The stroop effect

LinhNguyen

Monday, September 05, 2016

### The library:

* RCurl
* ggplot2
* gridExtra
* tidyr

### Background Information

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant’s task is to say out loud the color of the ink in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the congruent words condition, the words being displayed are color words whose names match the colors in which they are printed: for example RED, BLUE. In the incongruent words condition, the words displayed are color words whose names do not match the colors in which they are printed: for example PURPLE, ORANGE. In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

### The data:

The original data is stored in wide format, with 2 collumn "Congruent" and "Incongruent". Let's remove the capital character first:

## congruent incongruent  
## 1 12.079 19.278  
## 2 16.791 18.741  
## 3 9.564 21.214  
## 4 8.630 15.687  
## 5 14.669 22.803  
## 6 12.238 20.878

For the sake of flexibility, we shall add another format, a long one.

## type time  
## 1 congruent 12.079  
## 2 congruent 16.791  
## 3 congruent 9.564  
## 4 congruent 8.630  
## 5 congruent 14.669  
## 6 congruent 12.238

### The variables:

Independent variables:

* 1 measurement variable: the time it took to read a list of words.
* 1 nominal variable: type of words (congruent or incongruent).

Dependent variables: The time it takes to name the ink colors of a list of words.

### The hypothesis and the test:

The hypothesis:

Null hypothesis: There is no significant different between the means of the time it took to read a list of words for the two Congruent and Incongruent categories.

H0: μ1 = μ2

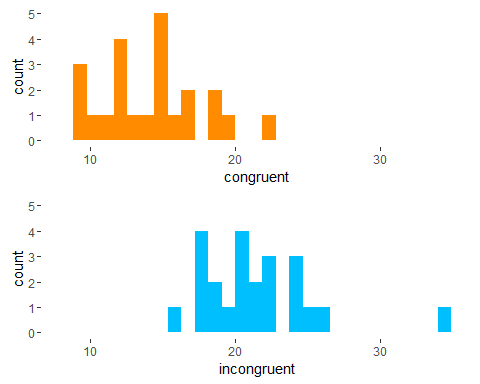
Alternative hypothesis: There is a significant different between the means of the time it took to read a list of words for the two Congruent and Incongruent categories.

Ha: μ1 != μ2

The proposed statistical test:

It is essential to look at the data (if you already have), or imagine how it would look like, before deciding on the kind of statistical test:

incongruent\_his <- ggplot(aes(x = congruent), data = sef) +   
 geom\_histogram(fill = "darkorange") +  
 theme(panel.background = element\_rect(fill = 'white'),  
 plot.background = element\_rect(fill = 'white')) +  
 xlim(8,35) +  
 ylim(0,5)  
congruent\_his <- ggplot(aes(x = incongruent), data = sef) +   
 geom\_histogram(fill = "deepskyblue") +  
 theme(panel.background = element\_rect(fill = 'white'),  
 plot.background = element\_rect(fill = 'white')) +  
 xlim(8, 35) +  
 ylim(0,5)  
  
grid.arrange(incongruent\_his, congruent\_his)



* The sample size: balanced design, 24 for each sample, considered as small.
* The distribution: light right-skewed for both sample

Based on the sample size, distribution, and number of variables, it is fitting to use Student's t–test for two samples, with the proven rate of false positive less than 5% even for very small sample, as long as the homogeneity of variance is true.

### Descriptive statistics

We could see it quite clearly with this one:

* The mean and the median are almost identical for the congruent, and varied a little bit for the incongruent (because of the 2 outliers, we do expec them to get closer, when removed).
* The min and max value from the 2 sample are quite far apart.

summary(sef)

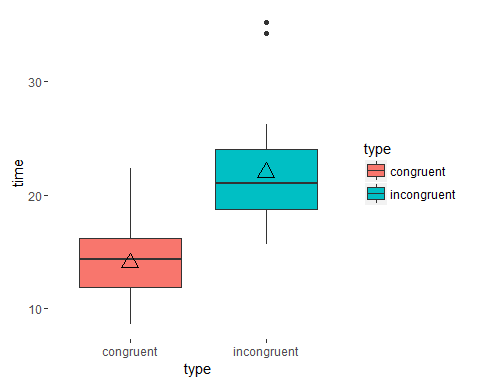
## congruent incongruent   
## Min. : 8.63 Min. :15.69   
## 1st Qu.:11.90 1st Qu.:18.72   
## Median :14.36 Median :21.02   
## Mean :14.05 Mean :22.02   
## 3rd Qu.:16.20 3rd Qu.:24.05   
## Max. :22.33 Max. :35.26

### The plot

The plots better visualized our point in the above section:

* The incongruent box plot are affected by the outliers (not heavy enough to remove it).
* The mean and median are close to each other.
* The 2 boxes are in 2 distinct different position.

ggplot(aes(x = type, y = time, fill = type), data =seflong) +  
 geom\_boxplot() +  
 stat\_summary(fun.y=mean, geom="point",shape=2,size=4,fill="yellow") +  
 theme(panel.background = element\_rect(fill = 'white'),  
 plot.background = element\_rect(fill = 'white'))



### The statistical test

t.test(time ~ type, data=seflong, var.equal=TRUE, conf.level=0.95)

##   
## Two Sample t-test  
##   
## data: time by type  
## t = -6.5323, df = 46, p-value = 4.595e-08  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -10.419121 -5.510462  
## sample estimates:  
## mean in group congruent mean in group incongruent   
## 14.05113 22.01592

Test result:

* t value: -6.5323
* df: 46
* p value: 4.595e-08
* Reject Null hypothesis.

Alternatite hypothesis: True.

There is a significant different between the means of the time it took to read a list of words for the two Congruent and Incongruent categories.

Final note:

Personally, I would see this test as part of the effort to study the power of the brain, which is the relationship between complexity of the information intake (incongruent words) and the speed of process (time it took to call the colour)