install.packages("tidyverse")

geom\_smooth(mapping = aes(x=displ,y=hwy,linetype=drv))+geom\_point(mapping = aes(x=displ,y=hwy,color=drv)) + xlab() + ylab() + ggtitle()

geom\_boxplot(aes(x=factor(x), y=y))

**## 95 % confidence interval for the data**

**## assuming normality and large n**

mean(y) + qnorm(.025)\*sqrt(var(y)/length(y))

mean(y) - qnorm(.025)\*sqrt(var(y)/length(y))

**## 95% confidence interval for the mean and a small unknown variance**

mean(y) + qt(.025,length(y)-1)\*sqrt(var(y)/length(y))

mean(y) - qt(.025,length(y)-1)\*sqrt(var(y)/length(y))

**bootstrap =**

b <- numeric(10000)

for(i in 1:10000) b[i] <- mean(sample(y,replace = T))

b <- sort(b); b[250]; b[9750]

**## test statistic**

z<-(mean(y)-30)/sqrt(var(y)/length(y))

**## p value**

p<- 1-pnorm(z)

**## 95% confidence interval for variance assuming normality**

(n-1)\*var(y)/qchisq(.025,n-1)

(n-1)\*var(y)/qchisq(.975,n-1)

pvalu <- 1-pchisq((n-1)\*var(y)/30,n-1)

**## paired test normal assumption**

t.test(v1,v2,paired = T,mu=0,alternative = "two.sided")

## **non parametric test for equal variances**

mood.test(men,women,alternative = "two.sided")

**##welches t test for equal means**

t.test(men,women,alternative = "two.sided",var.equal = F/T)

**## man whitney non parametric test for non normal data**

wilcox.test(men,women,alternative = "two.sided",exact = F)

**### randomization test**

xy <- c(x,y)

T<-numeric(10000)

for (i in 1:10000)T[i] <- sum(sample(xy,size = length(x),replace = F))

Tobs <- sum(x) ; table(T)

**## Tobs for the sample x was 0 , count the number of T's <= 0**

**##proportion test ##n = total**

prop.test(x=c(9,4),n=c(29,31),correct=F)

**##we are testing one sample whether p=.5**

binom.test(x=5067,n=10000,p=.5)

prop.test(x=5067,n=10000,p=.5,correct=F)

**## Fischers exact test: Look at all possible 2\*2 matrices ...**

fisher.test(matrix(c(17,139-17,31,141-31),ncol=2))

**## count data**

smoke.freq <- table(survey$Smoke)

smoke.prob <- c(.045,.795,.085,.075)

chisq.test(smoke.freq,p=smoke.prob)

**## make contingency table**

tbl <- table(survey$Smoke,survey$Exer)

**## test whether smoking and exercise are independent**

chisq.test(tbl)

**## chi square is imappropriate beacuase some cell counts r too low**

**## combine cells**

ctbl <- cbind(tbl[,"Freq"],tbl[,"None"] +tbl[,"Some"] )

chisq.test(ctbl)

**##If you dont have table data you can make a matrix**

color <- matrix(c(68,119,26,7,20,84,17,94,15,54,14,10,5,29,14,16),

nrow=4,byrow=T, dimnames=list(eye.color=c("Brown","Blue","Hazel","Green"),hair.color=c("Black","Brown","Red","Blonde")))

**##eye.color = colomn names**

chisq.test(color)

O <- c(229,211,93,35,7,0,0,1)

n<- c(0,1,2,3,4,5,6,7)

**## mean**

mn <- sum(O\*n)/sum(O)

**## expected**

E <- sum(O)\*mn^n\*exp(-mn)/factorial(n)

O <- c(O[1:4],sum(O[5:8]))

E <- c(E[1:4],sum(E[5:8]))

T <- sum(((O-E)^2)/E)

pv <- 1- pchisq(T,3)

**## If all cells are small and we cant combine cells to get a large enough cell count = Use fishers exact test**

Job <-matrix(c(1,2,1,0,3,3,6,1,10,10,14,9,6,7,12,11),4,4,dimnames = list(income = c("<15k","15-25k","25-40k",">40k"),satisfaction = c("veryD","LittleD","ModerateD","VeryS")))

fisher.test(Job)

Sm<-data$V2[which(data$V5==1)]