<- read.csv(file.choose(), header=TRUE)

attach()

model eqn: y=b + b1x+b2x+….b3x

Binary var:

LE2013L= (1\*(LE2013 <=74)) LE2013H = (1\*(LE2013>74))

check = data.frame(LE2013,LE2013L,LE2013H)

fix(check)

cut=cut(dataset$variable,br=c(1,9,11,12,13))

table(cut)

model3 = lm(HDI ~ cut)

#Factor anova interpretation:

# as the baseline for the model is intercept y, This means the average number of Sth the group with Sth is

# A scatter plot matrix of all explanatory variables and Y variable

varb = data.frame(x, , , , , Y)

pairs(varb,upper.panel=NULL)

**cor---bind**

attach(fds)

cor(cbind(),use="pairwise.complete.obs")

library(Rcmdr)

vif(model)

### Residuals VS fitted value

plot(residuals(model1) ~ fitted.values(model1), main="Residuals vs.Fitted Value")

AIC Critics step(model,direction='backward',criterion='AIC')

extractAIC(model)

Turkey diff teston groups

TukeyHSD(aov(Y~X))

# when the p-value > alpha (0.05) or when the CI does contain 0, we believe that there is not a significant difference between these two groups.

plot(TukeyHSD(aov(Y~X)))

#the lines that cross over 0 are not significantly different from eachother, while the ones that dont' touch the line ARE significantly different from eachother.

Binary var:

LE2013L= (1\*(LE2013 <=74))

LE2013H = (1\*(LE2013>74))

check = data.frame(LE2013,LE2013L,LE2013H)

fix(check)

cut=cut(fds**$**MEANYRSCH,br=c(1,9,11,12,13))

table(cut)

cut

#One factor anova

model6 = lm(HDI ~ factor(CHINRANK))

anova(model6)

TukeyHSD(aov(HDI ~ factor(CHINRANK)))

F-test t-criteria:

qf(.95, df1=2,df2=269, lower.tail=TRUE)

pf(F-t value on the dropped var from anova, df1=2, df2=269, lower.tail=FALSE)

anova(model2,model1)

**Points removals Steps:**

**dim() come first…also identify the actual model’s n and k**

**#n=30,k=8**

rstandard = rstandard(model11)

leverages = hatvalues(model11)

cooks = cooks.distance(model11)

rstandard[order(rstandard)]

leverages[order(leverages)]

cooks[order(cooks)]

Cooks criteria: qf(alpha,k+1,n-(k+1))

rstudent = rstudent(model1)

rstudent[order(rstudent)]

par(mfrow=c(1,2))

hist(rstandard)

hist(leverages)

layout(matrix(c(1,2,3,4,5,6,7,8,9,10,11,12),byrow=TRUE,ncol=6))

plot.new()

hist(Test1)

hist(Test2)

hist(GPA)

hist(CrHrs)

hist(jobs)

hist(Test3)

plot(Test1,Test3)

plot(Test2,Test3)

plot(GPA,Test3)

plot(CrHrs,Test3)

plot(jobs,Test3)

By checking the hist we use log to improve the normality and checking plots and sqr the variable to rule out the curve pattern.

## Confidence Interval VS Predicted Interval

newdata=data.frame(X=num)

predict(model1, newdata, interval="confidence", level=.95)

predict(model1, newdata, interval="prediction", level=.95)

i.e.expected result:

fit lwr upr

1 83.13821 77.46831 88.80811

## CI= if Xs has features described as above, we have 95% confidence that predicted LOSS would have an average of score at the range of () .

#Focus on the average of a group

##PI= If an individual has Xs values as described, we have 95% confidence that the predicted Y for that individual would have a value at () .

**TIME Series:**

plot(TIME,Y)

#the overall increase that looks like the linear trend shows that this is

#not stationary in the mean since the values of COST\_CLM are increasing

#with time

#The overall spread in the points is also increasing so it is not stationary

#in the variation, It has heterocedasticity

Y

Y1=Y[-1]

diff = Y1-Y[1:(length(Y)-1)]

#The above just shift values via 1 unit time in progress

#USING the diff instead of y’s original values often helps us make

#the obs be independent instead of the dependent originals,

# we can take the diff between each value and then those will fluctuate and not be related to the previous

#we want independence since it is a regression assumption

diff

plot(TIME[1: (n-1)],diff)

#stationary in the mean(it all hovers around or )

??#stationary in the variance

#Words:

#however,it is still not

#stationary in the variance.it has a magaphone shape and has a larger spread in the

#right end than in the left end.