Shirley Ang, Gavin Hatch, Gabe Hoose, Baylee Willis
Professor Christensen
STAT 230
December 8, 2021

Stroop Strife

Introduction

For our project, we are interested to see if college students from different major groups (Education, STEM and Other majors) will have a difference in their response time (measured in seconds) in the Stroop Task. Each participant will be asked to complete either Test A first, which involves naming the color of the font of each word as fast as possible, or Test B first, which involves reading the word regardless of the color of the font as fast as possible. These tasks will be explained in detail in the next section. Each person will be asked to complete each task once. We are interested in understanding whether there is a significant difference in response times between different majors. We expect that non-STEM majors will have faster response times and that STEM majors will perform worse on the overall task. However, we also expect there to be an interaction between test and major such that Test A will be easier for non-STEM majors and Test B will be easier for STEM majors.

Design and Data Collection

The experiment design that we used was the repeated measures design with one between block factor and one within block factor (SP/RM[1:1]). The between factor used was the type of majors (Education, STEM or Other) and the within block factor used was the type of test (Test A or Test B). The between block factor is randomly assigned to each participant at the different times they took the test.
This design measures the effect of blocking for person, major type, test type and the interaction between the major type and the test. The resulting statistical model is $y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{jk} + \varepsilon_{ijk}$; where μ is the grand mean of all the observations, α_i is the effect of the type of major (where I =3 where i=1 is education, i=2 is STEM i=3 is others), $\beta_{j(i)}$ is the effect of blocking by person (where j=1…19 and J=19) where the person is nested in the major, γ_k is the effect of the test (where K=2 where k=1 is Test A and k=2 is Test B), $(\alpha\gamma)_{jk}$ is the interaction between the two, and $\varepsilon_{ijk}$is the remaining error. Using this model we can predict the expected time for each of the given conditions.

Participants ← Major type
|
Time of the person's life ← Test type

Hypothesis 1: Education vs STEM vs Other
H0: the mean time effect for the different majors are the same.
H0: $\alpha_1=\alpha_2=\alpha_3$
HA: the mean time effect for different majors is not the same
HA: at least one α is different

Hypothesis 2: Test A vs Test B
H0: the mean time effect for the different tests is the same.
H0: $\gamma_1=\gamma_2$
HA: the mean time effect for different tests is not the same

HA: $\gamma 1$ does not equal $\gamma 2$

Hypothesis 3: Interaction between major and test
H0: The major and test factors do NOT interact.
H0: $\alpha\gamma 11 = \alpha\gamma 12 = \ldots \alpha\gamma 32$
HA: The major and test factors DO interact.
HA: at least one $\alpha\gamma$ is different

As mentioned, the 2 factors are the major of the participant and the type of test that the participant took. In total there will be 6 different combinations. We have 19 test subjects in total, the block being the person. To apply randomization for which test to be taken first, we flipped a coin to decide the order for that factor. If the results were heads, the subject will take Test A first then Test B. If the coin landed on tails, the order of the test would be to take Test B first then Test A.

The Stroop Task that we used for this experiment consisted of two separate tests, Test A and Test B. Each test had ten words, all of which were different names of colors written in a font color different from the written word. For Test A, the participants were asked to name the color of the font as quickly as possible and ignore the word that was written on the slide. During Test B, the participants were asked to read the word aloud as quickly as possible while ignoring the color of the font. For example, if the word "blue" came up on the screen in yellow font, the correct answer for Test A would be "yellow" while the correct answer for Test B would be "blue." Each participant was randomly assigned either to complete Test A or Test B first and then were given instructions on how to complete the assigned test. The test was administered using Google Slides, with test instructions preceding the test and one word prompt per slide, and the researcher conducting the experiment was only allowed to move to the next prompt if the participants gave the correct answer. Each researcher began the timer once the first word was shown on the screen and stopped the timer when they correctly gave the answer to the final prompt.

White

In the above example taken from our stroop task, participants would be expected to answer "purple" when taking Test A and "white" when taking Test B.

Data Analysis

Our CAZSIN assumptions have been met. Specifically, S (that standard deviations are the same) and N (normality of residuals). No transformation is needed.

Error: Name

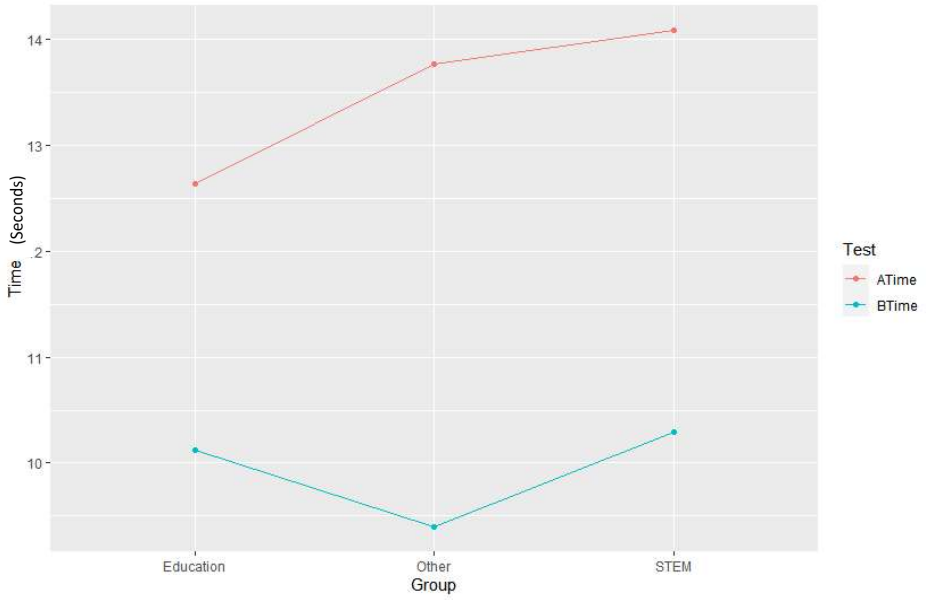
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	2	4.13	2.067	0.223	0.802
Residuals	16	148.09	9.256		

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Test	1	130.06	130.06	75.424	1.88e-07 ***
Group:Test	2	5.33	2.66	1.544	0.244
Residuals	16	27.59	1.72		

The major that each participant had did not significantly affect the time at which they completed the stroop task. Test type significantly impacted the time at which the participant completed the stroop task.

The mean time to complete Test A was 13.57 seconds and the mean time to complete Test B was 9.87 seconds.



Conclusions

We used an alpha level of .05 to determine statistical significance. The major group factor resulted with F-statistic of .223 and a p-value of .802 which is not statistically significant. There is not enough evidence that the major group has an impact on the speed of completing the stroop task.

The test type factor produced an F-statistic of 75.424 and a p-value of 1.88e-07 which is statistically significant. There is evidence that the type of test taken has an impact on the speed of completing the stroop task.

The interaction between major group and test type produced an F-statistic of 1.544 and a p-value of .244 which is not statistically significant. There is not enough evidence to claim an interaction between the major group and test type.

We found that major type does not seem to have an impact on the speed of completing the stroop task. However, the type of test does seem to impact the speed that the test is completed. It appears that the test which asked participants to read the word and ignore font color was completed faster than the test where participants were asked to identify the font color and ignore the written word. We can generalize our findings to college students in Utah Valley. A future study can be refined and expanded by having a larger sample size, having longer tests, having more spread of major and more specific major groups, and we could have added more factors as well like environment type, age, socioeconomic status, medical status (ADHD, Dyslexia, etc.).

Appendix

Name	Major	Test A Time	Test B Time	First Test	Group
Mary	Advertising	14.78	6.45	B	Other
Rollin	Statistics	10.59	8.64	B	STEM
Natalie	Art History	12.62	7.62	A	Other
Jillian	Business Education	12.01	8.51	A	Education
Erika	English	13.37	10.06	A	Other
Emily	Family Studies	11.69	8.06	A	Other
Nathan	Information Systems	12.86	10.01	B	STEM
Shelby	Elementary Education	7.03	6.82	B	Education
Delbert	Civil Engineering	15.54	9.63	B	STEM
Virginia	Psychology	15.37	11.28	A	STEM
Kelly	Dietics	15.17	8.99	A	STEM
Min Yong	Chemistry Education	12.3	9.84	A	Education
Alice	Culinary	13.7	7.84	B	Other
Hannah	Math Education	17.56	14.5	A	Education
Emma	English Education	14.3	10.95	A	Education
Sarah	Linguistics	14.02	11.55	B	Other
Mikinna	Marketing	13.23	11.25	B	Other
Jacob	Psychology	15.01	13.23	A	STEM
Nick	History	16.73	12.35	B	Other

Stroop Test Term Project Code:

```
#download necessary libraries (tidyverse and DescTools)
library(tidyverse)
library(DescTools)

stroop <- read_table("Name      Major      ATime  BTime Order Group
Mary Advertising 14.78 6.45     B      Other
Rollin   Statistics        10.59   8.64    B       STEM
Natalie  ArtHistory       12.62   7.62    A       Other
Jillian  BusinessEducation      12.01   8.51    A       Education
Erika    English 13.37   10.06   A       Other
Emily    FamilyStudies  11.69   8.06    A       Other
Nathan   InformationSystems     12.86   10.01   B       STEM
Shelby   ElementaryEducation    7.03    6.82    B       Education
Delbert CivilEngineering        15.54   9.63    B       STEM
VirginiaPsychology      15.37   11.28   A       STEM
```

```
Kelly    Dietetics          15.17   8.99    A       STEM
MinYong          ChemistryEducation      12.3    9.84    A       Education
Alice    Culinary           13.7    7.84    B       Other
Hannah MathEducation 17.56  14.5    A       Education
Emma     EnglishEducation           14.3    10.95   A       Education
Sarah    Linguistics        14.02   11.55   B       Other
Mikinna          Marketing          13.23   11.25   B       Other
Jacob    Psychology         15.01   13.23   A       STEM
Nick     History 16.73   12.35   B       Other
")

#check how it read in
stroop

#gather the data by Test type so that each experimental unit is listed twice (once for Test A time and once for Test B time)
stroopNew <- stroop %>%
  gather(ATime, BTime, key="Test", value="Time")
#check how it looks
stroopNew

#get anova tables and summaries
stroop.lm <- aov(Time ~ Group + Error(Name) + Test + Group:Test, data=stroopNew)
summary(stroop.lm)


#create an interaction plot
stroopNew.int <- stroopNew %>%
  group_by(Group, Test) %>%
  summarise(Time = mean(Time))
stroopNew.int %>%
  ggplot() +
  aes(x = Group, y = Time, color = Test) +
  geom_line(aes(group = Test)) +
  geom_point()

#means of each test type time
mean(stroop$ATime)
mean(stroop$BTime)



#the following code organizes the data to get standard deviations to check CAZSIN assumptions
OtherA <- stroopNew %>%
```

```
  filter(Group == "Other") %>%
  filter(Test == "ATime")

OtherB <- stroopNew %>%
  filter(Group == "Other") %>%
  filter(Test == "BTime")

STEMA <- stroopNew %>%
  filter(Group == "STEM") %>%
  filter(Test == "ATime")

STEMB <- stroopNew %>%
  filter(Group == "STEM") %>%
  filter(Test == "BTime")

EducationA <- stroopNew %>%
  filter(Group == "Education") %>%
  filter(Test == "ATime")

EducationB <- stroopNew %>%
  filter(Group == "Education") %>%
  filter(Test == "BTime")

sd(OtherA$Time)
sd(OtherB$Time)
sd(STEMA$Time)
sd(STEMB$Time)
sd(EducationA$Time)
sd(EducationB$Time)

#largest SD divided by smallest
3.838834/1.509472
#less than 3

#histograms of residuals to determine normality
hist(stroop.lm$Name$residuals)
hist(stroop.lm$Within$residuals)
#both are approximately normal

#therefore, no transformations are needed
```