Experimental Methods I - portfolio 1

Sebastian Scott Engen

9/26/2017

library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

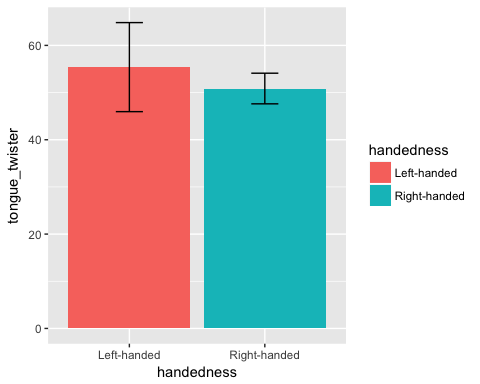
# We use the library function to Load the tidyverse package that have previously been installed from the online CRAN database. This allows us to use the GGPLOT2 functions  
  
setwd("/Users/FlowersnIce-cream/Google Drev/Hogwarts/R Studio/portfolios/Portfolio")  
# We set the working directory (WD for short) to the folder that contains our dataset  
  
load("2\_clean\_cogsci\_data.Rda")  
# Now we use the load function to get our RDA file with data about the classes personalities into RStudio.

# Portfolio 1: Datamining of the CogSci 2017 Personality Data

## 1) Who, on average, is faster at the tongue twister task, right-handers or left-handers?

(please provide numeric, visual, and verbal answer)

# We start by creating a new object called Faster\_Hands -> then we designate it's variables by first entering the 'personality\_data'  
# and then selecting and grouping all the entries in the handedness variable, which will enable us to  
faster\_hands <- personality\_data %>%  
 group\_by(handedness) %>%  
 summarise(tongue\_twister = mean(tongue\_twister),count = n())  
  
  
#handedness tongue\_twister count  
#Left-handed 55.39500 6  
#Right-handed 50.81698 53  
  
 # summarise the mean from the tongue twister variable for left handed people and right handed seperately. We finish off our DataFrame (DF for short) by adding a varible called count that gives us an insight to how many people are left and right handed seperately.   
 # with this data we now see that there are very little left-handers, which means that our mean for the lefthanders might me skewed.  
#Now we'll enter the personality data again -> not our DF 'Faster hands', since we want to have all the values when we find the confidence levels:  
ggplot(personality\_data, aes(x = handedness, y = tongue\_twister, fill=handedness)) + geom\_bar(stat = "summary", fun.y = mean) + geom\_errorbar(stat = "summary", fun.data = mean\_cl\_boot, width = 0.2)

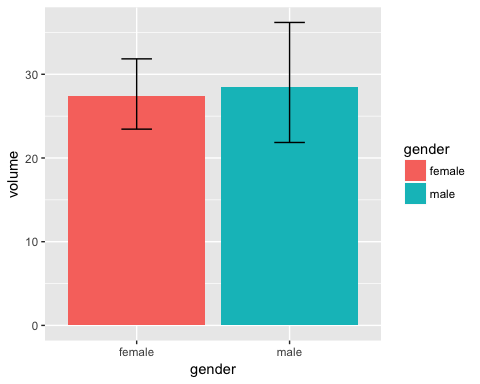


#Since the confidence levels overlap between left and right handers, we can't be sure wheter or not there is a significant difference between the means and thereby who in general is the fastest. This being said - The right handers was in this one event faster than the left-handers

## 2) Who prefers higher volumes of music, males or females?

(please provide numeric, visual, and verbal answer)

BoomBoom\_People <- personality\_data %>%   
 # We start by creating a new object called BoomBoom\_People -> then we designate it's variables by first entering the 'personality\_data'  
 group\_by(gender) %>%  
 # and then selecting and grouping all the entries in the gender variable, which will enable us to  
 summarise(volume = mean(volume), count = n())  
  
#gender volume count  
#female 27.44737 38  
#male 28.47619 21  
  
# summarise the mean from the volume variable for both genders. We finish off our DF by adding a varible called count that gives us an insight to how many people are from either gender  
 # with this data we now see that there are approx. the same amount of both genders - and we also see that the results look very similar, only varying by one point on the scale from 1 to a 100.  
#Now we'll enter the personality data and BooBoom\_People, since we want to have all the values when we find the confidence levels that'll gives us an insight to wether or not this difference of a single point is significant or not  
  
# graph code goes here  
ggplot(personality\_data, aes(x = gender, y = volume, fill=gender)) + geom\_bar(stat = "summary", fun.y = mean) + geom\_errorbar(stat = "summary", fun.data = mean\_cl\_boot, width = 0.2)

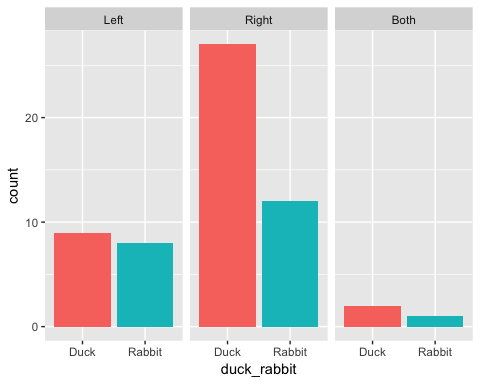


#Since the confidence levels overlap between the genders, we can't be sure wheter or not there is a significant difference between the means and thereby who in general likes the loudest music - men or women. This being said - the guys or "men" in this class at least, seem to like it a bit higher beats than their counterparts.

## 3) Is there a systematic relation between occular dominance and the rabbit-duck illusion?

E.g. does people with right occular dominance see the duck more than the rabbit? (please provide numeric and verbal answer. A graph is optional)

# code goes here  
Oc\_Dom\_SnurreSnub\_DonaldDuck <- personality\_data %>%   
 # We start by creating a new object called Oc\_Dom\_SnurreSnub\_DonaldDuck -> then we designate it's variables by first entering the 'personality\_data'  
 group\_by(ocular\_dominance, duck\_rabbit) %>%  
 # and then selecting and grouping all the entries in the ocular\_dominance and duck\_rabbit variable, which will enable us to  
 summarise(count= n())  
  
#ocular\_dominance duck\_rabbit count  
#Left Duck 9  
#Left Rabbit 8  
#Right Duck 27  
#Right Rabbit 12  
#Both Duck 2  
#Both Rabbit 1  
  
# use the summarise function to just show us the count of people voting the different variable outcomes.  
 # To see our results visually we make a bar graph, coloured by selection of duck\_rabbit and divided into graph-windows by handedness:  
ggplot(Oc\_Dom\_SnurreSnub\_DonaldDuck, aes(duck\_rabbit, count, fill = duck\_rabbit)) + geom\_bar(stat = "summary", fun.y = mean) + facet\_wrap(~ocular\_dominance) + theme(legend.position = "none")



# It quickly becomes clear that there aren't really any change in preferation in the left-handed group, but that there are a great preferation for Ducks amongst right-handers. The both-handed people seem to prefeer the ducks aswell, but since the group is so small, it's hard to draw any real conclusions.

## 4) Do we observe differences in variance (measured by standard deviation) in music consumption between those who have Danish as their native language and those who speaks a different native language?

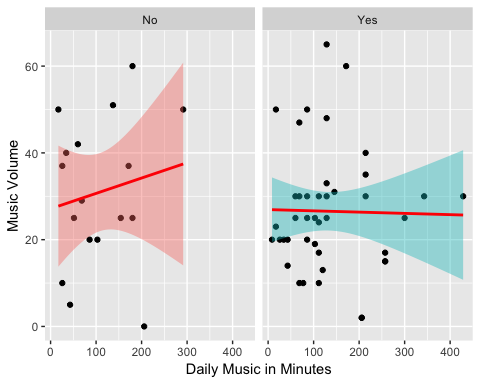
(please provide numeric and verbal answer. Optionally you can also provide a graph)

HeadbangingByNationality <- personality\_data %>%   
 # We start by creating a new object called HeadbangingByNationality -> then we designate it's variables by first entering the 'personality\_data'  
 group\_by(native\_speaker\_DK) %>%  
 # and then selecting and grouping all the entries in the native\_speaker variable, which will enable us to  
 summarise(Mean\_music = mean(hours\_music), Standard\_Deviation = sd(hours\_music), count = n())  
  
#native\_speaker\_DK Mean\_music Standard\_Deviation count  
#No 12.70000 12.09270 10  
#Yes 15.02041 10.15195 49  
  
# use the summarise function to just show us both, the number of native and unnative speakers, plus the mean listening time and the accompanying standard deviation for both.  
# The variance is lower for the native speakers compared to the non-native, but though this would suggest that the danes are more similar in their listening tendencies, it might just aswell be caused by the fact that there isn't as many non-native speakers. Any one outlier would have a bigger effect on the non-native speakers than the native.

## 5) Calculate the average music consumption per day for those who can touch the floor and those who can't. Do you observe a relation between music consumption and preference for volume in the two groups (just from looking at the data and graph)?

(please provide numeric, visual, and verbal answer)

music\_days <- mutate(personality\_data, music\_pr\_day = hours\_music/7\*60, yes\_no = ifelse(touch\_floor %in% c("Yes of course", "Yes"),"Yes", "No"))  
#We start of by mutating our data, which means that we add a new column, and add our new variable of daily music in minutes (where we take all the participants weekly hours of music listening and convert it into daily minutes) and also a variable where the anwsers in the touch\_floor variable is summed up into yes's and no's  
  
MusicManiacs\_ToDegreeOf\_FloorSweeping <- music\_days %>%   
 #Now we enter our new mutated dataset  
 group\_by(yes\_no) %>%  
 #Group our data by the yes\_no variable we just created  
 summarise(Daily\_Music\_Minutes = mean(music\_pr\_day), Vol = mean(volume), count = n())  
  
#yes\_no Daily\_Music\_Minutes Vol count  
#No 107.8992 30.94118 17  
#Yes 132.4490 26.54762 42  
  
#The we collapse our dataset and summarise the mean of our mutated variable music\_pr\_day and the volume variable. We also add a column for the number of people in either camp (Yes or No)  
 # By looking at our newly formed DF MusicManiacs\_ToDegreeOf\_FloorSweeping, we quickly see that those who can touch the floor listen to more music. Nevertheless we also see that those musicmaniacs who can't touch the floor tend to prefer higher volume than their peers. Glimpsing at the count variable, we must again notice that there are far more who can touch the floor than cannot, which do influence the spread of data and thereby our means (simply beacuse of bigger influence from outliers)  
#let's have a look at the actual spread:  
  
# graph code goes here  
ggplot(music\_days, aes(music\_pr\_day, volume, fill = yes\_no)) + facet\_wrap(~yes\_no) + geom\_point()+ geom\_smooth(method = "lm", colour = "Red")+ theme(legend.position = "none")+ labs(x = "Daily Music in Minutes", y = "Music Volume")



#First we draw a plot of our music\_days DF, and set the x-axis to music\_pr\_day and y-axis to volume, we use the facetwrap function to divide the data into two plots (one for yes and one for no). Then we add all the datapoints and use the geom\_smooth function to get a regressionline surrounded by a 95% confidence interval.  
#We now see clearly that the two confidence intervals overlap and that we thereby can't significantly distinguish them, furthermore we see the actaul spread and how there's a lot my points to cluster around the mean in the group that can touch the floor. Going beyond this, it seems that the preference for high volume has an inverse relationship in the no-group and the opposite in the yes-group, meaning that among the people who can't touch the floor the preference for loud music goes up as the minutes does, while for the people can touch the floor prefer to quiet it a bit down if they listen for longer periods.