###########Q1

#(a)

logret<-read.table("http://www.ams.sunysb.edu/~pfkuan/Teaching/AMS597/Data/d\_logret\_6stocks.txt",header=T)

attach(logret)

fit1<-lm(Pfizer~Exxon+Citigroup)

fit1$coefficients

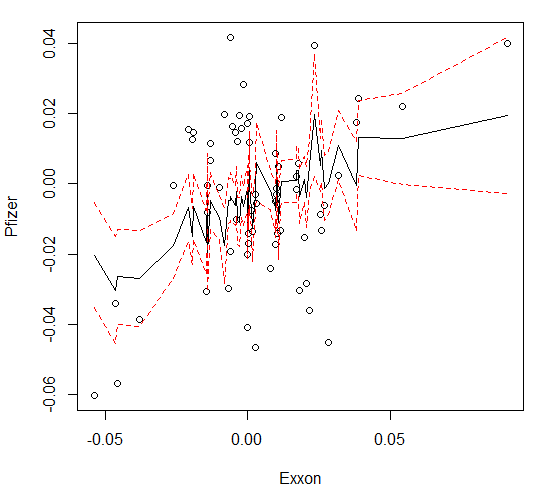
#The estimated coefficients for Exxon is 0.287636 and Citigroup is 0.185977 and estimate for intercept is -0.005257.

#(b)

pc<-predict(fit1, int="c")

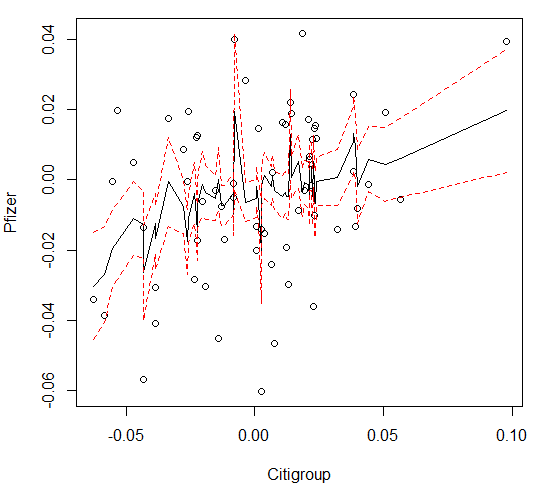
plot( Exxon,Pfizer)

matlines(sort(Exxon), pc[order(Exxon),], lty=c(1,2,2), col=c("black", "red", "red"))



plot( Citigroup,Pfizer)

matlines(sort(Citigroup), pc[order(Citigroup),], lty=c(1,2,2), col=c("black", "red", "red"))



#(c)

anova(fit1)

#Analysis of Variance Table

#Response: Pfizer

# Df Sum Sq Mean Sq F value Pr(>F)

#Exxon 1 0.0041609 0.0041609 9.2462 0.003475 \*\*

#Citigroup 1 0.0019516 0.0019516 4.3369 0.041496 \*

#Residuals 61 0.0274506 0.0004500

# the p-values are less than 0.05, so reject the null hypothesis (coefficient β = 0). Conclude that the regression effects are significant

#(d)

fit2<-lm(Pfizer~-1+Exxon+Citigroup)

fit2$coefficients

#the estimated coefficients for Exxon is 0.25097 and Citigroup is 0.18812.

#(e)

cor(Pfizer,Exxon) # 0.3520965

cor.test(Pfizer,Exxon)

#p-value is very small 0.004328<0.05, so we reject null, so the correlation is not zero

################# Q2

#(a)

logret<-read.table("http://www.ams.sunysb.edu/~pfkuan/Teaching/AMS597/Data/d\_logret\_6stocks.txt",header=T)

attach(logret)

group1<-c(Citigroup, AmerExp)

group2<-c(Exxon, GenMotor)

group3<-c(Intel)

y<-c(group1,group2)

group<-c(rep(1,length(group1)),rep(2,length(group2)))

ydata1<-data.frame(y=y,group=factor(group))

anova(lm(y~group,data=ydata1))

#P= 0.7681>0.05, so fail to reject null. Not sufficient evidence to conclude that that means for Groups 1 and 2 are different.

#(b)

y2<-c(group1,group2,group3)

group<-c(rep(1,length(group1)),rep(2,length(group2)),rep(3,length(group3)))

ydata2<-data.frame(y=y2,group=factor(group))

anova(lm(y~group,data=ydata2))

#p=0.5642>0.05, so fail to reject null. Not sufficient evidence to conclude that means for Groups 1-3 are different.

################Q3

#(a)

attach(ChickWeight)

anova(lm(weight~factor(Time)+Diet,data=ChickWeight))

#Analysis of Variance Table

#Response: weight

# Df Sum Sq Mean Sq F value Pr(>F)

#factor(Time) 11 2067050 187914 147.391 < 2.2e-16 \*\*\*

#Diet 3 129721 43240 33.916 < 2.2e-16 \*\*\*

#Residuals 563 717785 1275

#Both the p-values of Time and Diet are far less than 0.05, reject null, it can be concluded that both Time and Diet are statistically significant.

#(b)

fit<-aov(weight~Diet,data=ChickWeight[Time==2,])

TukeyHSD(fit)

# $Diet

# diff lwr upr p adj

# 2-1 2.15 -1.2811583 5.581158 0.3508353

# 3-1 3.15 -0.2811583 6.581158 0.0823673

# 4-1 4.55 1.1188417 7.981158 0.0050463

# 3-2 1.00 -2.9619603 4.961960 0.9068420

# 4-2 2.40 -1.5619603 6.361960 0.3806460

# 4-3 1.40 -2.5619603 5.361960 0.7825293

# For the pair 4-1, p=0.0050463<0.05, reject null, it can be concluded that the weight means of Diet 1 and 4 are significant different

################Q4

#(a)

prop.test(length(Pfizer[Pfizer>0]), length(Pfizer), 0.55)

#####The p-value= 0.05303>0.05, so fail to reject null. Conclude that the proportion of positive returns of Pfizer is 0.55

#(b)

prop.test(length(Intel[Intel>0]), length(Intel),0.55, alternative="greater")

#pvalue=0.8812>0.05, so fail to reject null. Not sufficient evidence to conclude that proportion of Intel positive returns is greater than 0.55.

#(c)

prop.test(c(length(Pfizer[Pfizer>0]), length(Intel[Intel>0])), c(length(Pfizer), length(Intel)))

#pvalue=0.7221>0.05, so fail to reject null. Conclude that the proportions of positive returns of Pfizer and Intel are same.

#(d)

g1<-c(Citigroup,AmerExp)

g2<-c(Exxon,GenMotor)

g3<-c(Intel)

n1<-sum(g1<(-0.1))

n2<-sum(g1>=-0.1 & g1<0)

n3<-sum(g1>=0 & g1<0.1)

n4<-sum(g1>=0.1)

n5<-sum(g2<(-0.1))

n6<-sum(g2>=-0.1 & g2<0)

n7<-sum(g2>=0 & g2<0.1)

n8<-sum(g2>=0.1)

n9<-sum(g3<(-0.1))

n10<-sum(g3>=-0.1 & g3<0)

n11<-sum(g3>=0 & g3<0.1)

n12<-sum(g3>=0.1)

returns<-matrix(c(n1,n2,n3,n4,n5,n6,n7,n8,n9,n10,n11,n12),nrow=3,byrow=T)

colnames(returns)=c("r<-0.1","-0.1<=r<0","0<=r<0.1","r>=0.1")

rownames(returns)=c("Group 1","Group 2","Group 3")

returns

chisq.test(returns)

#pvalue is 0.0002055<0.05, so reject null. Conclude that they are not independent, group and return range effects are dependent

################Q5

library(MASS)

attach(mcycle)

bic <- NULL

for (i in 1:30){

polyfit <- lm(accel~poly(times,i,raw=T))

bic[i] <- BIC(polyfit)

}

fit5=lm(accel~poly(times,which(bic==min(bic))[1],raw=T))

summary(fit5)

fit5a = step(fit5,k=log(133))

#Selected the fitted model with has the smallest value of Bayesian information criterion

par(mfrow=c(2,2))

plot(fit5)

# From the top-left plot, the mean of residuals of the selected model is zero. As the fitted values along x increase, the residuals are approximately flat, so the disturbances are homoscedastic. The plot on the bottom-left also shows the homoscedasticity of residuals. The normal QQ plot in top-right indicates that the residuals are normally distributed. The residuals vs leverage plot in the bottom-right shows the spread is almost uniform and no point has excess leverage.

