## Homework 5

I start by setting the working directory to the source file location.

My first step is to read in the data.

dat<-read.csv("homework5.csv")

Then I can check the structure of the data.

str(dat)

## 'data.frame': 64 obs. of 2 variables:  
## $ music: Factor w/ 2 levels "metal","none": 2 2 2 2 2 2 2 2 2 2 ...  
## $ wpm : num 23.5 44.1 42.3 24.4 42.8 ...

By checking the structure I can see that there are two variables. The first is the independant variable of music, which has two levels: metal and none. The second is the dependent variable of words per minute.

Now I can find the mean words per minute for the two levels of the IV music.

mean(dat$wpm[dat$music=="metal"])

## [1] 48.03665

mean(dat$wpm[dat$music=="none"])

## [1] 42.44641

This reveals that the mean words per minute for the condition music:metal is 48.04, and the mean words per minute for the condition music:none is 42.45. The higher mean word per minute for the music:metal condition seems to indicate that participants in the condition music:metal typed more quickly than those in the music:none condition.

I can now run an independent t-test to find out if there actually is a difference in mean words per minute between the two levels of music metal and none.

t.test(dat$wpm[dat$music=="metal"],  
 dat$wpm[dat$music=="none"],  
 var.equal=TRUE)

##   
## Two Sample t-test  
##   
## data: dat$wpm[dat$music == "metal"] and dat$wpm[dat$music == "none"]  
## t = 2.3317, df = 62, p-value = 0.02298  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.7977514 10.3827312  
## sample estimates:  
## mean of x mean of y   
## 48.03665 42.44641

The results of the t-test reveal that there is a significant difference between the two conditions. Words per minute in the condition music:metal was higher than in the condition music:none t(62) = 2.33, p = .02298.

I can also run an independent t-test to determine if the alternative hypothesis is true: that participants in the condition music:metal type faster than the participants in the condition music:none.

t.test(dat$wpm[dat$music=="metal"],  
 dat$wpm[dat$music=="none"],  
 var.equal=TRUE,  
 alternative="greater")

##   
## Two Sample t-test  
##   
## data: dat$wpm[dat$music == "metal"] and dat$wpm[dat$music == "none"]  
## t = 2.3317, df = 62, p-value = 0.01149  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 1.586923 Inf  
## sample estimates:  
## mean of x mean of y   
## 48.03665 42.44641

The results of the test confirm that participants in the condition music:metal type faster than participants in the condition music:none t(62) = 2.33, p = .01149

I can create a bar graph to display the results by loading the ggplot2, gplots, and dplyr libraries.

library(ggplot2)  
library(gplots)

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

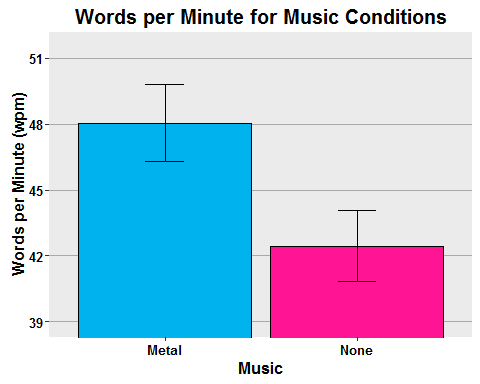
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

Then I can use dplyr to analyze the data. I can find the means and standard errors of the means and I can summarize the data for the different groups.

temp<-dat%>%group\_by(music)%>%  
 summarize(means=mean(wpm), sems=sd(wpm)/sqrt(length(wpm)))

I am able to create a figure from this summarized data.

f<-ggplot(temp,aes(x=factor(music),y=means))+  
 geom\_bar(stat="identity", color="black",  
 fill=c("deepskyblue2", "deeppink"))+  
 geom\_errorbar(aes(ymax=means+sems,  
 ymin=means-sems),  
 width=.2)+  
 ggtitle("Words per Minute for Music Conditions")+  
 labs(x="Music", y="Words per Minute (wpm)")+  
 scale\_x\_discrete(breaks=c("metal","none"),  
 labels=c("Metal","None"))+  
 theme(plot.title=element\_text(size=15,  
 face="bold",  
 vjust=.5))+  
 theme(axis.title.x=element\_text(size=12,  
 face="bold",  
 vjust=-.25))+  
 theme(axis.title.y=element\_text(size=12,  
 face="bold",  
 vjust=.25))+  
 theme(axis.text.x=element\_text(size=10,  
 face="bold",  
 color="black"))+  
 theme(axis.text.y=element\_text(size=10,  
 face="bold",  
 color="black"))+  
 coord\_cartesian(ylim=c(min(temp$means)-2\*max(temp$sems),  
 max(temp$means)+2\*max(temp$sems)))+  
 theme(panel.border=element\_blank(),  
 axis.line=element\_line())+  
 theme(panel.grid.major.x=element\_blank())+  
 theme(panel.grid.major.y=element\_line(color="darkgrey"))+  
 theme(panel.grid.minor.y=element\_blank())  
f

 Another test I am able to perform on the data is a two-tailed paired t-test.

t.test(dat$wpm[dat$music=="metal"],  
 dat$wpm[dat$music=="none"],  
 var.equal=TRUE,  
 paired=TRUE)

##   
## Paired t-test  
##   
## data: dat$wpm[dat$music == "metal"] and dat$wpm[dat$music == "none"]  
## t = 12.153, df = 31, p-value = 2.51e-13  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 4.652093 6.528390  
## sample estimates:  
## mean of the differences   
## 5.590241

The results reveal that those in the music:metal condition typed more quickly than those in the music:none condition t(31) = 12.153, p < .05.