Test a Perceptual Phenomenon

Q1. What is our independent variable? What is our dependent variable?

Ans :-

1. Here the Congruent words and In-Congruent words are the independent parameters

2. While the Time taken to name the ink colour is the dependent variable

Here we can observer from the data that incongruent words take more time to be named as compared to incongruent **as depicted in the sample** As depicted in this snapshot of the sample data we are considering

Q2. What is an appropriate set of hypotheses for this task? What kind of   statistical test do you expect to perform? Justify your choices.

Ans:-

Here we will perform Hypotheses Testing

Here we will consider the Null Hypothesis and a Alternative Hypothesis

1. **Null Hypothesis**:- In this Hypothesis we generally make a statement without any proof or data and we don't consider any relationship between the two variables it lakes observation

2. Null Hypothesis is **indirect and implicit** and is denoted by **H0(H-zero)** and they carry '=' sign in their formulation

3. **Mathematical form** $H0 : \mu = k, H0 : \mu <= k, H0 : \mu >= k$

4. **Alternative Hypothesis** is based on the statistical results and observation where we establish some relationship between the two variables based on the observation

5. Alternative hypotheses is **Direct and explicit** and is denoted by **H1(H-one) or Ha** and they carry '$\neq$' sign in mathematical formulation

6. **Mathematical form** $H1 : \mu \neq k, H1 : \mu > k, H1 : \mu < k$

7. Here μ represents the parameters we are working with like proporti**on etc**

8. **'k' or** μI represents a content value or the intervention of the test

9. Here we will use a **Two Tailed Test(Non-Directional statistical Test)** Since we don't have the idea about the direction in which values will move

10. Here the general trends is that the Congruent words require less time to be recognised as compared to the in-congruent words so here we are simply comparing the means of the sample sets from the two category to rectify this fact hence here we want to ensure that the values that we are comparing lies in the 95% samples hence we use two tail Testing.

11. **Here we are assuming** the given sample data set as a population set for performing these tests while in real-world scenarios its a sample of the much bigger population set.

Q3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

Ans:- Following Descriptive statistics have been calculated about the given data set

1. **<u>Central Tendency :-</u>** Here we have calculated the Mean and the Median for the data *(Mode cannot be calculated since the frequency of different items is 1 only hence technically all are mode !)

2. Here we can observer that the Congruent words have lower mean (less response time) while the incongruent words have higher mean(higher response time) which is as expected from the data

3 . Median also shows a similar trend

4. **<u>Variability:-</u>** Here we have calculated the range of the values ,variance and Standard Deviation

5. As expected the range boundary values of the Congruent words is less than that of the incongruent words

6. A similar trend is observed in the case of variance and Standard deviation which indicates a steeper curve for Congruent and a less steeper for the Incongruent curves.

7. Here since this is a sample of a bigger population a more accurate Standard Deviation will be calculated with Bessels correction hence to increase the value of SD to compensate for those points that are not covered in the sample under consideration

Q4. Provide one or two visualisations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

Ans :-
1. Here the First scatter plot and the histogram represents the Congruent Data set.

2. Here Second scatter plot and the histogram represents the In-Congruent Data Set

3. Here we can observer that the histogram for Congruent words represent more of a normal curve with mode in range 9-17 s

4. Hence it is in line with the general trend that maximum people take only 9-17s in identifying the Congruent words

5. While if we talk about the histogram for in-Congruent words it is a positively skewed curve where median is in the range 20-25 seconds which is again support are general trend of most of the people talking approx. 20-25 s to identify in-congruent words which is greater than that of congruent word

Standard Deviation	12.66902907	23.01175704
With Bessels correction	3.559357958	4.797057122

6.   Mode(Congruent) < Mode(In-Congruent)

7. Again in the scatter plot we can see the range for max points in congruent words is 8-20s
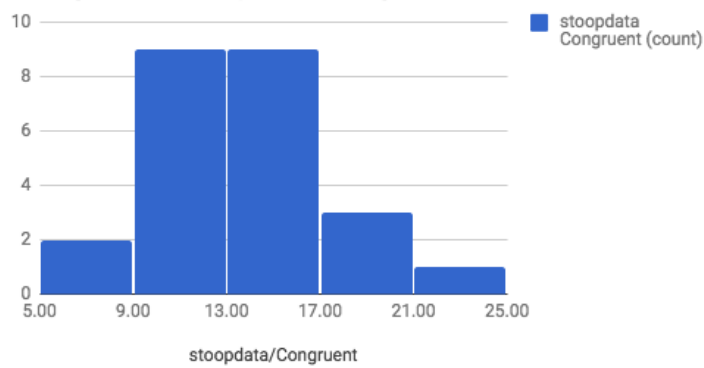
8. While that of the In-Congruent words is 15-30s

9. Hence this is also in line with the general trend of the users as depicted by Histogram
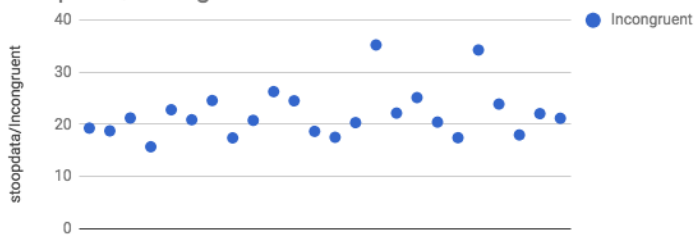
stoopdata/Congruent
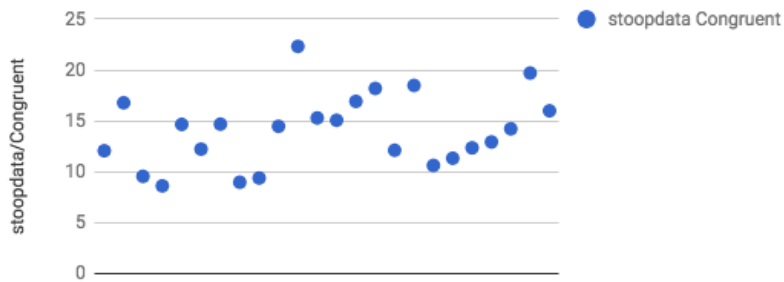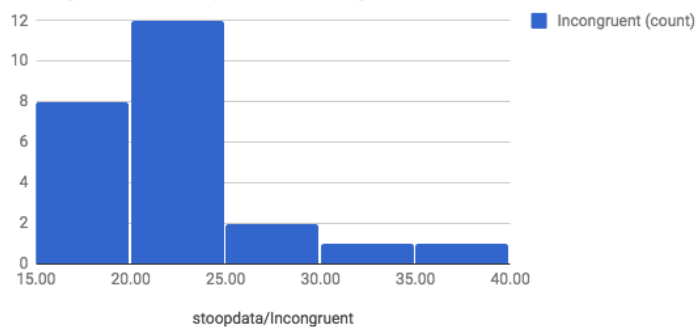

Congruent Plot

Histogram of stoopdata/Congruent

Congruent Plot

stoopdata/Incongruent

InCongruent Plot

Histogram of stoopdata/Incongruent

InCongruent Plot

Z values (Congruent [C])	Z-Values(In-Congruent [IC])	Here we are taking Alpha = 0.05%			
		Critical Z Values = +1.96 to -1.96			
Congruent [C]	In-Congruent [IC]	Congruent [C] (ratio)	Congruent [C] (Percentage	In-Congruent [IC] (ratio)	In-Congruent [IC] (Percentage)
-1.093840508	-1.178112331	0.1379	13.79%	0.119	11.90%
-1.342834417	-0.4696145579	0.0918	9.18%	0.3228	32.28%
-1.34219268	-0.01825828603	0.0901	9.01%	0.496	49.60%
0.03061084908	-0.2358322219	0.512	51.20%	0.409	40.90%
0.9464977213	0.3639739265	0.8264	82.64%	0.6406	64.06%
0.07091192501	0.9845649644	0.5279	52.79%	0.8365	83.65
0.04421567089	0.5287275466	0.516	51.60%	0.6985	69.85%
0.1484337399	2.465917393	0.5557	55.57%	0.9931	99.31

Here all the sample values pass the two tail non-directional test where except one highleted in red

We can concluse this from the % values also as most of the values lies in medum region
one values has a % of 99.31 which lies in the alpha region of 5%

Hence we accept the null Hupothesis as the trend continues for all the mean values sets compared

Q5. 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?

Ans

Two tailed (non-directional testing was performed on the above sample)

Following steps were performed separately for the congruent and in-congruent words

Step-1 here we randomly selected sample size of 5 and calculated the means for each sample size

Step-2 Now we plotted the curve distribution for these mean values which was a normal curve as expected by the **Central Limit Theorem**

Here we can approximate that the curves we obtained from the means are some what Normal Curves
Here both the Congruent and In-Congruents Histogram plots form a Normal Distribution in accordance with the Centeral Limit Theoram

Plot for randomly picked samples

Randomly Picked mean values/Congru...

Randomly Picked mean values/In-Congrue...
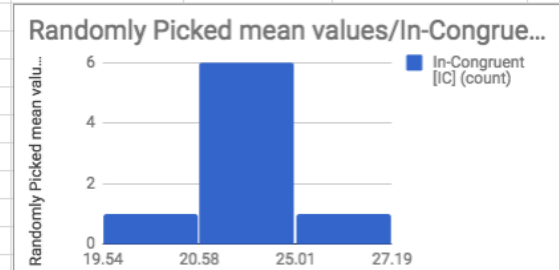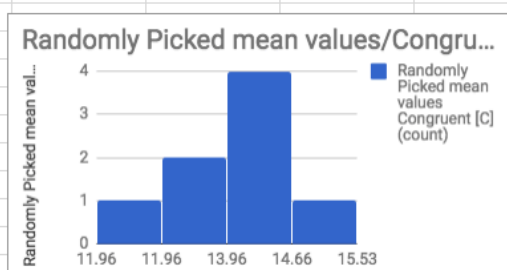
Step-3 Now on the basis of the Central Limit Theorem we calculated the Standard error for the mean values.

Step-4 Now we calculated the Z-score for the following means (Refer to the spread sheet for detailed information)

Step-5 Now we randomly selected 8 mean values from our set of mean values and performed Two Tailed Testing (Non-Directional testing) for α = 0.05

And corresponding (Critical Z value)Z* = 1.96

Here we also notice that the mean of mean distribution is same as population mean

Tests:-

1. Null Hypothesis :-
    $$\mu = \mu I$$
The Trend Remains same :-Time Taken to recognise Congruent words remains lower than that of the In-congruent words

2. Alternative Hypothesis :-
    $$\mu \neq \mu I$$
The Trend Reverses :- Time Take to recognise the Congruent Words was higher than that of the In-Congruent Words

Results:-

Here after Performing the Two tail testing on the Congruent and In-congruent Word recognition time sample of randomly selected users

We conclude that they depict the same trend of (Time Taken to recognise Congruent words remains lower than that of the In-congruent words)

And all of these mean values are well inside the 95% Zone based on the Two tail testing hence they occurred by chance which can be deduced by looking their **P values [Ratios or Percentage Values]**

Hence we fail to disprove the Null Hypothesis

Conclusion :-

Hence we Conclude that it takes less time recognise Congruent words(Words named and coloured the same) while it takes a bit more time to recognise In-Congruent Words.

* Please refer to the spread sheet present in the same folder of submission

Q6. 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!

Ans:-
1 .The ability of our brain to create a image for remembering different objects is the cause since on looking at a colour our brain can response very quickly due to perceptive memory while for processing the same color name on different color takes time

2. Yes jumbled up images and words ie if we display a image of a ship and call it a tree or vice versa it should show similar kind of impact as here also our brain relies on perceptive memory

3. It can also be done by interchanging the names of two people for a day and then see the difference your selves !!

Resources
1. https://keydifferences.com/difference-between-null-and-alternative-hypothesis.html
2. Wikipedia:- https://www.wikipedia.org/

Please refer the spread sheet for the values and results

Spread sheet can also be accessed here

https://docs.google.com/spreadsheets/d/1zMsd9yFfl3KSY2NDLIH5oglNnkGiZGjxr_SauzPHqlc/edit?usp=sharing