# code 7 Regression

# text figure

plot(c(0,10),c(0,100),xlab="",ylab="",type="n")

lines(c(0,10),c(80,10),lwd=2)

# intercept = 80

lines(c(0,0),c(0,80),col="green")

lines(c(0,-10),c(80,80),col="red")

# slope = -7

lines(c(2,8),c(24,24),col="brown")

lines(c(2,2),c(66,24),col="blue")

# tannin example

reg.data <- read.csv("c:\\temp\\tannin.csv")

attach(reg.data)

names(reg.data)

plot(tannin,growth,pch=21,bg="blue")

lm(growth~tannin)

abline(lm(growth~tannin),col="green")

fitted <- predict(lm(growth~tannin))

fitted

lines(c(0,0),c(12,11.7555555))

# residuals

for (i in 1:9)

lines (c(tannin[i],tannin[i]),c(growth[i],fitted[i]),col="red")

# estimating the maximum likelihood slope

b <- seq(-1.43,-1,0.002)

sse <- numeric(length(b))

for (i in 1:length(b)) {

a <- mean(growth)-b[i]\*mean(tannin)

residual <- growth - a - b[i]\*tannin

sse[i] <- sum(residual^2)

}

plot(b,sse,type="l",ylim=c(19,24))

arrows(-1.216,20.07225,-1.216,19,col="red")

abline(h=20.07225,col="green",lty=2)

lines(b,sse)

b[which(sse==min(sse))]

# corrected sums of squares

SSX <- sum(tannin^2)-sum(tannin)^2/length(tannin)

SSY <- sum(growth^2)-sum(growth)^2/length(growth)

SSXY <- sum(tannin\*growth)-sum(tannin)\*sum(growth)/length(tannin)

# box 7.5 figure

plot(c(0,10),c(0,10),xlab="",ylab="",type="n")

abline(h=5,lty=2)

lines(c(0,10),c(8,2))

text(2,6.2,expression(hat(y) - bar(y)))

text(2,8.45,expression(y - hat(y)))

arrows(7,5,7,9.5,code=3,length=0.1)

arrows(1,5,1,7.4,code=3,length=0.1)

arrows(1,9.5,1,7.4,code=3,length=0.1)

points(1,9.5,pch=16)

text(8,7.4,expression(y - bar(y)))

text(0.2,5,expression(bar(y)))

text(.2,7.4,expression(hat(y)))

text(.2,9.5,"y")

# regreesion model in R

model <- lm(growth~tannin)

summary(model)

summary.aov(model)

par(mfrow=c(2,2))

plot(model)

# a non-linear relationship

par(mfrow=c(1,1))

data <- read.csv("c:\\temp\\decay.csv")

attach(data)

names(data)

plot(time,amount,pch=21,col="blue",bg="green")

abline(lm(amount~time),col="red")

summary(lm(amount~time))

plot(time,log(amount),pch=21,col="blue",bg="red")

abline(lm(log(amount)~time),col="blue")

model <- lm(log(amount)~time)

summary(model)

par(mfrow=c(1,1))

plot(time,amount,pch=21,col="blue",bg="green")

xv <- seq(0,30,0.25)

yv <- 94.38536 \* exp(-0.068528 \* xv)

lines(xv,yv,col="red")

# shapes of quadratic relationships

par(mfrow=c(2,2))

curve(4+2\*x-0.1\*x^2,0,10,col="red",ylab="y")

curve(4+2\*x-0.2\*x^2,0,10,col="red",ylab="y")

curve(12-4\*x+0.3\*x^2,0,10,col="red",ylab="y")

curve(4+0.5\*x+0.1\*x^2,0,10,col="red",ylab="y")

model2 <- lm(amount~time)

model3 <- lm(amount~time+I(time^2))

summary(model3)

AIC(model2,model3)

anova(model2,model3)

# non-linear regression using nls

deer <- read.csv("c:\\temp\\jaws.csv")

attach(deer)

names(deer)

par(mfrow=c(1,1))

plot(age,bone,pch=21,bg="lightgrey")

model <- nls(bone~a-b\*exp(-c\*age),start=list(a=120,b=110,c=0.064))

summary(model)

model2 <- nls(bone~a\*(1-exp(-c\*age)),start=list(a=120,c=0.064))

anova(model,model2)

av <- seq(0,50,0.1)

bv <- predict(model2,list(age=av))

lines(av,bv,col="blue")

summary(model2)

null.model <- lm(bone ~ 1)

summary.aov(null.model)

# generalized additive models GAM

library(mgcv)

hump <- read.csv("c:\\temp\\hump.csv")

attach(hump)

names(hump)

model <- gam(y~s(x))

plot(model,col="blue")

points(x,y-mean(y),pch=21,bg="red")

summary(model)