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

attach(risk)

#file name is 'RiskSurvey'

model1 = lm(FIRMCOST ~ ASSUME + CAP + SIZELOG + INDCOST + SOPH +CENTRAL)

summary(model1)

Call:

lm(formula = FIRMCOST ~ ASSUME + CAP + SIZELOG + INDCOST + SOPH +

CENTRAL)

Residuals:

Min 1Q Median 3Q Max

-20.083 -7.665 -1.297 2.368 83.735

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 59.7646 19.0653 3.135 0.002567 \*\*

ASSUME -0.3004 0.2221 -1.353 0.180810

CAP 5.4985 3.8482 1.429 0.157763

SIZELOG -6.8361 1.9229 -3.555 0.000704 \*\*\*

INDCOST 23.0775 8.3039 2.779 0.007092 \*\*

SOPH -0.1367 0.3468 -0.394 0.694777

CENTRAL 0.1329 1.4413 0.092 0.926826

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.56 on 66 degrees of freedom

Multiple R-squared: 0.256, Adjusted R-squared: 0.1883

F-statistic: 3.784 on 6 and 66 DF, p-value: 0.002673

#Compare with the assumption made

#in Page 4 at CH6

# CAP AND CENTRAL BOTH had positive coefficient where we predicted

# negative ones. None of THE 4 VARIABLES hypothesized about were

#significant.THE ONLY 2 Varaibles that are significant are the ones

#we didn't expect

#F p value is small,which is good,but both Rsqrts are small

#which is bad and R2 adj is quite a bit lower than R2 which is also bad

rstandard = rstandard(model1)

leverages = hatvalues(model1)

par(mfrow=c(1,2))

hist(rstandard)

hist(leverages)

dim(risk)

[1] 73 7

rstandard[order(rstandard)]

10 34 9 12 26 33

-1.597929990 -0.986892570 -0.954674609 -0.928301322 -0.882970845 -0.862184859

54 44 68 49 64 73

-0.840954336 -0.802673605 -0.802074925 -0.714009087 -0.702470594 -0.696452766

13 38 24 21 36 30

-0.671934920 -0.658245500 -0.649152946 -0.610045129 -0.596585160 -0.591978593

39 23 3 41 56 17

-0.547705187 -0.513060941 -0.510217504 -0.499627944 -0.494081188 -0.385207548

51 35 28 11 52 4

-0.384843662 -0.363075137 -0.320228249 -0.229436408 -0.224527219 -0.202634162

18 53 60 70 29 5

-0.137786550 -0.124190049 -0.107626097 -0.102486302 -0.102421041 -0.102047813

67 31 48 58 1 62

-0.097636320 -0.088474680 -0.065041774 -0.053555148 -0.050858274 -0.022020075

7 22 47 2 43 27

-0.005230534 0.001262388 0.011449326 0.012124919 0.033291479 0.045088381

25 69 46 14 65 63

0.067661417 0.093924981 0.109583796 0.109687498 0.114233635 0.140866164

37 40 55 42 19 66

0.173635228 0.184345493 0.217286224 0.237156830 0.266358588 0.426460100

45 20 32 8 61 59

0.461771408 0.464372361 0.471528354 0.472954642 0.481703158 0.604343222

57 50 6 71 16 72

0.709263450 0.771541522 1.224971226 1.321946495 1.715467499 3.889170728

15

5.961560722

> #Observation 72 and 15 are outliers, since they exceeds the +2 or -2 cutoff

> leverages[order(leverages)]

14 28 2 48 29 13 43

0.02499847 0.02650646 0.02828247 0.03353386 0.03503212 0.03608470 0.03800023

71 18 8 17 9 62 4

0.03990550 0.04028582 0.04062954 0.04099559 0.04268917 0.04571901 0.04616034

46 3 41 26 63 22 59

0.04695947 0.04735682 0.04808586 0.04894971 0.04993808 0.05099831 0.05242242

56 25 30 60 50 58 53

0.05546918 0.05672534 0.05770825 0.05952930 0.06205475 0.06234805 0.06301324

35 5 15 21 44 69 33

0.06751317 0.06769558 0.06908635 0.07027440 0.07058245 0.07125245 0.07282206

6 68 34 61 66 39 36

0.07298473 0.07357456 0.07391675 0.07451308 0.07543641 0.07576611 0.07773663

19 51 31 40 20 11 23

0.08093866 0.08345730 0.08463513 0.08466978 0.09155068 0.09205094 0.09282739

47 7 64 1 52 42 38

0.09329216 0.09423400 0.09602756 0.10132107 0.10278498 0.10577766 0.10765120

24 55 65 54 49 37 70

0.11242658 0.11509600 0.11545317 0.12041742 0.12140219 0.12275521 0.12371342

27 12 45 67 72 32 10

0.13249131 0.15461426 0.16544594 0.16743582 0.19042529 0.19410633 0.25468578

73 16 57

0.28824795 0.29846901 0.71805901

#High leverage cut off is 3(k+1)/73= 3(6+1)/73=.28767

#obs 73, 16 ,and 57 exceed the cutoff and have high leverage values

> model1 = lm(FIRMCOST ~ ASSUME + CAP + SIZELOG + INDCOST + SOPH +CENTRAL, subset=-c(72,15,57,16,73))

> summary(model1)

Call:

lm(formula = FIRMCOST ~ ASSUME + CAP + SIZELOG + INDCOST + SOPH +

CENTRAL, subset = -c(72, 15, 57, 16, 73))

Residuals:

Min 1Q Median 3Q Max

-13.815 -2.962 -1.020 1.331 19.942

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 28.93532 8.39546 3.447 0.00103 \*\*

ASSUME 0.72419 0.35375 2.047 0.04495 \*

CAP -0.58818 1.61048 -0.365 0.71621

SIZELOG -2.87910 0.88703 -3.246 0.00190 \*\*

INDCOST 7.17056 3.85760 1.859 0.06788 .

SOPH 0.04411 0.14062 0.314 0.75481

CENTRAL -0.61670 0.59004 -1.045 0.30006

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.817 on 61 degrees of freedom

Multiple R-squared: 0.3208, Adjusted R-squared: 0.254

F-statistic: 4.801 on 6 and 61 DF, p-value: 0.0004532

> #has a lower p-value which is good, has a higher Rsqrt which is good

> #We see that ASSUME is now postive(which is bad) but significant(which

> # is good)SOPH is now positive (bad),CENTRAL is now negative(good)

> #but neither of those 2 or CAP ARE SIGNIFICANT still

> hist(FIRMCOST)

> hist(ASSUME)

> hist(CAP)

> hist(SIZELOG)

> # FIRMCOST and ASSUME ARE both right skewed, CAP ONLY takes 0 and 1

> # values,SIZELOG IS symmetric

> hist(INDCOST)

> hist(SOPH)

> # INDCOST is right skewed,but not as bad as FIRMCOST OR ASSUME,so we will

> # leave it alone, SOPH is a little left skewed,but not bad

> hist(CENTRAL)

> #right skewed,but may be taking on values on a scale,we will also

> #leave this alnoe

> lnFIRMCOST = log(FIRMCOST)

> lnASSUME = log(ASSUME)

> model3 = lm(lnFIRMCOST ~ lnASSUME + CAP + SIZELOG + INDCOST + SOPH +CENTRAL)

> lnASSUME = log(ASSUME+.000000000000000000000000000001)

> model3 = lm(lnFIRMCOST ~ lnASSUME + CAP + SIZELOG + INDCOST + SOPH +CENTRAL)

> #It wud not let us take the log of ASSUME COZ A COUPLE OF

> #Observations had 0 values.by adding a tiny amt,we don't really change the value

> #but force the program to work

> summary(model3)

Call:

lm(formula = lnFIRMCOST ~ lnASSUME + CAP + SIZELOG + INDCOST +

SOPH + CENTRAL)

Residuals:

Min 1Q Median 3Q Max

-1.77065 -0.50403 -0.00123 0.44579 3.10929

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.061587 1.118139 6.315 2.62e-08 \*\*\*

lnASSUME 0.022098 0.009956 2.219 0.029900 \*

CAP -0.032078 0.217554 -0.147 0.883227

SIZELOG -0.685163 0.115709 -5.921 1.27e-07 \*\*\*

INDCOST 1.760525 0.479283 3.673 0.000482 \*\*\*

SOPH -0.009728 0.020973 -0.464 0.644285

CENTRAL -0.033907 0.086212 -0.393 0.695365

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8531 on 66 degrees of freedom

Multiple R-squared: 0.554, Adjusted R-squared: 0.5135

F-statistic: 13.66 on 6 and 66 DF, p-value: 5.122e-10

> #p-value is still significant,R sqrt are even higher,ASSUME and

> #the other 2 variables are still significant,none of the others are

> #all are negatives that we expected now except ASSUME

rstandard = rstandard(model3)

leverages = hatvalues(model3)

> hist(rstandard)

> hist(leverages)

> rstandard[order(rstandard)]

10 34 45 42 25 9

-2.296402604 -2.149585742 -1.852971661 -1.812942718 -1.644183100 -1.536195770

23 47 52 44 73 35

-1.499348381 -1.207713409 -1.004684662 -0.991060717 -0.950111009 -0.801998606

3 13 24 26 21 28

-0.771296932 -0.765892505 -0.757626158 -0.706428387 -0.656718194 -0.624951416

68 12 49 41 51 64

-0.611656871 -0.575933416 -0.559750354 -0.558007664 -0.441046721 -0.409502183

57 58 33 32 56 38

-0.400736452 -0.341657205 -0.323957397 -0.262442642 -0.221887473 -0.221490733

7 16 54 11 18 36

-0.154171132 -0.151427922 -0.134644644 -0.088719817 -0.042452259 -0.020528650

30 37 4 17 5 62

-0.001488499 0.011634381 0.014776893 0.051538154 0.079190833 0.080639658

60 48 40 22 53 39

0.095055460 0.166535522 0.270722111 0.299146478 0.309953890 0.348126658

43 29 69 2 20 1

0.390155976 0.394415187 0.405199651 0.476953080 0.527644337 0.550680603

46 67 14 63 65 19

0.558248532 0.587203818 0.682976682 0.721982709 0.744716341 0.887571752

55 66 27 72 59 8

0.893909880 0.910780250 0.916517834 0.955412757 0.993487464 1.068550001

70 31 61 71 6 50

1.149353466 1.238160870 1.376052076 1.682655577 1.772725690 1.852302939

15

3.774374086

> #cutoff =2 and -2

#We merely search 2 ends

> #observation 10,34,15 are outliers

> leverages[order(leverages)]

14 28 2 48 29 13 43

0.02522523 0.02585900 0.02825236 0.03411564 0.03495592 0.03636627 0.03787498

71 17 8 18 26 9 62

0.03971074 0.03975708 0.04063605 0.04125953 0.04264641 0.04296076 0.04576109

4 46 3 41 63 59 22

0.04704931 0.04706093 0.05006603 0.05023708 0.05165406 0.05175090 0.05181779

6 56 25 30 60 53 33

0.05553796 0.05597497 0.05616459 0.05838275 0.06180791 0.06414438 0.06545869

58 68 15 34 50 35 5

0.06592782 0.06700179 0.06758190 0.06774979 0.06783097 0.06800771 0.06903930

44 69 39 21 61 36 66

0.07024424 0.07185812 0.07197929 0.07291230 0.07487836 0.07491100 0.07788112

31 51 40 19 57 23 11

0.07894664 0.07998202 0.08459230 0.08552845 0.08859103 0.09345488 0.09379502

47 20 1 64 52 38 7

0.09909339 0.09930506 0.09959871 0.10228851 0.10265395 0.10604700 0.10607945

65 42 55 49 24 54 73

0.11321338 0.11376261 0.11486571 0.11561951 0.11914820 0.12250458 0.12385366

27 70 37 45 12 72 10

0.12673845 0.12808624 0.14350217 0.16733662 0.16898656 0.19049187 0.25573558

16 67 32

0.26700114 0.56355574 0.57534950

> #cutoff is still .28767,obs 67 and 32 are high leverage values

> # is good)SOPH is now positive (bad),central is now negative(good)

> model4 = lm(lnFIRMCOST ~ lnASSUME + CAP + SIZELOG + INDCOST + SOPH +CENTRAL,-c(67,32,10,34,15))

model4 = lm(lnFIRMCOST ~ lnASSUME + CAP + SIZELOG + INDCOST + SOPH +CENTRAL,subset=-c(67,32,10,34,15))

> summary(model4)

Call:

lm(formula = lnFIRMCOST ~ lnASSUME + CAP + SIZELOG + INDCOST +

SOPH + CENTRAL, subset = -c(67, 32, 10, 34, 15))

Residuals:

Min 1Q Median 3Q Max

-1.44056 -0.43247 -0.00076 0.43228 1.53867

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.794997 1.057506 6.425 2.23e-08 \*\*\*

lnASSUME 0.095035 0.060128 1.581 0.119

CAP -0.177675 0.192151 -0.925 0.359

SIZELOG -0.672300 0.107350 -6.263 4.21e-08 \*\*\*

INDCOST 1.957882 0.454386 4.309 6.07e-05 \*\*\*

SOPH 0.002572 0.018123 0.142 0.888

CENTRAL -0.077998 0.075617 -1.031 0.306

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7082 on 61 degrees of freedom

Multiple R-squared: 0.6164, Adjusted R-squared: 0.5786

F-statistic: 16.33 on 6 and 61 DF, p-value: 4.125e-11

> #Good things- R squrt are high again

> #Good things- R squrt are higher again

> #Bad things-SOPH is back to being positive and lnASSUME IS now not

> #significant