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

#scatter matrix plot

library(tidyverse)

library(GGally)

nyc <- read.csv("C:/Teaching@cofc/Math 550/Chapter 5/nyc.csv")

library(car)

scatterplotMatrix(~Price+Food+Decor+Service+East, data=nyc)

pairs(~Price+Food+Decor+Service+East, data=nyc, lower.panel=panel.smooth)

lm<-lm(Price~Food+Decor+Service+East,data=nyc)

sres<-rstandard(lm)

plot(sres~nyc$Food)

plot(sres~nyc$Decor)

plot(sres~nyc$Service)

plot(sres~nyc$East)

plot(nyc$Price~fitted(lm))

abline(a=0,b=1)

#valid model

#generated data 1 to support x1=1+0.5x2

#y=(x1+2x2)^2+e

# Fit model y~x1,x2 e with constant variability

quadraticwithconstanterror <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/quadraticwithconstanterror.txt")

attach(quadraticwithconstanterror)

pairs(~y+x1+x2, lower.panel=panel.smooth)

lmsg1<-lm(y~x1+x2)

plot(rstandard(lmsg1)~x1)

plot(rstandard(lmsg1)~x2)

plot(rstandard(lmsg1)~fitted(lmsg1))

plot(y~fitted(lmsg1))

#generated data 2 to support x1=1+0.5x2

#y=x1+x2+e

# Fit model y~x1,x2 e increasing variability

nonlinearxincreasingerror <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/nonlinearxincreasingerror.txt")

attach(nonlinearxincreasingerror)

pairs(~y+x1+x2, lower.panel=panel.smooth)

lmsg2<-lm(y~x1+x2)

plot(rstandard(lmsg2)~x1)

plot(rstandard(lmsg2)~x2)

plot(rstandard(lmsg2)~fitted(lmsg2))

plot(y~fitted(lmsg2))

#if the condition(s) are not satisfied then the standardized residual plots don’t provide direct information

caution <- read.csv("C:/Teaching@cofc/Math 550/Chapter 6/caution.txt", sep="")

# in the generated data, E(Y|X)=|x1|/(2+(1.5+x2)^2)=g1(x1)/g2(x2)

#Note two functions not one

#observations of X1 and X2 were sample from an elliptical distribution

# note linearity among the predictors is a stronger conditions than

#elliptical distribution

#errors from normal (0,1)

pairs(~y+x1+x2, data=caution)

cor(caution$x1,caution$x2)

lmg<-lm(caution$y~caution$x1+caution$x2)

sreg<-rstandard(lmg)

plot(sreg~caution$x1)

plot(sreg~caution$x2)

plot(sreg~fitted(lmg))

plot(caution$y ~fitted(lmg))

# y vs fitted y shows y is not a single function of fitted y.

nonlinearx <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/nonlinearx.txt")

#x1 is equally spaced from -3 and 3 and errors are normally distributed with #std=0.1

#generated from Y=x1+3x2^2+e

#E(X2|X1)=sin(X1)

pairs(~y+x1+x2,data=nonlinearx)

lmg2<-lm(y~x1+x2,data=nonlinearx)

rseg2<-rstandard(lmg2)

plot(rseg2~nonlinearx$x1)

plot(rseg2~nonlinearx$x2)

plot(rseg2~fitted(lmg2))

plot(nonlinearx$y~fitted(lmg2))

# missing a periodic function of x1 based on standardized residual plot is not true

library(car)

nyc <- read.csv("C:/Teaching@cofc/Math 550/Chapter 5/nyc.csv")

scatterplotMatrix(~Price+Food+Decor+Service+East, data=nyc)

attach(nyc)

plot(Price~Food)

plot(Price~Decor)

plot(Price~Service)

plot(Price~East)

# these plots look at the effect of a given predictor on *Y* , Price, ignoring the effects of the other predictors on Price

# added variable plots

avPlots(lm(Price~Food+Decor+Service+East, data=nyc))

#transformation-Inverse response plots

defects <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/defects.txt")

scatterplotMatrix (~Defective+Temperature+Density+Rate,data=defects)

lmm<-lm(Defective~ Temperature+Density+Rate, data=defects)

restd<-rstandard(lmm)

plot(restd~Temperature,data=defects)

plot(restd~Density,data=defects)

plot(restd~Rate,data=defects)

plot(defects$Defective~fitted(lmm))

#explanation of inverse response plot

attach(defects)

newdata = data.frame(Temperature = Temperature, Density = Density, Rate= Rate )

predicty <- predict(lmm,newdata)

yno<-Defective^(-1)

ynh<- Defective ^(-0.5)

ynt <- Defective ^(-1/3)

yz<-log(Defective)

yt <- Defective ^(1/3)

yh<- Defective ^(0.5)

yo<- Defective

lamno<-lm(predicty ~yno)

rssno<- sum(resid(lamno)^2)

lamnh<-lm(predicty ~ynh)

rssnh<- sum(resid(lamnh)^2)

lamnt<-lm(predicty ~ynt)

rssnt<- sum(resid(lamnt)^2)

lamz<-lm(predicty~yz)

rssz<- sum(resid(lamz)^2)

lamo<-lm(predicty ~yo)

rsso<- sum(resid(lamo)^2)

lamh<-lm(predicty ~yh)

rssh<- sum(resid(lamh)^2)

lamt<-lm(predicty~yt)

rsst<- sum(resid(lamt)^2)

library(alr3)

lam<-invResPlot(lmm, lambda=c(-1,-1/2,-1/3,-1/4,0,1/4,1/3,1/2,1))

lam$lambda

plot(lam$RSS~lam$lambda)

transd<-sqrt(defects$Defective)

tlmm<-lm(transd~ Temperature+Density+Rate, data=defects)

trestd<-rstandard(tlmm)

plot(trestd~Temperature,data=defects)

plot(trestd~Density,data=defects)

plot(trestd~Rate,data=defects)

plot(transd~fitted(tlmm))

#Box-Cox method

install.packages("psych")

library("psych")

lambda<-rep(0,100)

RSSnl<-rep(0,100)

for(i in 1:100)

{ lambda[i]<- -3+(i-1)\*6/99

ynl<-geometric.mean(Defective)^(1-lambda[i])\*(Defective^lambda[i]-1)/lambda[i]

lmnl<-lm(ynl~ Temperature+Density+Rate)

RSSnl[i]<-sum(resid(lmnl)^2)

}

best.lambda<-lambda [which(RSSnl==min(RSSnl))]

#transformation-Box-cox method

library(faraway)

library(MASS)

bc<-boxcox(lmm, lambda=seq(-3,3,1/10))

bestlambda<-bc$x[which(bc$y==max(bc$y))]

ndefective<-sqrt(defects$Defective)

lmn<-lm(ndefective~Temperature+Density+Rate, data=defects)

restdn<-rstandard(lmn)

plot(restdn~Temperature,data=defects)

plot(restdn~Density,data=defects)

plot(restdn~Rate,data=defects)

plot(restdn~fitted(lmn))

plot(lmn)

summary(lmn)

avPlots(lm(ndefective~Temperature+Density+Rate, data=defects))

magazines <- read.csv("C:/Teaching@cofc/Math 550/Chapter 6/magazines.csv")

scatterplotMatrix (~AdRevenue + AdPages + SubRevenue + NewsRevenue,data= magazines)

pc<-powerTransform(cbind(magazines$AdPages, magazines$SubRevenue, magazines$NewsRevenue)~1)

summary(pc)

library(alr3)

logap<-log(magazines$AdPages ,base=exp(1))

logsr<-log(magazines$SubRevenue,base=exp(1))

lognr<-log(magazines$NewsRevenue,base=exp(1))

scatterplotMatrix (~logap + logsr + lognr)

lmag<-lm(AdRevenue~ logap+ logsr+lognr, data=magazines)

plot(magazines$ AdRevenue~fitted(lmag))

lam<-invResPlot(lmag, lambda=c(-1,-1/2,-1/3,-1/4,0,1/4,1/3,1/2,1))

lam$lambda

plot(lam$RSS~lam$lambda)

logar<- log(magazines$AdRevenue,base=exp(1))

scatterplotMatrix (~logar+logap+logsr+lognr)

lmagapproach1<-lm(logar~ logap+ logsr+lognr)

summary(lmagapproach1)

restdn<-rstandard(lmagapproach1)

plot(restdn~logap)

plot(restdn~logsr)

plot(restdn~lognr)

plot(logar~fitted(lmagapproach1))

# add line y=a+bx

abline(a=0,b=1)

plot(lmagapproach1)

avPlots(lmagapproach1)

pca<-powerTransform(cbind(magazines$AdRevenue, magazines$AdPages, magazines$SubRevenue, magazines$NewsRevenue)~1)

summary(pca)

circulation <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/circulation.txt")

#Skip this example

pairs(~Sunday+Weekday+Tabloid.with.a.Serious.Competitor, data=circulation)

logsun<-log(circulation$Sunday ,base=exp(1))

logweek<-log(circulation$Weekday,base=exp(1))

pairs(~logsun+logweek+ circulation$Tabloid.with.a.Serious.Competitor)

logmod<-lm(logsun~logweek+ Tabloid.with.a.Serious.Competitor, data=circulation)

summary(logmod)

resmod<-rstandard(logmod)

plot(resmod~logweek)

plot(resmod~ circulation$Tabloid.with.a.Serious.Competitor)

plot(resmod~fitted(logmod))

plot(logsun~fitted(logmod))

abline(0,1)

plot(logmod)

logweekn<-log(210000,base=exp(1))

with<-predict(logmod,list(logweek=logweekn, Tabloid.with.a.Serious.Competitor=1), interval = "prediction", level = 0.95)

without<-predict(logmod,list(logweek=logweekn, Tabloid.with.a.Serious.Competitor=0), interval = "prediction", level = 0.95)

exp(with)

exp(without)

avPlots(logmod)

#for multicollinearity

bridge <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/bridge.txt")

scatterplotMatrix (~Time+DArea+CCost+Dwgs+Length+Spans,data=bridge)

attach(bridge)

library(car)

pca<-powerTransform(cbind(bridge$Time, bridge$DArea, bridge$CCost, bridge$Dwgs, bridge$Length, bridge$Spans)~1)

summary(pca)

logt<-log(bridge$Time,base=exp(1))

logd<-log(bridge$DArea,base=exp(1))

logc<-log(bridge$CCost,base=exp(1))

logl<-log(bridge$Length,base=exp(1))

logs<-log(bridge$Spans,base=exp(1))

logdw<- log(bridge$Dwgs,base=exp(1))

scatterplotMatrix (~logt+logd+logc+logl+logs)

log<-lm(logt~logd+logc+logdw+logl+logs)

reslog<-rstandard(log)

plot(reslog~logt)

plot(reslog~logd)

plot(reslog~logc)

plot(reslog~logl)

plot(reslog~logs)

plot(reslog~logdw)

plot(logt~fitted(log))

abline(0,1)

plot(log)

summary(log)

cor(logd,logc)

cor(logd,logl)

cor(logd,logs)

cor(logd,logdw)

cor(logc,logl)

cor(logc,logs)

cor(logc,logdw)

cor(logs,logdw)

avPlots(log)

lmd<-lm(logd~ logc+logdw+logl+logs)

out<-summary(lmd)

1/(1-out$r.squared)

vif(log)

storks <- read.delim("C:/Teaching@cofc/Math 550/Chapter 6/storks.txt")

lm<-lm(Babies~Storks,data=storks)

summary(lm)

attach(storks)

plot(Babies~Storks)

abline(lm)

summary(lm)

lmw<-lm(Babies~Women,data=storks)

plot(Babies~Women)

abline(lmw)

summary(lmw)

lms<-lm(Women~Storks,data=storks)

plot(Women~Storks)

abline(lms)

summary(lms)

lmm<-lm(Babies~Storks+Women,data=storks)

summary(lmm)

summary(lmm)

cars04 <- read.csv("C:/Teaching@cofc/Math 550/Chapter 6/cars04.csv")

pairs(~SuggestedRetailPrice+EngineSize+Cylinders+Horsepower+HighwayMPG+Weight+WheelBase+Hybrid, data=cars04)

lmwo<-lm(SuggestedRetailPrice~EngineSize+Cylinders+Horsepower+HighwayMPG+Weight+WheelBase+Hybrid,data=cars04)

sres<-rstandard(lmwo)

plot(sres~ cars04$EngineSize)

plot(sres~ cars04$ Cylinders)

plot(sres~ cars04$ Horsepower)

plot(sres~ cars04$ HighwayMPG)

plot(sres~ cars04$ Weight)

plot(sres~ cars04$ WheelBase)

plot(sres~ cars04$ Hybrid)

plot(sres~fitted(lmwo))

outhat<-hatvalues(lmwo)

Leverage<-I(outhat>2\*(7+1)/234)

Outlier<-I(abs(sres)>2)

Badlev<-Leverage\*Outlier

plot(lmwo)

pc<-powerTransform(cbind(cars04$EngineSize, cars04$ Cylinders, cars04$ Horsepower, cars04$ HighwayMPG, cars04$ Weight, cars04$ WheelBase)~1)

summary(pc)

attach(cars04)

**tEngineSize <- (EngineSize)^.25**

**tCylinders <- log(Cylinders,** base=exp(1)**)**

**tHorsepower <- log(Horsepower,** base=exp(1)**)**

**tHighwayMPG <- 1/HighwayMPG**

**tWheelBase <- log(WheelBase,** base=exp(1)**)**

lmwt<-lm(SuggestedRetailPrice~tEngineSize+tCylinders+tHorsepower+tHighwayMPG+Weight+tWheelBase+Hybrid,data=cars04)

lam<-invResPlot(lmwt, lambda=c(-1,-1/2,-1/3,-1/4,0,1/4,1/3,1/2,1))

lam$lambda

plot(lam$RSS~lam$lambda)

**tSuggestedRetailPrice <- log(SuggestedRetailPrice,** base=exp(1)**)**

**#estimate percentage effects**

lmt<-lm(tSuggestedRetailPrice~tEngineSize+tCylinders+tHorsepower+tHighwayMPG+Weight+tWheelBase+Hybrid,data=cars04)

rest<-rstandard(lmt)

plot(rest~fitted(lmt))

plot(lmt)

vif(lmt)

lmtr<-lm(tSuggestedRetailPrice~tEngineSize+tCylinders+tHorsepower +Weight +Hybrid,data=cars04)

plot(lmtr)

library(car)

nyc <- read.csv("C:/Teaching@cofc/Math 550/Chapter 5/nyc.csv")

scatterplotMatrix(~Price+Food+Decor+Service+East, data=nyc)

cars04 <- read.csv("C:/Teaching@cofc/Math 550/Chapter 6/cars04.csv")

scatterplotMatrix(~SuggestedRetailPrice+EngineSize+Cylinders+Horsepower+HighwayMPG+Weight+WheelBase+Hybrid,data=cars04)

scatterplotMatrix(~SuggestedRetailPrice+EngineSize+Cylinders+Horsepower+HighwayMPG+Weight+WheelBase+Hybrid, smooth=FALSE, data=cars04)