**25 Jan: Data: Cardata 2005, Foldername: PGAabhi2, filename: PGAabhi2**

**Basic command, T test, Anova, Chisquare Test, data manipulation**

cardata=read.csv(file.choose())

head(cardata)

tail(cardata)

str(cardata)

class(cardata) # gives the class of data

ncol(cardata)

nrow(cardata)

colnames(cardata) # gives column names

rownames(cardata)

summary(cardata)

summary(cardata$Price)

summary(cardata$Mileage)

summary(cardata$Make)

sd(cardata$Price)

sd(cardata$Mileage)

describe(cardata) #all info on central tendency, dispersion and shape need psych package

hist(cardata$Price)

hist(cardata$Mileage)

boxplot(cardata$Price,horizontal = T)

boxplot(cardata$Mileage, horizontal = T)

plot(density(cardata$Price))

plot(density(cardata$Mileage))

plot(density(log(cardata$Price))) # logrithimic transformation

str(cardata)

cardata1=cardata[-c(3:6)] # column deletion

cov(cardata1)

cor(cardata1)

cor.plot(cardata1)

pairs(cardata)

table(cardata$Make)

table(cardata$Type)

table(cardata$Make,cardata$Type)

table(cardata$Make,cardata$Sound)

table(cardata$Make,cardata$Leather)

table(cardata$Make,cardata$Sound,cardata$Leather)

round(prop.table(table(cardata$Make)),digits = 4)\*100

round(prop.table(table(cardata$Model)),digits = 4)\*100

pie(round(prop.table(table(cardata$Model)),digits = 4)\*100)

aggregate(cardata$Price~cardata$Sound,FUN = mean)

aggregate(cardata$Price~cardata$Sound,FUN = sd)

aggregate(cbind(Price=cardata$Price,Mile=cardata$Mileage)~cardata$Sound+cardata$Leather+cardata$Make,FUN = mean)

aggregate(cardata$Price~cardata$Sound,FUN = mean) # mean prices for cvars with and without sound system

t.test(cardata$Price~cardata$Sound)

#Test the Hypothesis that the cruise control Has Significant Impact on Mileage of Car:

str(cardata)

aggregate(cardata$Mileage~cardata$Cruise,FUN = mean)

t.test(cardata$Mileage~cardata$Cruise)

aggregate(cardata$Price~cardata$Leather,FUN = mean)

t.test(cardata$Price~cardata$Leather)

aggregate(cardata$Price~cardata$Make,FUN = mean) #Anova

summary(aov(cardata$Price~cardata$Make))

aggregate(cardata$Mileage~cardata$Type, FUN = mean)

summary(aov(cardata$Mileage~cardata$Type))

# chi-square test:

# Thus a relationship exist between the Make and Type variables in the data

table(cardata$Make,cardata$Type)

chisq.test(table(cardata$Make,cardata$Type)) # table is always an input command for chi sq test

# Thus a relationship exist between the Type and Cruise variables in the data

table(cardata$Type,cardata$Cruise)

chisq.test(table(cardata$Type,cardata$Cruise))

#chisquare test of proportion:

#Test weather cars with and without sound systems are in equal proportions

table(cardata$Sound)

chisq.test(table(cardata$Sound)) # by default this command check for equal probabilities

chisq.test(table(cardata$Sound), p=c(0.3,0.7)) # chsq test for given probabilities since P value=.196, which is >0.05 so accept null i.e. the cars with and without sound systems are not in equal proportions

#Data manipulation (creating new variables based on existing variables)

summary(cardata$Price)

cardata$PriceType=ifelse(cardata$Price>=35000,"High Price",ifelse(cardata$Price>=15000,"Medium Price","Low Price"))

summary(cardata$PriceType)

table(cardata$PriceType)

table(cardata$Price,cardata$PriceType)

summary(cardata$Mileage)

cardata$usetype=ifelse(cardata$Mileage>=30000,"Very old",ifelse(cardata$Mileage>=20000,"Old","Recent"))

table(cardata$usetype)

#chisquare test on this:

table(cardata$PriceType,cardata$usetype)

chisq.test(table(cardata$PriceType,cardata$usetype))

cor(cardata$Price,cardata$Mileage)

str(cardata)

cardata$Cylinder=as.factor(cardata$Cylinder) # as.factor converts numerical into character

str(cardata)

cardata$Liter=as.factor(cardata$Liter)

cardata$Doors=as.factor(cardata$Doors)

cardata$Cruise=as.factor(cardata$Cruise)

cardata$Sound=as.factor(cardata$Sound)

cardata$Leather=as.factor(cardata$Leather)

str(cardata)