**6304 Module 5 Live Lecture**

**R Script File**

**rm(list=ls())**

**library(rio)**

**# Hotel Restaurant Revenue**

**restaurant=import("6304 Module 5 Data Sets.xlsx",**

**sheet="Restaurant",skip=2)**

**colnames(restaurant)=tolower(make.names(colnames(restaurant)))**

**attach(restaurant)**

**names(restaurant)**

**plot(rooms.occupied,revenue,pch=19,main="Restaurant Revenue")**

**cor(rooms.occupied,revenue)**

**restaurant.out=lm(revenue~rooms.occupied,data=restaurant)**

**summary(restaurant.out)**

**abline(restaurant.out,lwd=3,col="red")**

**#Linearity**

**plot(restaurant$revenue,restaurant.out$fitted.values,**

**pch=19,main="Restaurant, Actuals v. Fitteds",**

**xlim=c(0,12000),ylim=c(0,12000))**

**abline(0,1,col="red",lwd=3)**

**#Normaility**

**qqnorm(restaurant.out$residuals,pch=19,**

**main="Restaurant, Normality of Residuals")**

**qqline(restaurant.out$residuals,lwd=3,col="red")**

**hist(restaurant.out$residuals,col="red",**

**main="Restaurant, Histogram of Residuals")**

**plot(density(restaurant.out$residuals),lwd=3,**

**main="Restaurant, Density Plot of Residuals")**

**moments::skewness(restaurant.out$residuals)**

**moments::kurtosis(restaurant.out$residuals)**

**#Equality of Variances**

**plot(restaurant$revenue,rstandard(restaurant.out),**

**pch=19,main="Restaurant Standardized Residuals",**

**sub="By Revenue")**

**abline(0,0,lwd=3,col="red")**

**plot(rstandard(restaurant.out),**

**main="Restaurant Standardized Residuals",**

**pch=19,sub="By Order")**

**abline(0,0,col="red",lwd=3)**

**#Some stuff to make it pretty.**

**plot(restaurant$rooms.occupied,restaurant$revenue,**

**pch=19,**

**main=paste("Restaurant Revenue r=",**

**round(cor(restaurant$rooms.occupied,**

**restaurant$revenue),3)))**

**abline(restaurant.out,lwd=3,col="red")**

**# Look at the pattern.**

**par(mfrow=c(3,1))**

**plot(restaurant$rooms.occupied,restaurant$revenue,**

**pch=19,main="Original Data")**

**abline(restaurant.out,lwd=3,col="red")**

**plot(restaurant$revenue,restaurant.out$fitted.values,**

**pch=19,main="Actuals v. Fitteds")**

**abline(0,1,lwd=3,col="red")**

**plot(restaurant$revenue,rstandard(restaurant.out),**

**main="Stdized Residuals",**

**pch=19)**

**abline(0,0,col="red",lwd=3)**

**par(mfrow=c(1,1))**

**leverage=hat(model.matrix(restaurant.out))**

**plot(leverage,pch=19,ylim=c(0,.5))**

**abline(3\*mean(leverage),0,col="red",lwd=3)**

**restaurant[leverage>3\*mean(leverage),]**

**restaurant[which.max(leverage),]**

**# New Data Set -- Warehouse Costs**

**rm(list=ls())**

**warehouse=import("6304 Module 5 Data Sets.xlsx",**

**sheet="Warehouse Cost",skip=2)**

**colnames(warehouse)=tolower(make.names(colnames(warehouse)))**

**attach(warehouse)**

**names(warehouse)**

**warehouse.out=lm(cost.000~sales.000+orders,data=warehouse)**

**summary(warehouse.out)**

**#Linearity**

**plot(warehouse$cost.000,warehouse.out$fitted.values,**

**pch=19,main="Warehouse Actuals v. Fitted")**

**abline(0,1,col="red",lwd=3)**

**cor(warehouse$cost.000,warehouse.out$fitted.values)**

**#Normality**

**qqnorm(warehouse.out$residuals,pch=19,**

**main="Warehouse Normality Plot")**

**qqline(warehouse.out$residuals,lwd=3,col="red")**

**hist(warehouse.out$residuals,col="red",**

**main="Warehouse Residuals Histogram")**

**plot(density(warehouse.out$residuals),lwd=3,**

**main="Warehouse Residuals Density Plot")**

**moments::skewness(warehouse.out$residuals)**

**moments::kurtosis((warehouse.out$residuals))**

**# Overlaying the Normal Curve & the Histogram**

**hist(warehouse.out$residuals,col="red",**

**main="Warehouse Residuals Histogram",**

**freq=FALSE)**

**curve(dnorm(x,mean(warehouse.out$residuals),**

**sd(warehouse.out$residuals)),**

**from=min(warehouse.out$residuals),**

**to=max(warehouse.out$residuals),lwd=3,**

**add=TRUE)**

**#Equality of Variances**

**plot(warehouse$cost.000,rstandard(warehouse.out),**

**pch=19,main="Warehouse Residual Plot")**

**abline(0,0,col="red",lwd=3)**

**#Identifying high leverage points.**

**leverage=hat(model.matrix(warehouse.out))**

**plot(leverage,pch=19,ylim=c(0,.5))**

**abline(3\*mean(leverage),0,col="red",lwd=3)**

**#A Prediction**

**maryann=data.frame(sales.000=300,orders=3000)**

**predict(warehouse.out,maryann,interval="predict")**

**predict(warehouse.out,maryann,interval="confidence")**

**predict(warehouse.out,maryann,interval="none")**

**predict(warehouse.out,maryann)**

**#Making a Mistake**

**maryann=data.frame(sales.00=300,orders=3000)**

**predict(warehouse.out,maryann,interval="predict")**

**# MPG Data**

**rm(list=ls())**

**cars=import("6304 Module 5 Data Sets.xlsx",sheet="MPG")**

**colnames(cars)=tolower(make.names(colnames(cars)))**

**attach(cars)**

**plot(horsepower,mpg,pch=19,main="MPG and Horsepower")**

**plot(weight,mpg,pch=19,main="MPG and Weight")**

**plot(cars,pch=19)**

**#A simple regression first**

**cars.out=lm(mpg~horsepower,data=cars)**

**summary(cars.out)**

**plot(horsepower,mpg,pch=19,main="MPG and Horsepower")**

**abline(cars.out,lwd=3,col="red")**

**plot(cars$mpg,rstandard(cars.out),pch=19,**

**main="Residual Plot")**

**abline(0,0,col="red",lwd=3)**

**#A data transform.**

**#Squaring the horsepower variable.**

**#The hard way to do it.**

**cars$horsepower2=cars$horsepower^2**

**#And conducting the regression.**

**cars2.out=lm(mpg~horsepower+horsepower2,data=cars)**

**summary(cars.out)**

**summary(cars2.out)**

**#How's the fit?**

**par(mfrow=c(1,2))**

**plot(cars$mpg,cars.out$fitted.values,pch=19,**

**main="Main Effects Model")**

**abline(0,1,col="red",lwd=3)**

**plot(cars$mpg,cars2.out$fitted.values,pch=19,**

**main="Squared Term Model")**

**abline(0,1,col="red",lwd=3)**

**par(mfrow=c(1,1))**

**#The easy way to do it.**

**#First let's clean up the data frame.**

**cars=cars[,-4]**

**cars2=lm(mpg~horsepower+I(horsepower^2))**

**summary(cars2.out)**

**#So let's throw in everything.**

**cars3.out=lm(mpg~horsepower+weight+I(horsepower^2)+**

**I(weight^2),data=cars)**

**summary(cars3.out)**

**#No identifiable nonlinear relationship with weight.**

**cars3.out=lm(mpg~horsepower+weight+I(horsepower^2),**

**data=cars)**

**summary(cars3.out)**

**#What about an interaction?**

**cars4.out=lm(mpg~horsepower+weight+I(horsepower^2)+**

**horsepower:weight,data=cars)**

**summary(cars4.out)**

**#Cars3.out is the best model fit.**

**#Cars3 Linearity**

**plot(cars$mpg,cars3.out$fitted.values,pch=19,**

**main="Cars3 Actual v. Forecast")**

**abline(0,1,lwd=3,col="red")**

**#Cars3 Normality**

**qqnorm(cars3.out$residuals,pch=19,**

**main="Cars3 QQ Plot")**

**qqline(cars3.out$residuals,lwd=3,col="red")**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Histogram")**

**plot(density(cars3.out$residuals),lwd=3,**

**main="Cars3 Residuals Density Plot")**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Density Overlay",freq = FALSE)**

**points(density(cars3.out$residuals),lwd=3)**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Better Density Overlay",freq = FALSE)**

**points(density(cars3.out$residuals),type="l",lwd=3,**

**main="Cars3 Residuals Density Plot")**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Normal Curve Overlay",freq = FALSE)**

**curve(dnorm(x,mean(cars3.out$residuals),**

**sd(cars3.out$residuals)),**

**from=min(cars3.out$residuals),**

**to=max(cars3.out$residuals),lwd=3,**

**add=TRUE)**

**#Lets make this prettier.**

**par(mfrow=c(2,2))**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Histogram")**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Density Overlay",freq = FALSE)**

**points(density(cars3.out$residuals),lwd=3)**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Better Density Overlay",freq = FALSE)**

**points(density(cars3.out$residuals),type="l",lwd=3,**

**main="Cars3 Residuals Density Plot")**

**hist(cars3.out$residuals,col="red",**

**main="Cars3 Residuals Normal Curve Overlay",freq = FALSE)**

**curve(dnorm(x,mean(cars3.out$residuals),**

**sd(cars3.out$residuals)),**

**from=min(cars3.out$residuals),**

**to=max(cars3.out$residuals),lwd=3,**

**add=TRUE)**

**par(mfrow=c(1,1))**

**#Cars3 Equality of Variances**

**plot(cars$mpg,rstandard(cars3.out),pch=19,**

**main="Cars3 Stdized Residuals v. Actuals")**

**abline(0,0,col="red",lwd=3)**

**plot(cars3.out$fitted.values,rstandard(cars3.out),pch=19,**

**main="Cars3 Stdized Residuals v. Fitted Values")**

**abline(0,0,col="red",lwd=3)**

**plot(rstandard(cars3.out),pch=19,**

**main="Cars3 Stdized Residuals In Order")**

**abline(0,0,col="red",lwd=3)**

**#Child Abuse with Binary Variables**

**rm(list=ls())**

**abuse1=import("6304 Module 5 Data Sets.xlsx",**

**sheet="Child Abuse with Binaries")**

**colnames(abuse1)=tolower(make.names(colnames(abuse1)))**

**attach(abuse1)**

**no.binary.out=lm(reported.victims~pop.under.18,data=abuse1)**

**summary(no.binary.out)**

**with.binary.out=lm(reported.victims~pop.under.18+se.state,data=abuse1)**

**summary(with.binary.out)**

**par(mfrow=c(1,2))**

**plot(abuse1$reported.victims,no.binary.out$fitted.values,**

**main = "No Binary Variable",pch=19)**

**abline(0,1,col="red",lwd=3)**

**plot(abuse1$reported.victims,with.binary.out$fitted.values,**

**main="With Binary Variable",pch=19)**

**abline(0,1,col="red",lwd=3)**

**par(mfrow=c(1,1))**

**#A better way to model a categorical variable.**

**abuse2=import("6304 Module 5 Data Sets.xlsx",**

**sheet="Child Abuse with Binaries 2")**

**colnames(abuse2)=tolower(make.names(colnames(abuse2)))**

**attach(abuse2)**

**better.binary.out=lm(reported.victims~pop.under.18+se.state,data=abuse2)**

**summary(better.binary.out)**

**#House Appraisals**

**rm(list=ls())**

**house=import("6304 Module 5 Data Sets.xlsx",**

**sheet="House Appraisals")**

**colnames(house)=tolower(make.names(colnames(house)))**

**str(house)**

**house$garage=as.factor(house$garage)**

**house$baths=as.factor(house$baths)**

**attach(house)**

**house.out=lm(appraised.value~land.acres+house.size.sqft+**

**age+rooms+baths+garage,data=house)**

**#OR**

**house.out=lm(appraised.value~.-address,data=house)**

**summary(house.out)**

**#Linearity**

**plot(house$appraised.value,house.out$fitted.values,**

**pch=19,main="House Data Actuals v. Fitted")**

**abline(0,1,lwd=3,col="red")**

**#Normality**

**par(mfrow=c(1,3))**

**qqnorm(house.out$residuals,pch=19,**

**main="House Data, Residuals QQ Plot")**

**qqline(house.out$residuals,lwd=3,col="red")**

**hist(house.out$residuals,col="red",**

**main="House Data, Residuals Histogram")**

**plot(density(house.out$residuals),lwd=3,**

**main="House Data, Residuals Density Plot")**

**moments::skewness(house.out$residuals)**

**moments::kurtosis(house.out$residuals)**

**#par(mfrow=c(1,1))**

**par(mfrow=c(3,1))**

**qqnorm(house.out$residuals,pch=19,**

**main="House Data, Residuals QQ Plot")**

**qqline(house.out$residuals,lwd=3,col="red")**

**hist(house.out$residuals,col="red",**

**main="House Data, Residuals Histogram")**

**plot(density(house.out$residuals),lwd=3,**

**main="House Data, Residuals Density Plot")**

**moments::skewness(house.out$residuals)**

**moments::kurtosis(house.out$residuals)**

**par(mfrow=c(1,1))**

**#Equality of Variances**

**plot(house.out$fitted.values,rstandard(house.out),pch=19)**

**abline(0,0,col="red",lwd=3)**