Brian Chase
June 16th, 2017

Statistics: The Science of Decisions Project Answers

1. The independent variable is whether the ink colors are congruent or incongruent with the words. The dependent variables is the time it takes to name the ink colors in equally-sized lists.

2. $H0: \mu_I = \mu_C$

   The **null hypothesis** (H0) states that there is no difference between incongruent and congruent population mean times.

   μ_I denotes incongruent population mean and μ_C denotes congruent population mean.

   $Ha: \mu_I \neq \mu_C$

   The alternative hypothesis states that there is a statistically significant change (either positive or negative) between incongruent and congruent population mean times.

   I'd expect to perform a two-tailed test since the direction of the change is unknown and both directions (faster or slower) would be an interesting conclusion.

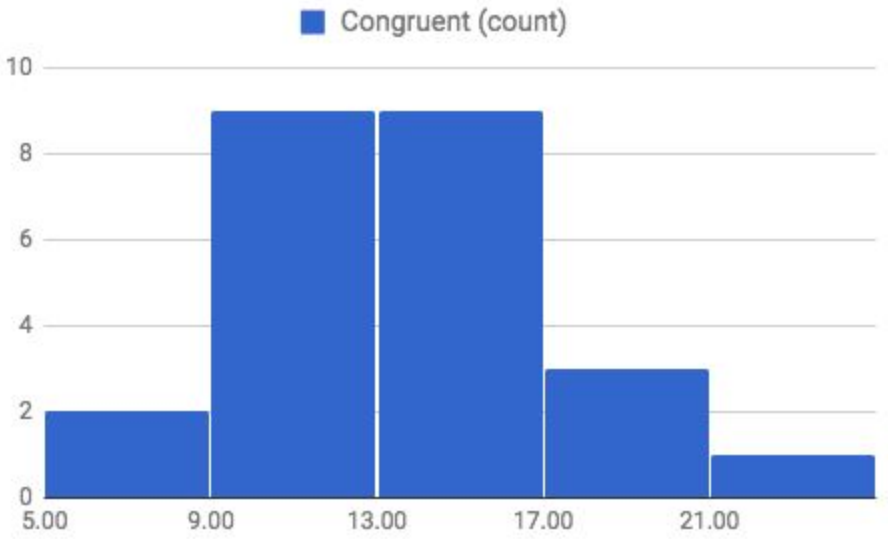
   Since the population standard deviation is unknown (and sample size < 30), I chose to perform a two-tailed t-test with two dependent samples.

   The reason each pair of congruent and incongruent time entries are dependent on each other is because they are related by the person performing them. Each person does both the congruent and incongruent timed challenge.

3. The average of the times to name the ink colors are lower for the congruent case at an average of 14.625 seconds. And the standard deviation is lower at 3.914 for the congruent case---meaning there was more variation in times for the incongruent case.

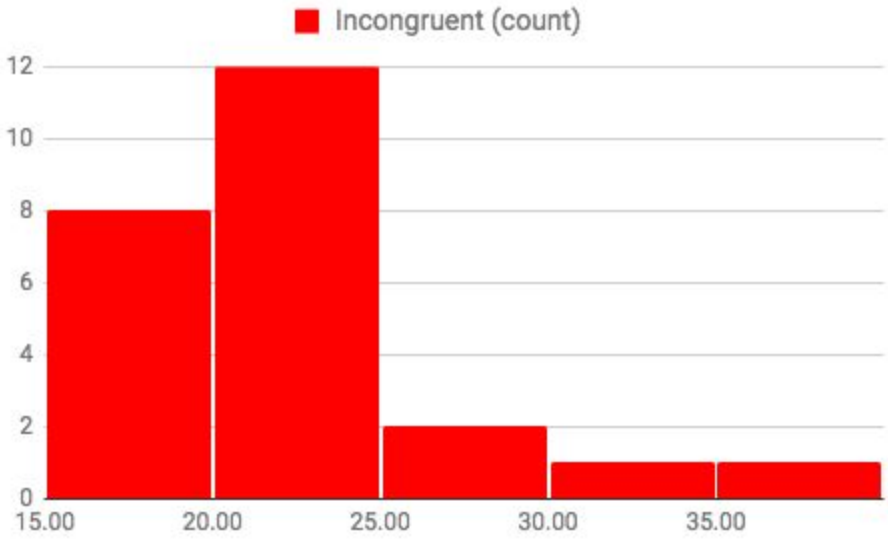
	Congruent	Incongruent
Mean	14.051125	22.01591667
StdDev	3.559357958	4.797057122

4. For the congruent distribution, the data looks normally distributed around the mean with a small tail for the longer time



s.

For the incongruent case, there is a longer tail of values for the larger values past 25 seconds--with most people in the 15-25 second bins.



5. I chose to perform a two-tailed t-test since the result could be a slower or faster time. The samples are also dependent on each other, since it is a person doing two different tests. The result of the test is that we reject the null hypothesis---information provided below.

t-statistic = 8.020706944
t-critical value (at 99.9% confidence level / alpha (two-tailed) = 0.001) **= +-3.768**
P-value = 0.00000004

Since the t-statistic > t-critical value, we reject the null hypothesis. The calculated p-value suggests that this result could occur due to sampling error 0.000004% of tests giving great confidence that the effect is real.

I would then infer that having incongruent coloring of words would take longer to name the ink colors than congruent coloring of words.

From a cursory glance at the data, I would have expected it to be statistically significant. Seems so obvious after the fact---and yet I did not know about the effect before doing the exercise.

Sources used

https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg

https://www.ck12.org/book/CK-12-Probability-and-Statistics-Advanced-Second-Edition/section/8.5/

http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-correctly-interpret-p-values

http://www.danielsoper.com/statcalc/calculator.aspx?id=8

https://en.wikipedia.org/wiki/Stroop_effect