Data Analyst

P1 project

JuHyung Son

I got stroop data from Udacity

https://www.google.com/url?q=https://drive.google.com/file/d/0B9Yf01UaIbUgQXpYb2NhZ29yX1U/view?usp%3Dsharing&sa=D&ust=1473405265697000&usg=AFQjCNE8kXIU2zcQlqgL1fJ2bqUTuHB8Ww

In this data, there are 2 columns, congruent and incongruent and 24 rows.

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

Independent variable is whether it is congruent or incongruent.

Dependent variable is the time that takes to name the color of inks.

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

null hypotheses is $H_0: \mu_{con} = \mu_{incon}$

alternative hypothesis is $H_1: \mu_{con} < \mu_{incon}$

μ_{con}=the population mean of the time in congruent test.

μ_{incon}=the population mean of the time in incongruent test.

Null hypothesis means that the population mean of the time to take stroop test in congruent and incongruent is same. In other words, there`s no difference between 2 tests.

Alternative hypothesis means that the population mean of the time to take stroop test in congruent is shorter than test in incongruent.

In this test, paired t-test is appropriate.

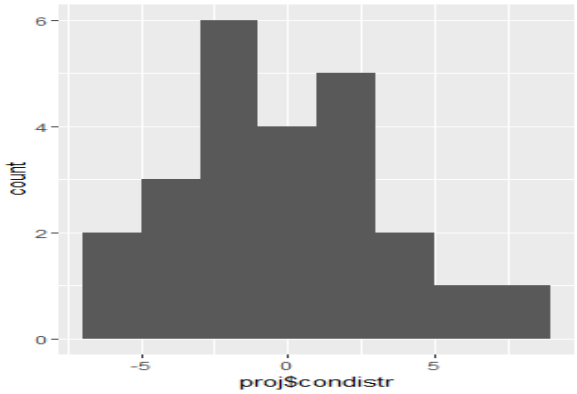
Because each person take the test twice and compare it. So this is dependent sample.

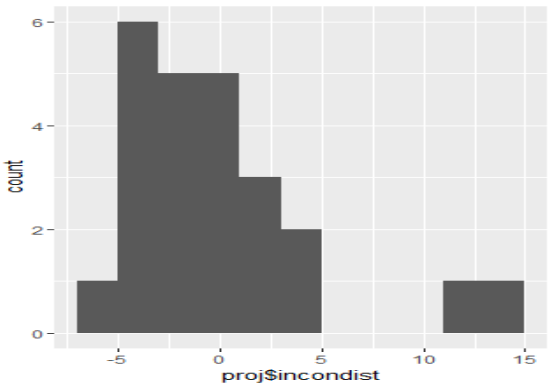
The reason why t-test is appropriate than z-test is z-test is used when we know population parameter. But here, we don`t know. And t-test is appropriate when we don`t know population parameter.

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

	Congruent	Incongruent
Mean	14.05	22.01
Mean of difference	7.96	
Sd	4.86	
se	0.993	

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.





This is histogram of error in samples.

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?

Difference between the two observation: con-incon
-7.199 -1.950 -11.650 -7.057 -8.134
-8.640 -9.880 -8.407 -11.361 -11.802
-2.196 -3.346 -2.437 -3.401 -17.055
-10.028 -6.644 -9.790 -6.081 -21.919
-10.950 -3.727 -2.348 -5.153

Mean difference: -7.96
Sd of differences: 4.86
Se: 0.993
t-statistic: -8.016
t-critical=-1.714
p-value=2.052e-08
95% CI: (-6.257,-9.662)

I can see t-statistic<t-critical so I reject the null hypothesis.
And this is what I expected! The population mean of the time in incongruent test is longer than congruent test.

Reference
Udacity: Statistics