sets of data (congruent and incongruent data) used to produce the difference scores are linked in the population through repeated measurement, natural association, or matching which in this case is the congruent and the Incongruent word tests on the same participants once for each type.
 - Congruent and Incongruent Samples should be random samples from a population. That is, Random sampling from a defined population.

Q. 3

Measure of central tendency:

$$\text{Mean } (X) = X(\text{congruent}) - X(\text{incongruent}) = -7.964791667$$

or

$$\text{Mean } (X) = \text{Sum}(A_i - B_i)/N = -7.964791667$$

where $X(\text{congruent}) = 14.051125$ = Mean of Congruent time data

$X(\text{incongruent}) = 22.01591667$ = Mean of Incongruent time data

A_i = i (th) term in the congruent data set

B_i = i (th) term in the congruent data set

N = Total number of participants = 24

df = degree of freedom = $N-1$ = 23

Measure of variability:

$$\begin{aligned}\text{Standard deviation}(SD) &= \sqrt{(\text{Sum of squared deviations}) / (N-1)} \\ &= \sqrt{544.33044 / 23} \\ &= 4.86482691\end{aligned}$$

where SS = Sum of squared deviations $\Rightarrow \text{Sum}([C(i) - I(i)] - X(d))$, for $i = 1$ to N

$C(i)$ = i (th) value of the congruent data

$I(i)$ = i (th) value of the incongruent data

$X(d)$ = difference between the means of congruent and incongruent data

N = Total number of participants

$$\begin{aligned}\text{Variance} &= \text{Sum of Squared Deviations} / (N - 1) \\ &= 544.33044 / 23 \\ &= 23.66654087\end{aligned}$$

(and accordingly SE (Standard Error) = SD/\sqrt{N} = 0.993028635)

Q. 4

plots (distribution of the sample data)

--> CODE AT LAST

Q. 5

Statistical test:

- X_c = Mean of the Congruent time, X_i = Mean of the Incongruent time
- $u(c)$ = Population mean of the congruent time, $u(i)$ = Population mean of the Incongruent time
- Point Estimate = $X_c - X_i = -7.964791667$
- t-statistic = $(X - u) / SE$, ($X \Rightarrow X_c - X_i$) and ($u \Rightarrow u(c) - u(i) \Rightarrow 0$ (Null hypothesis))
 - = $(X_c - X_i) / SE$
 - = $(-7.964791667) / SE$
 - = $-7.964791667 / 0.993028635$
 - = -8.020706944

Confidence level: 95% confidence interval

- with 2.5% on the left and 2.5% on the right of the distribution, we are left with 95% in the middle.
- So, CI = 95%, with (-2.069 and +2.069) t-statistics.

Alpha-level: 0.05 df = N-1 = 23

Critical Statistic value: -2.069, +2.069

Accept/Reject the null hypothesis:

- Since our t-statistic value, -4.116143531, falls in the critical region. So, we will **reject** the Null-Hypothesis (H_0).
- reason: -4.116143531 (t-statistic) < -2.069 (t-critical Statistic) So, it falls in the critical region.
- Hence, we accept the Alternate Hypothesis (H_a). As, we can clearly see that the time taken to name the Congruent words is lot less than naming the Incongruent words.

Thus, there is **significant difference** between the time taken to speak congruent and incongruent ink words. We can also say that the type of words (congruent/incongruent) had the causal effects on the speaking time of the participants.

Did the results match up with my expectation?

- Yes, as I expected that due to the extra thinking time, to identify the ink color of the incongruent words, there will be some difference between both the time results measured.

Q. 6

Factors for the effects observed:

- Maybe it is possible that

Alternative / similar task with similar effect:

- the qwerty and the alphabetical keypads example that was given in the lessons.
- In that example also, the users were used to use qwerty keypads so that's why they were making more errors using with the alphabetical keypad as they were not used to it prior the task.

```

In [4]: # Q 4
# plots to visualize the distribution of sample data

#

import matplotlib.pyplot as plt

%matplotlib inline

# data 1 is the congruent data
data_1 = [12.079, 16.791, 9.564, 8.63, 14.669, 12.238, 14.692, 8.987, 9.401, 14.48, 22.328, 15.298, 15.073, 16.929, 18.2, 12.13, 18.495, 10.639, 11.344, 12.369, 12.944, 14.233, 19.71, 16.004]

# data 2 is the incongruent data
data_2 = [19.278, 18.741, 21.214, 15.687, 22.803, 20.878, 24.572, 17.394, 20.762, 26.282, 24.524, 18.644, 17.51, 20.33, 35.255, 22.158, 25.139, 20.429, 17.425, 34.288, 23.894, 17.96, 22.058, 21.157]

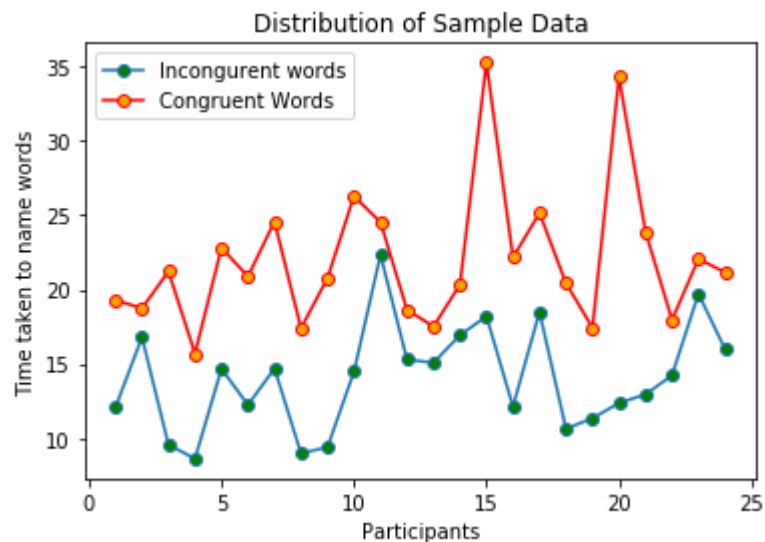
# plot for the congruent data
plt.plot(range(1, 25), data_1, marker='o',
         markerfacecolor='green', markersize=6)

# plot for the incongruent data
plt.plot(range(1, 25), data_2, '-', color='red', marker='o',
         markerfacecolor='orange', markersize=6)

fig = plt.gcf()
plt.title('Distribution of Sample Data')
plt.xlabel('Participants')
plt.ylabel('Time taken to name words')
plt.legend(['Incongruent words', 'Congruent Words'], loc='upper left')
plt.show()

# Incongruent words time is always higher than the congruent words time

```



Time taken by the participants for naming Incongruent words remains always higher than naming the congruent words.

Materials referred during the project:

- [plt.plot\(\)](https://matplotlib.org/api/_as_gen/matplotlib.pyplot.plot.html) method for plotting data in python
- Statistics Placement Advisor (<https://www.udacity.com/course/viewer#!/c-ud134-nd/l-4446458586/>)