P1: Test a perceptual Phenomenon

1. What is our independent variable? What is our dependent variable?
   Answer -
   Independent variable: the chosen number of **participants** (subject), and treatment:
   the list (chosen size) of **congruent words** (chosen words),
   the list (chosen size) of **incongruent words** (chosen words)
   Dependent variable: **time** it takes to name the ink colors in equally-sized lists of congruent words and incongruent words

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.
   Answer -
   μ_c = *Mean (read time) for the list of Congruent words*
   μ_i = *Mean (read time) for the list of Incongruent words*
   Null hypothesis H_0: $\mu_c \geq \mu_i$ One-tailed directional statistical hypothesis
   Alternative hypothesis H_a: $\mu_c < \mu_i$ One-tailed (directional) statistical test
   We use one-tailed directional statistical test to either prove or reject the Null hypothesis, we choose less conservative directional test vs. conventional two-tailed test, because I am aware that incongruent words are harder to make right (in my own experiment), and there is clear indication of this from the sampling data.

   This is a perfect case for **dependent t-test for paired samples**. Reason being:
   - We don't know population mean μ or standard deviation σ
   - We have only 2 within subject (same group of person) sample data set on 2 different conditions, Congruent words list OR Incongruent words list.

3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.
   Answer -
   To descriptive statistics (M, SD), Mean or μ is a typical measure of central tendency, even though the mean, median and mode are all valid measures of it.
   Likewise standard deviation SD or σ is a typical measure of variability, even though 3 other frequently used measures of variability: the range, IQR and variance.
   Result Sections:
   Sample mean of Congruent: $\overline{X}_C = \sum_{i=2}^{25} A_i$ see A_i in table column A below
   Sample mean of Incongruent: $\overline{X}_I = \sum_{i=2}^{25} B_i$ see B_i in table column B below
   Sample mean of Difference (Congruent$_i$ - Incongruent$_i$): $\overline{X}_D = \sum_{i=2}^{25} C_i$, C_i in table column C
   we have: Mean = $\overline{X}_D$ = ***-7.964791667*** = $\overline{X}_C - \overline{X}_I$, and
   Bessel's Correction Standard Deviation for 24 samples SD=stdev(c2:c24) = ***4.86482691***

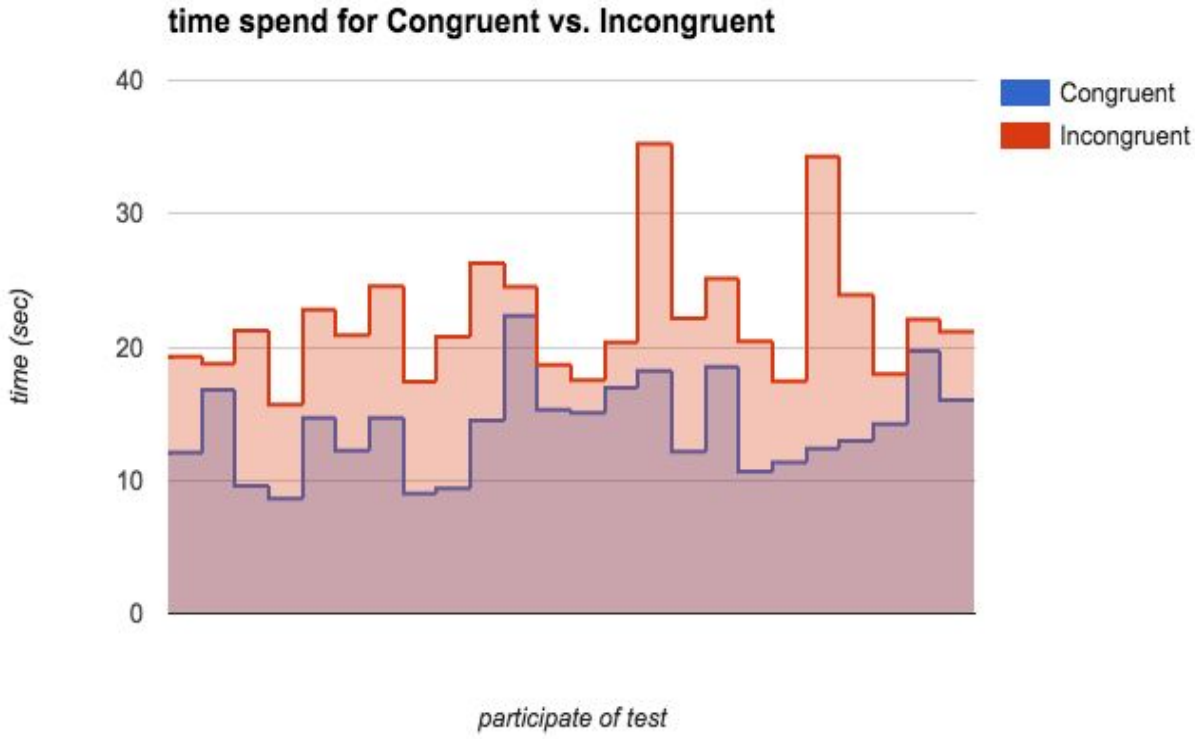
fx =stdev(C2:C25)

	A	B	C	D	E
1	Congruent	Incongruent	Congruent - Incongruent	Mean=Average(c2:c25)	SD=stdev(c2:c25)
2	12.079	19.278	-7.199	-7.964791667	4.86482691
3	16.791	18.741	-1.95		
4	9.564	21.214	-11.65		
5	8.63	15.687	-7.057		
6	14.669	22.803	-8.134		
7	12.238	20.878	-8.64		
8	14.692	24.572	-9.88		
9	8.987	17.394	-8.407		
10	9.401	20.762	-11.361		
11	14.48	26.282	-11.802		
12	22.328	24.524	-2.196		
13	15.298	18.644	-3.346		
14	15.073	17.51	-2.437		
15	16.929	20.33	-3.401		
16	18.2	35.255	-17.055		
17	12.13	22.158	-10.028		
18	18.495	25.139	-6.644		
19	10.639	20.429	-9.79		
20	11.344	17.425	-6.081		
21	12.369	34.288	-21.919		
22	12.944	23.894	-10.95		
23	14.233	17.96	-3.727		
24	19.71	22.058	-2.348		
25	16.004	21.157	-5.153		

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.

Answer -

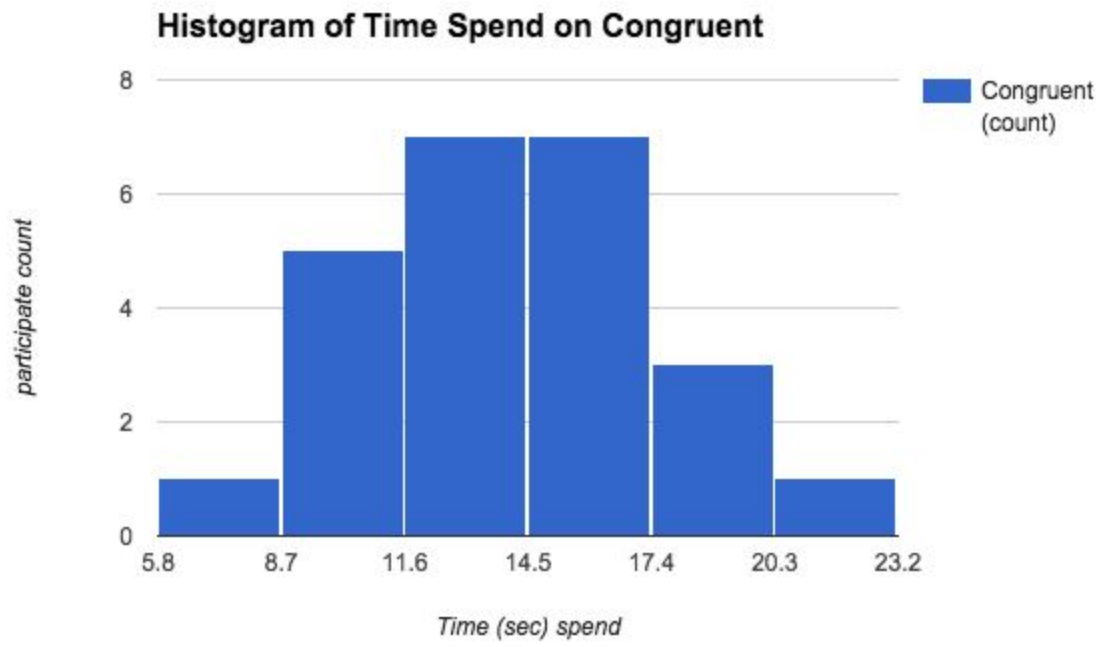
Plot 1. below is a bar graph vs. histogram. It depicts the absolute time it takes for congruent vs. incongruent by each participant (x-axis), all time starts from 0 in y-axis. Higher outlines from incongruent bar charts clearly tell that it takes more time for people to go through incongruent words list vs. congruent.

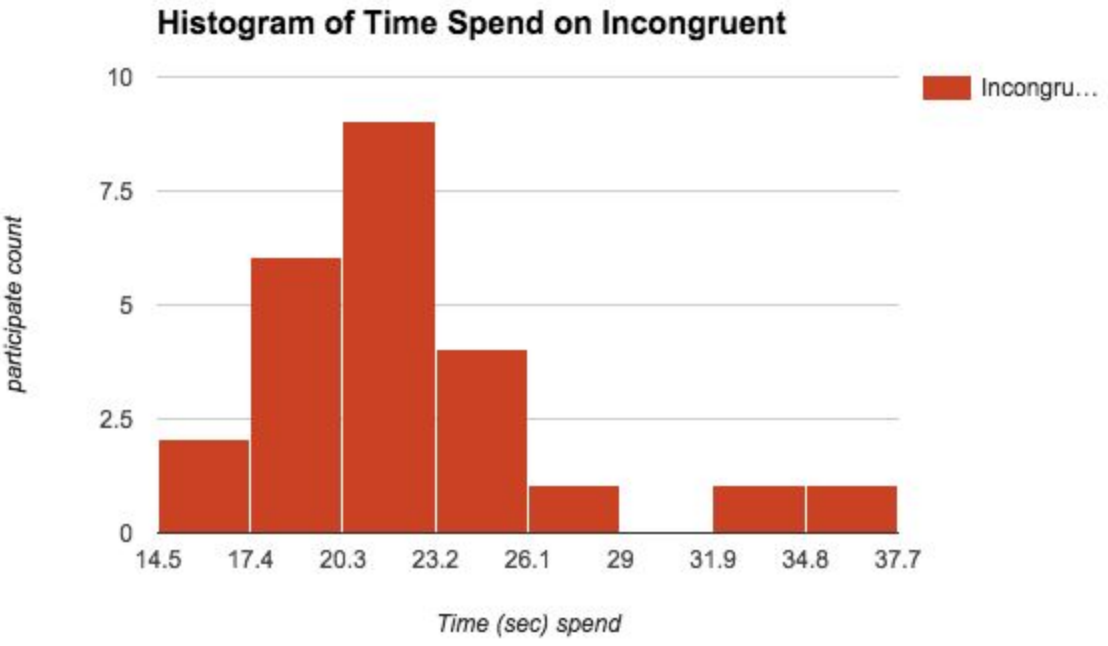


Plot 2. Histogram, I found it is possible to plot back-to-back histogram (as excel) or 2 histograms on the same chart using current google sheets. So I plotted two histogram charts: Congruent and Incongruent respectively.

I used the same bucket size: 2.9 in x-axis for both histograms, and I placed them up&down together, look forward to seeing some statistical difference between them easily:

Attention: x-axis on both histograms below is not aligned.





From these 2 histograms, we can assume they are likely (not so sure due to small sample data size) symmetric t-distribution. Importantly the central tendency of ‘Incongruent’ sample set is at higher value (20.3 to 23.2 seconds) than ‘Congruent’ (~14.5 seconds) samples. This indicates for the same subjects (participants), read ‘incongruent’ words take longer time than ‘congruent’. Next we run some inferential statistics test to either reject or fail to reject this hypothesis (in 2.)

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?
Answer- statistical test: dependent t-test for paired sample with one-tailed directional hypothesis
H_0: $\mu_c \geq \mu_i$, H_a: $\mu_c < \mu_i$ (M_D , S_D below are from 3.)
df = n-1 = 23, M_D = -7.964791667, S_D = 4.86482691, **$t_{statistics} = \frac{M_D}{S_D/\sqrt{n}}$ = -8.020707**
Choose a very high (more conservative) confidence level α = 98%,
using t-table, at df = 23, we have **$t_{critical}$ = -2.50**
In APA style, we have
t(23) = -8.02, p < .0001, one-tailed.
Since t(23) << **$t_{critical}$** ,by conventional criteria, this difference is considered to be extremely statistically significant. we have a very much smaller probability p to support H_0, so we reject the null hypothesis. In other words the statistics test results do NOT match up with H_0 expectations. We are confidence to say: population $mean - \mu_c$ from congruent test is smaller than population $mean - \mu_i$ from incongruent test.
In addition, we have ***Cohen's d*** = $\frac{M_D}{S_D}$ **= -1.637**, and (didn't compute Confidence Interval as not asked) $r^2 = t^2/(t^2 + df) = 73.66\%$, so ~74% of the difference in time is contributed to Congruent vs. Incongruent.

6. 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!
Answer -
$r^2 = 73.66\%$ tell us that the word ink color type Congruent vs. Incongruent is mainly responsible for the effects observed, at a ~74% coefficient of determination. Other than this Stroop task, another similar dependent t-test for paired samples would be to change the test order: i.e. read loud Incongruent list words first, than Congruent list words, everything else is controlled the same.
Furthermore, we can also design an experimental Independent t-test for two samples from two different group of participants (between-subject designs), e.g. we can have 100 people only test with Congruent words vs. another 100 people only test with Incongruent words. Similar effect may be observed.

Thanks for your review!

P.S.

Tools and software used:

google docs, google sheets, google charts
graphpad.com (use it to compute p from t and df, in question 5.)
t-table online at https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg