**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")**

**# Combined Plot**

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

**# Linearity**

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

**pch=19,xlim=c(0,10000),ylim=c(0,10000),**

**main="Actuals v. Fitteds, Restaurant Revenue")**

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

**# Normality**

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

**main="Normality Plot, Restaurant Revenue")**

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

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

**main="Residuals, Restaurant Revenue",**

**probability = TRUE)**

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

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

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

**to=max(restaurant.out$residuals)**

**,lwd=3,col="blue",add=TRUE)**

**# Equality of Variances**

**plot(restaurant.out$fitted.values,rstandard(restaurant.out),**

**pch=19,main="Equality of Variances, Restaurant Revenue")**

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

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

**#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")**

**# 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)**

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

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

**# Combined Plot**

**# Home Made Four-In-One Plot**

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

**#Linearity**

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

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

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

**#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",**

**freq=FALSE,ylim=c(0,.1))**

**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)**

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

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

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

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

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

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

**#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)**

**# 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 does the fit change?**

**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))**

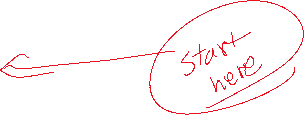
**# The easy way to do it.**

**# 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.**

**cars3a.out=lm(mpg~horsepower+weight+I(horsepower^2),**

**data=cars)**

**summary(cars3a.out)**

**# What about an interaction?**

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

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

**summary(cars4.out)**

**#Cars3a.out is the best model fit.**

**#Cars3a Combined Plot**

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

**plot(cars$mpg,cars3a.out$fitted.values,pch=19,**

**main="Cars3a Actual v. Forecast")**

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

**#Cars3a Normality**

**qqnorm(cars3a.out$residuals,pch=19,**

**main="Cars3a QQ Plot")**

**qqline(cars3a.out$residuals,lwd=3,col="red")**

**hist(cars3a.out$residuals,col="red",**

**main="Cars3a Residuals Normal Curve Overlay",freq = FALSE)**

**curve(dnorm(x,mean(cars3a.out$residuals),**

**sd(cars3a.out$residuals)),**

**from=min(cars3a.out$residuals),**

**to=max(cars3a.out$residuals),lwd=3,**

**add=TRUE)**

**plot(cars3a.out$fitted.values,rstandard(cars3a.out),**

**pch=19,main="Cars3a Equality of Variances")**

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

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

**# Firearms with Binary Variables**

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

**firearms1=import("6304 Module 5 Data Sets.xlsx",**

**sheet="Firearms with Binaries")**

**colnames(firearms1)=tolower(make.names(colnames(firearms1)))**

**firearms1.out=lm(firearm.deaths~population+se.state,data=firearms1)**

**summary(firearms1.out)**

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

**firearms2=import("6304 Module 5 Data Sets.xlsx",**

**sheet="Firearms with Binaries 2")**

**colnames(firearms2)=tolower(make.names(colnames(firearms2)))**

**str(firearms2)**

**firearms2$se.state=as.factor(firearms2$se.state)**

**str(firearms2)**

**firearms2.out=lm(firearm.deaths~population+se.state,data=firearms2)**

**summary(firearms2.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)**

**str(house)**

**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)**

**# Why Categorical Variables?**

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

**numeric.model=lm(appraised.value~garage,data=house)**

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

**categorical.model=lm(appraised.value~garage,data=house)**

**summary(numeric.model)**

**summary(categorical.model)**