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

attach(fds)

a.1)

model1 = lm(HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +CHINRANK + DL)

summary(model1)

Call:

lm(formula = HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +

CHINRANK + DL)

Residuals:

Min 1Q Median 3Q Max

-0.0256140 -0.0006778 0.0000785 0.0009954 0.0037186

Coefficients:

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

(Intercept) -3.135e-03 3.562e-03 -0.880 0.3800

LE2013 3.560e-05 5.639e-05 0.631 0.5286

MEANYRSCH 1.118e-05 1.624e-04 0.069 0.9452

EYRSCH 1.895e-04 1.535e-04 1.234 0.2188

GNI2013 1.162e-08 1.906e-08 0.610 0.5429

HDI2012 9.992e-01 8.418e-03 118.692 < 2e-16 \*\*\*

CHINRANK 9.994e-04 1.507e-04 6.634 3.84e-10 \*\*\*

DLlow 1.848e-03 1.281e-03 1.443 0.1507

DLmedium 1.841e-03 6.658e-04 2.765 0.0063 \*\*

DLvery high -1.511e-03 7.372e-04 -2.049 0.0419 \*

---

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

Residual standard error: 0.002288 on 177 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 9.617e+04 on 9 and 177 DF, p-value: < 2.2e-16

> step(model1)

Start: AIC=-2264.28

HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 + CHINRANK +

DL

Df Sum of Sq RSS AIC

- MEANYRSCH 1 0.000000 0.000926 -2266.3

- GNI2013 1 0.000002 0.000928 -2265.9

- LE2013 1 0.000002 0.000928 -2265.9

- EYRSCH 1 0.000008 0.000934 -2264.7

<none> 0.000926 -2264.3

- DL 3 0.000067 0.000993 -2257.2

- CHINRANK 1 0.000230 0.001157 -2224.8

- HDI2012 1 0.073728 0.074655 -1445.5

Step: AIC=-2266.27

HDI ~ LE2013 + EYRSCH + GNI2013 + HDI2012 + CHINRANK + DL

Df Sum of Sq RSS AIC

- GNI2013 1 0.000002 0.000929 -2267.8

- LE2013 1 0.000003 0.000929 -2267.7

- EYRSCH 1 0.000009 0.000935 -2266.6

<none> 0.000926 -2266.3

- DL 3 0.000067 0.000994 -2259.2

- CHINRANK 1 0.000241 0.001168 -2225.0

- HDI2012 1 0.153296 0.154223 -1311.8

Step: AIC=-2267.82

HDI ~ LE2013 + EYRSCH + HDI2012 + CHINRANK + DL

Df Sum of Sq RSS AIC

- LE2013 1 0.000002 0.000930 -2269.5

- EYRSCH 1 0.000007 0.000935 -2268.5

<none> 0.000929 -2267.8

- DL 3 0.000068 0.000996 -2260.6

- CHINRANK 1 0.000239 0.001168 -2227.0

- HDI2012 1 0.182819 0.183748 -1281.0

Step: AIC=-2269.46

HDI ~ EYRSCH + HDI2012 + CHINRANK + DL

Df Sum of Sq RSS AIC

- EYRSCH 1 0.000005 0.000936 -2270.4

<none> 0.000930 -2269.5

- DL 3 0.000069 0.000999 -2262.1

- CHINRANK 1 0.000237 0.001168 -2229.0

- HDI2012 1 0.242415 0.243346 -1230.5

Step: AIC=-2270.38

HDI ~ HDI2012 + CHINRANK + DL

Df Sum of Sq RSS AIC

<none> 0.00094 -2270.4

- DL 3 0.00007 0.00100 -2263.3

- CHINRANK 1 0.00024 0.00118 -2229.1

- HDI2012 1 0.32767 0.32861 -1176.3

Call:

lm(formula = HDI ~ HDI2012 + CHINRANK + DL)

Coefficients:

(Intercept) HDI2012 CHINRANK DLlow DLmedium DLvery high

-0.002026 1.005070 0.000997 0.001961 0.001859 -0.001280

> #we choose HDI=-.002026+1.00507HDI2012+.000997CHINRANK+.00196DLlow+.001859DLmedium-.00128DLveryhigh

a.2)

> summary(model1)

Call:

lm(formula = HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +

CHINRANK + DL)

Residuals:

Min 1Q Median 3Q Max

-0.0256140 -0.0006778 0.0000785 0.0009954 0.0037186

Coefficients:

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

(Intercept) -3.135e-03 3.562e-03 -0.880 0.3800

LE2013 3.560e-05 5.639e-05 0.631 0.5286

MEANYRSCH 1.118e-05 1.624e-04 0.069 0.9452

EYRSCH 1.895e-04 1.535e-04 1.234 0.2188

GNI2013 1.162e-08 1.906e-08 0.610 0.5429

HDI2012 9.992e-01 8.418e-03 118.692 < 2e-16 \*\*\*

CHINRANK 9.994e-04 1.507e-04 6.634 3.84e-10 \*\*\*

DLlow 1.848e-03 1.281e-03 1.443 0.1507

DLmedium 1.841e-03 6.658e-04 2.765 0.0063 \*\*

DLvery high -1.511e-03 7.372e-04 -2.049 0.0419 \*

---

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

Residual standard error: 0.002288 on 177 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 9.617e+04 on 9 and 177 DF, p-value: < 2.2e-16

> model1a = lm(HDI ~ LE2013 + EYRSCH + GNI2013 + HDI2012 +CHINRANK + DL)

> summary(model1a)

Call:

lm(formula = HDI ~ LE2013 + EYRSCH + GNI2013 + HDI2012 + CHINRANK +

DL)

Residuals:

Min 1Q Median 3Q Max

-0.0256059 -0.0006744 0.0000838 0.0009932 0.0037212

Coefficients:

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

(Intercept) -3.131e-03 3.552e-03 -0.882 0.37922

LE2013 3.347e-05 4.696e-05 0.713 0.47693

EYRSCH 1.861e-04 1.451e-04 1.283 0.20114

GNI2013 1.099e-08 1.672e-08 0.658 0.51155

HDI2012 9.996e-01 5.824e-03 171.628 < 2e-16 \*\*\*

CHINRANK 1.002e-03 1.471e-04 6.811 1.44e-10 \*\*\*

DLlow 1.855e-03 1.274e-03 1.457 0.14699

DLmedium 1.843e-03 6.634e-04 2.777 0.00607 \*\*

DLvery high -1.505e-03 7.303e-04 -2.061 0.04076 \*

---

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

Residual standard error: 0.002281 on 178 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 1.088e+05 on 8 and 178 DF, p-value: < 2.2e-16

> model1b = lm(HDI ~ EYRSCH + GNI2013 + HDI2012 +CHINRANK + DL)

> summary(model1b)

Call:

lm(formula = HDI ~ EYRSCH + GNI2013 + HDI2012 + CHINRANK + DL)

Residuals:

Min 1Q Median 3Q Max

-0.0257487 -0.0006970 0.0001211 0.0009940 0.0039285

Coefficients:

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

(Intercept) -1.829e-03 3.042e-03 -0.601 0.54841

EYRSCH 1.555e-04 1.384e-04 1.124 0.26264

GNI2013 8.450e-09 1.631e-08 0.518 0.60499

HDI2012 1.002e+00 4.942e-03 202.707 < 2e-16 \*\*\*

CHINRANK 9.913e-04 1.461e-04 6.783 1.66e-10 \*\*\*

DLlow 1.799e-03 1.269e-03 1.417 0.15811

DLmedium 1.832e-03 6.623e-04 2.766 0.00627 \*\*

DLvery high -1.454e-03 7.257e-04 -2.003 0.04668 \*

---

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

Residual standard error: 0.002278 on 179 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 1.247e+05 on 7 and 179 DF, p-value: < 2.2e-16

> model1c = lm(HDI ~ EYRSCH + HDI2012 +CHINRANK + DL)

> summary(model1c)

Call:

lm(formula = HDI ~ EYRSCH + HDI2012 + CHINRANK + DL)

Residuals:

Min 1Q Median 3Q Max

-0.0257243 -0.0007185 0.0001166 0.0009361 0.0039286

Coefficients:

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

(Intercept) -0.0020905 0.0029934 -0.698 0.48586

EYRSCH 0.0001354 0.0001325 1.021 0.30845

HDI2012 1.0026744 0.0046299 216.563 < 2e-16 \*\*\*

CHINRANK 0.0009864 0.0001456 6.777 1.69e-10 \*\*\*

DLlow 0.0018640 0.0012606 1.479 0.14097

DLmedium 0.0018339 0.0006610 2.775 0.00611 \*\*

DLvery high -0.0012957 0.0006572 -1.971 0.05021 .

---

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

Residual standard error: 0.002274 on 180 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 1.461e+05 on 6 and 180 DF, p-value: < 2.2e-16

> model1d = lm(HDI ~ HDI2012 +CHINRANK + DL)

> summary(model1d)

Call:

lm(formula = HDI ~ HDI2012 + CHINRANK + DL)

Residuals:

Min 1Q Median 3Q Max

-0.0257853 -0.0007353 0.0001548 0.0010882 0.0038902

Coefficients:

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

(Intercept) -0.0020263 0.0029931 -0.677 0.49929

HDI2012 1.0050701 0.0039923 251.751 < 2e-16 \*\*\*

CHINRANK 0.0009970 0.0001452 6.867 1.02e-10 \*\*\*

DLlow 0.0019610 0.0012572 1.560 0.12054

DLmedium 0.0018592 0.0006606 2.814 0.00543 \*\*

DLvery high -0.0012804 0.0006571 -1.948 0.05291 .

---

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

Residual standard error: 0.002274 on 181 degrees of freedom

Multiple R-squared: 0.9998, Adjusted R-squared: 0.9998

F-statistic: 1.752e+05 on 5 and 181 DF, p-value: < 2.2e-16

#each time we delete the variable with the largest p-value.

#we finally have lm(formula = HDI ~ HDI2012 + CHINRANK + DL) as our model in the backward approach

b.

> rstandard = rstandard(model1)

> rstandard[order(rstandard)]

185 32 119 9 55

-11.540513756 -1.608970199 -1.489775140 -1.128530698 -1.082456638

129 133 130 177 29

-0.954218961 -0.945472684 -0.931420907 -0.921335595 -0.918265215

53 124 120 4 154

-0.898381097 -0.869983175 -0.805812218 -0.671456988 -0.616421434

148 95 10 27 24

-0.606722455 -0.586794196 -0.568834404 -0.567117021 -0.565308849

50 6 115 26 30

-0.546824803 -0.541268539 -0.534654434 -0.514445259 -0.509596234

20 25 59 52 79

-0.499345571 -0.490489244 -0.488366358 -0.468879430 -0.455720487

176 166 42 28 125

-0.449785607 -0.447754192 -0.445735055 -0.422693786 -0.415375656

131 65 58 101 83

-0.415047883 -0.386953758 -0.379549005 -0.372740288 -0.354420426

167 138 1 78 57

-0.327325185 -0.319688025 -0.319317580 -0.315318180 -0.313863799

84 85 31 51 77

-0.313769253 -0.308598538 -0.301566718 -0.293972193 -0.288712797

73 155 80 38 104

-0.287157109 -0.285988414 -0.281675384 -0.281292475 -0.264831674

11 158 47 103 182

-0.245328857 -0.241027986 -0.236056124 -0.226831329 -0.218966520

123 64 149 107 37

-0.218112580 -0.213505368 -0.193600285 -0.186318744 -0.169756604

106 110 169 157 3

-0.165950850 -0.154291001 -0.143002718 -0.131610111 -0.127237280

108 122 146 8 22

-0.118686198 -0.118394770 -0.098148408 -0.070741874 -0.068177598

145 63 56 21 168

-0.056028165 -0.046993146 -0.045303612 -0.044652626 -0.041256963

179 81 75 128 132

-0.032603120 -0.028183841 -0.006752356 0.006416878 0.010874039

159 23 90 117 88

0.017611466 0.024090979 0.025549747 0.030239886 0.030377511

109 2 118 135 67

0.030561684 0.031667437 0.032600995 0.035086993 0.050502631

61 134 33 54 143

0.055214433 0.059987567 0.065769922 0.066736472 0.069772611

170 70 86 142 121

0.078114158 0.088197147 0.101185555 0.110495540 0.113217650

71 93 60 7 66

0.126277322 0.135441213 0.136191667 0.136435470 0.155814731

94 89 43 91 13

0.155978072 0.156336429 0.162605113 0.168197607 0.170886297

92 165 140 87 102

0.177084000 0.186800811 0.192877242 0.235458821 0.243593839

72 174 100 152 160

0.247195255 0.248910784 0.286853101 0.297508297 0.298593386

5 15 180 16 113

0.312067277 0.321696327 0.325090120 0.338507862 0.342629453

172 12 151 14 136

0.343542519 0.370908394 0.391359578 0.394271206 0.401976529

105 178 114 19 144

0.402400241 0.433422691 0.437573956 0.440832544 0.448444524

153 69 137 116 68

0.455597484 0.462401256 0.468595765 0.469871777 0.491287356

181 49 127 35 41

0.509936614 0.524591008 0.528096622 0.528538033 0.561314541

139 36 74 40 39

0.575002604 0.575226814 0.577269937 0.577806183 0.581723364

62 184 82 147 98

0.587606173 0.598123349 0.598558753 0.600630610 0.604127821

44 45 48 164 99

0.612484117 0.623505242 0.632816055 0.644498140 0.646301609

97 187 96 18 150

0.653163520 0.656137641 0.692609074 0.722538785 0.751042077

76 162 156 161 163

0.752480055 0.777683626 0.783699829 0.834144881 0.840122146

17 46 34 186 112

0.860978271 0.865252931 0.883366940 0.886342269 0.887066349

141 171 175 111 126

0.925238233 0.932850498 1.259801879 1.259864464 1.299716092

183 173

1.593653078 1.660624122

> #above showed 185 is the outlier.

> #as it is smaller than -2

> leverages = hatvalues(model1)

> leverages[order(leverages)]

82 84 77 92 22 21 20

0.02060151 0.02144284 0.02160794 0.02174873 0.02191423 0.02290050 0.02393424

78 86 58 67 61 10 172

0.02399873 0.02430651 0.02450520 0.02521457 0.02565843 0.02569252 0.02582169

73 14 89 8 169 24 85

0.02585606 0.02600274 0.02709617 0.02713815 0.02715195 0.02724161 0.02761914

71 62 12 98 102 26 93

0.02781538 0.02839641 0.02868209 0.02899565 0.02903892 0.02951848 0.02956316

54 69 25 59 99 6 52

0.02984615 0.02987094 0.03045337 0.03067992 0.03078314 0.03082927 0.03126157

111 68 4 115 16 17 19

0.03158563 0.03204964 0.03250996 0.03251182 0.03281634 0.03304340 0.03309850

63 15 94 65 108 160 165

0.03315526 0.03330771 0.03371435 0.03382770 0.03387910 0.03437251 0.03458617

28 176 125 29 121 157 164

0.03472795 0.03521986 0.03550429 0.03554976 0.03617889 0.03671189 0.03683785

177 113 5 39 127 27 3

0.03704796 0.03740546 0.03747445 0.03763774 0.03772090 0.03772268 0.03779475

117 124 44 100 51 129 81

0.03804970 0.03843450 0.03852670 0.03870247 0.03871173 0.03928449 0.03972405

87 173 139 33 38 60 152

0.04090374 0.04186320 0.04191790 0.04197419 0.04211879 0.04217170 0.04221546

90 135 72 34 179 107 74

0.04242442 0.04255525 0.04270902 0.04297100 0.04312889 0.04319238 0.04325720

181 32 168 133 138 70 53

0.04331471 0.04410792 0.04415086 0.04442081 0.04447001 0.04485521 0.04491264

132 43 75 79 101 122 147

0.04518836 0.04520103 0.04532282 0.04578955 0.04586814 0.04594766 0.04613218

13 155 64 23 36 116 88

0.04631407 0.04673681 0.04679588 0.04694022 0.04730284 0.04744874 0.04752415

45 91 143 57 41 83 66

0.04768038 0.04826638 0.04856770 0.04895869 0.04923076 0.04942378 0.04956994

47 166 171 76 48 178 120

0.04983379 0.05059073 0.05073673 0.05202051 0.05216916 0.05266955 0.05276876

35 137 161 40 110 49 154

0.05286287 0.05293424 0.05298092 0.05298142 0.05334892 0.05430075 0.05469686

114 126 105 140 42 158 50

0.05548763 0.05577592 0.05582855 0.05597066 0.05619434 0.05647024 0.05728549

7 128 150 175 180 185 170

0.05731602 0.05757660 0.05798809 0.05803933 0.05831646 0.05872982 0.05896318

167 142 112 184 109 1 103

0.05934085 0.06008927 0.06010443 0.06045997 0.06094039 0.06377173 0.06383597

80 174 162 95 130 56 37

0.06576784 0.06609703 0.06766745 0.06767664 0.06830406 0.06921548 0.06974557

159 30 96 2 134 163 11

0.07006619 0.07035698 0.07076294 0.07247482 0.07690920 0.07803748 0.08019210

151 123 153 145 136 106 146

0.08094556 0.08103176 0.08117402 0.08282378 0.08344581 0.08385036 0.08389221

183 118 97 104 187 141 149

0.08449647 0.08829200 0.08968916 0.09077543 0.09200627 0.09542092 0.09594336

119 148 156 131 182 9 186

0.09622094 0.10577420 0.10926850 0.11238284 0.11265417 0.11846499 0.12200167

18 55 46 144 31

0.13822431 0.14245038 0.15748973 0.18140629 0.35676149

> dim(fds)

[1] 187 9

> 3\*(9+1)/187

[1] 0.1604278

> #3(k+1)/n are the cutoff

> #obs ,and exceed the cutoff and have high leverage values

> #observation 144 and 31 ,and exceed the cutoff and therefore are high leverage

> #points

> CooksD = cooks.distance(model1)

> CooksD [order(CooksD)]

75 128 132 159 23 90 81

2.164567e-07 2.515634e-07 5.596166e-07 2.336940e-06 2.858472e-06 2.892118e-06 3.285927e-06

117 88 21 179 135 109 67

3.617087e-06 4.604312e-06 4.673048e-06 4.791077e-06 5.471820e-06 6.061312e-06 6.597364e-06

63 2 168 61 118 22 54

7.572943e-06 7.835871e-06 7.862204e-06 8.028308e-06 1.029265e-05 1.041436e-05 1.370169e-05

8 56 33 143 86 145 134

1.395989e-05 1.526229e-05 1.895221e-05 2.485076e-05 2.550623e-05 2.834752e-05 2.998171e-05

70 170 71 121 108 93 169

3.653026e-05 3.823260e-05 4.562334e-05 4.811572e-05 4.939702e-05 5.588372e-05 5.707482e-05

3 157 122 89 92 142 60

6.359054e-05 6.601293e-05 6.750815e-05 6.807040e-05 6.971756e-05 7.805484e-05 8.166473e-05

94 146 7 165 43 66 110

8.488606e-05 8.821482e-05 1.131786e-04 1.250105e-04 1.251713e-04 1.266238e-04 1.341581e-04

13 91 107 102 77 84 37

1.418149e-04 1.434726e-04 1.567096e-04 1.774644e-04 1.840910e-04 2.157332e-04 2.160570e-04

73 140 64 87 78 106 85

2.188660e-04 2.205652e-04 2.237893e-04 2.364453e-04 2.444759e-04 2.520563e-04 2.704964e-04

72 47 172 160 100 158 38

2.726189e-04 2.922502e-04 3.128292e-04 3.173672e-04 3.312836e-04 3.476954e-04 3.479209e-04

51 103 15 58 5 16 152

3.480177e-04 3.508485e-04 3.565732e-04 3.618837e-04 3.791574e-04 3.887932e-04 3.901233e-04

149 155 12 14 123 174 113

3.977693e-04 4.009989e-04 4.062401e-04 4.150033e-04 4.194848e-04 4.384980e-04 4.561849e-04

138 57 65 11 80 182 20

4.756370e-04 5.071224e-04 5.242470e-04 5.247250e-04 5.585430e-04 6.087091e-04 6.114248e-04

125 28 83 180 69 19 101

6.351299e-04 6.428078e-04 6.531102e-04 6.544760e-04 6.583508e-04 6.652324e-04 6.679060e-04

167 1 104 52 176 82 59

6.758967e-04 6.945317e-04 7.002240e-04 7.094579e-04 7.385337e-04 7.536212e-04 7.548811e-04

25 68 26 10 24 6 105

7.556587e-04 7.991738e-04 8.049799e-04 8.532620e-04 8.949513e-04 9.319412e-04 9.574629e-04

115 79 62 178 166 98 127

9.605986e-04 9.965963e-04 1.009130e-03 1.044435e-03 1.068309e-03 1.089856e-03 1.093221e-03

116 114 181 42 137 27 39

1.099753e-03 1.124842e-03 1.177331e-03 1.182942e-03 1.227307e-03 1.260804e-03 1.323482e-03

99 151 139 136 44 74 4

1.326668e-03 1.348974e-03 1.446560e-03 1.471119e-03 1.503191e-03 1.506680e-03 1.514978e-03

35 49 164 41 36 147 50

1.559159e-03 1.580136e-03 1.588687e-03 1.631451e-03 1.642898e-03 1.744740e-03 1.817025e-03

153 40 45 30 131 154 48

1.833777e-03 1.867796e-03 1.946423e-03 1.965366e-03 2.181076e-03 2.198603e-03 2.204134e-03

184 95 17 124 76 29 177

2.302153e-03 2.499447e-03 2.533157e-03 3.025269e-03 3.107175e-03 3.108087e-03 3.265843e-03

150 34 120 96 129 53 161

3.472249e-03 3.503746e-03 3.617334e-03 3.653051e-03 3.723252e-03 3.795305e-03 3.892636e-03

133 97 148 187 162 144 171

4.155447e-03 4.203336e-03 4.354243e-03 4.362390e-03 4.389498e-03 4.456577e-03 4.651145e-03

112 31 111 163 130 156 18

5.031982e-03 5.043973e-03 5.176974e-03 5.974132e-03 6.360105e-03 7.534382e-03 8.373606e-03

141 175 126 186 32 173 46

9.030342e-03 9.778992e-03 9.978579e-03 1.091629e-02 1.194548e-02 1.204890e-02 1.399469e-02

9 55 183 119 185

1.711501e-02 1.946370e-02 2.344046e-02 2.362919e-02 8.309878e-01

> 187-10

[1] 177

> qf(.95,10,177)

[1] 1.884525

> #k+1=10,n-(k+1)=177

> #No observation are high enough to exceed 1.8845

> #There are no influential point

c.1)

#Remove all the identified points in Part b)

model2 = lm(HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +CHINRANK + DL,subset=-c(185,144,31))

summary(model2)

Call:

lm(formula = HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +

CHINRANK + DL, subset = -c(185, 144, 31))

Residuals:

Min 1Q Median 3Q Max

-0.0037406 -0.0007559 0.0000668 0.0009160 0.0032060

Coefficients:

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

(Intercept) 4.739e-03 1.897e-03 2.498 0.01341 \*

LE2013 9.088e-06 3.003e-05 0.303 0.76254

MEANYRSCH 1.271e-04 8.329e-05 1.526 0.12877

EYRSCH 1.511e-04 7.893e-05 1.914 0.05725 .

GNI2013 2.522e-08 1.160e-08 2.174 0.03104 \*

HDI2012 9.903e-01 4.431e-03 223.487 < 2e-16 \*\*\*

CHINRANK 9.586e-04 7.679e-05 12.483 < 2e-16 \*\*\*

DLlow 6.970e-05 6.547e-04 0.106 0.91533

DLmedium 9.404e-04 3.351e-04 2.806 0.00559 \*\*

DLvery high -7.485e-04 3.796e-04 -1.972 0.05021 .

---

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

Residual standard error: 0.001143 on 174 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 3.713e+05 on 9 and 174 DF, p-value: < 2.2e-16

c.2)

> step(model2)

Start: AIC=-2483.1

HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 + CHINRANK +

DL

Df Sum of Sq RSS AIC

- LE2013 1 0.000000 0.000228 -2485.0

<none> 0.000227 -2483.1

- MEANYRSCH 1 0.000003 0.000230 -2482.7

- EYRSCH 1 0.000005 0.000232 -2481.3

- GNI2013 1 0.000006 0.000234 -2480.2

- DL 3 0.000033 0.000260 -2464.1

- CHINRANK 1 0.000204 0.000431 -2367.4

- HDI2012 1 0.065271 0.065498 -1443.1

Step: AIC=-2485

HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012 + CHINRANK + DL

Df Sum of Sq RSS AIC

<none> 0.000228 -2485.0

- MEANYRSCH 1 0.000004 0.000231 -2484.2

- EYRSCH 1 0.000005 0.000233 -2482.8

- GNI2013 1 0.000007 0.000234 -2481.7

- DL 3 0.000033 0.000260 -2466.1

- CHINRANK 1 0.000204 0.000431 -2369.4

- HDI2012 1 0.130325 0.130552 -1318.2

Call:

lm(formula = HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012 + CHINRANK +

DL, subset = -c(185, 144, 31))

Coefficients:

(Intercept) MEANYRSCH EYRSCH GNI2013 HDI2012 CHINRANK DLlow

5.006e-03 1.129e-04 1.402e-04 2.372e-08 9.912e-01 9.579e-04 6.170e-05

DLmedium DLvery high

9.403e-04 -7.255e-04

> #we have lm(formula = HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012 + CHINRANK +

> DL, subset = -c(185, 144, 31))

> #as our stepwise's solution

> summary(model2)

Call:

lm(formula = HDI ~ LE2013 + MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +

CHINRANK + DL, subset = -c(185, 144, 31))

Residuals:

Min 1Q Median 3Q Max

-0.0037406 -0.0007559 0.0000668 0.0009160 0.0032060

Coefficients:

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

(Intercept) 4.739e-03 1.897e-03 2.498 0.01341 \*

LE2013 9.088e-06 3.003e-05 0.303 0.76254

MEANYRSCH 1.271e-04 8.329e-05 1.526 0.12877

EYRSCH 1.511e-04 7.893e-05 1.914 0.05725 .

GNI2013 2.522e-08 1.160e-08 2.174 0.03104 \*

HDI2012 9.903e-01 4.431e-03 223.487 < 2e-16 \*\*\*

CHINRANK 9.586e-04 7.679e-05 12.483 < 2e-16 \*\*\*

DLlow 6.970e-05 6.547e-04 0.106 0.91533

DLmedium 9.404e-04 3.351e-04 2.806 0.00559 \*\*

DLvery high -7.485e-04 3.796e-04 -1.972 0.05021 .

---

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

Residual standard error: 0.001143 on 174 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 3.713e+05 on 9 and 174 DF, p-value: < 2.2e-16

> model2 = lm(HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012 +CHINRANK + DL,subset=-c(185,144,31))

> summary(model2)

Call:

lm(formula = HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012 + CHINRANK +

DL, subset = -c(185, 144, 31))

Residuals:

Min 1Q Median 3Q Max

-0.0037508 -0.0007536 0.0000691 0.0009228 0.0032122

Coefficients:

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

(Intercept) 5.006e-03 1.675e-03 2.989 0.00320 \*\*

MEANYRSCH 1.129e-04 6.856e-05 1.647 0.10144

EYRSCH 1.402e-04 7.009e-05 2.000 0.04702 \*

GNI2013 2.372e-08 1.047e-08 2.266 0.02469 \*

HDI2012 9.912e-01 3.131e-03 316.619 < 2e-16 \*\*\*

CHINRANK 9.579e-04 7.656e-05 12.512 < 2e-16 \*\*\*

DLlow 6.170e-05 6.524e-04 0.095 0.92476

DLmedium 9.403e-04 3.343e-04 2.813 0.00547 \*\*

DLvery high -7.255e-04 3.709e-04 -1.956 0.05207 .

---

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

Residual standard error: 0.00114 on 175 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 4.199e+05 on 8 and 175 DF, p-value: < 2.2e-16

c.3)

> model2 = lm(HDI ~ EYRSCH + GNI2013 + HDI2012 +CHINRANK + DL,subset=-c(185,144,31))

> summary(model2a)

Call:

lm(formula = HDI ~ EYRSCH + GNI2013 + HDI2012 + CHINRANK + DL,

subset = -c(185, 144, 31))

Residuals:

Min 1Q Median 3Q Max

-0.0037283 -0.0007546 0.0000307 0.0008967 0.0031791

Coefficients:

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

(Intercept) 4.091e-03 1.588e-03 2.577 0.01079 \*

EYRSCH 1.260e-04 6.990e-05 1.802 0.07320 .

GNI2013 1.786e-08 9.895e-09 1.805 0.07274 .

HDI2012 9.942e-01 2.567e-03 387.376 < 2e-16 \*\*\*

CHINRANK 9.871e-04 7.483e-05 13.191 < 2e-16 \*\*\*

DLlow 1.970e-04 6.504e-04 0.303 0.76231

DLmedium 9.738e-04 3.353e-04 2.905 0.00415 \*\*

DLvery high -6.957e-04 3.723e-04 -1.869 0.06330 .

---

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

Residual standard error: 0.001146 on 176 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 4.753e+05 on 7 and 176 DF, p-value: < 2.2e-16

model2b = lm(HDI ~ HDI2012 +CHINRANK + DL,subset=-c(185,144,31))

summary(model2b)

Call:

lm(formula = HDI ~ HDI2012 + CHINRANK + DL, subset = -c(185,

144, 31))

Residuals:

Min 1Q Median 3Q Max

-0.0041244 -0.0007757 0.0000914 0.0008579 0.0030804

Coefficients:

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

(Intercept) 3.513e-03 1.556e-03 2.258 0.02519 \*

HDI2012 9.976e-01 2.076e-03 480.521 < 2e-16 \*\*\*

CHINRANK 9.844e-04 7.528e-05 13.076 < 2e-16 \*\*\*

DLlow 4.189e-04 6.481e-04 0.646 0.51886

DLmedium 1.001e-03 3.380e-04 2.961 0.00348 \*\*

DLvery high -3.878e-04 3.395e-04 -1.142 0.25489

---

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

Residual standard error: 0.001156 on 178 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 6.537e+05 on 5 and 178 DF, p-value: < 2.2e-16

#Thus,we have lm(formula = HDI ~ HDI2012 + CHINRANK + DL, subset = -c(185,

144, 31)) as our model3 (model2b)

> model3=model2b

> model4 = lm(HDI ~ HDI2012 +CHINRANK,subset=-c(185,144,31))

> anova(model4,model3)

Analysis of Variance Table

Model 1: HDI ~ HDI2012 + CHINRANK

Model 2: HDI ~ HDI2012 + CHINRANK + DL

Res.Df RSS Df Sum of Sq F Pr(>F)

1 181 0.00026379

2 178 0.00023784 3 2.5946e-05 6.4725 0.00035 \*\*\*

---

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

#Thus,we want model4:lm(formula = HDI ~ HDI2012 + CHINRANK, subset = -c(185,

144, 31)) as our model3 (model2b)

c.4)

s is 0.002288 in model1,.00114 in model2 and .001156 in model3.

Thus,model1 has the worst and model3 and model2 are nearly the same.

> #From model1 to model2,R sqr and R sqr adj both improved a bit

> #F-p value remains significant < alpha=.05.

> #HDI2012 and CHINRANK remains significant by their p-values

> #In model2,intercpt and GNI2013 become significant by alpha=.05

> #There is very little improvement from model2 to model3

d.1)

**cor(cbind(**LE2013, MEANYRSCH, EYRSCH, GNI2013 , HDI2012 ,CHINRANK , DL**))**

LE2013 MEANYRSCH EYRSCH GNI2013 HDI2012

LE2013 1.00000000 0.72873530 0.755600107 0.60790147 0.90160224

MEANYRSCH 0.72873530 1.00000000 0.799798926 0.55966668 0.89719271

EYRSCH 0.75560011 0.79979893 1.000000000 0.59328254 0.89455721

GNI2013 0.60790147 0.55966668 0.593282539 1.00000000 0.72620484

HDI2012 0.90160224 0.89719271 0.894557215 0.72620484 1.00000000

CHINRANK -0.07297559 0.06930749 0.001250648 -0.06538076 -0.03438763

DL 0.31065119 0.29892314 0.350477429 0.50800201 0.36248640

CHINRANK DL

LE2013 -0.072975587 0.310651188

MEANYRSCH 0.069307491 0.298923144

EYRSCH 0.001250648 0.350477429

GNI2013 -0.065380759 0.508002010

HDI2012 -0.034387635 0.362486396

CHINRANK 1.000000000 -0.001120913

DL -0.001120913 1.000000000

There are high correlations between HDI2012/LE2013=.9016,LE2013/ MEANYRSCH = 0.728735 andLE2013/EYRSCH = 0.7477,HDI2012/MEANYRSCH = .89719 and EYRSCH/MEANYRSCH = .79980. AND many other median correlation. Multicollinearity may be a big problem here.

> library(Rcmdr)

> vif(model1)

GVIF Df GVIF^(1/(2\*Df))

LE2013 8.814486 1 2.968920

MEANYRSCH 8.766902 1 2.960895

EYRSCH 6.770118 1 2.601945

GNI2013 4.397268 1 2.096966

HDI2012 61.690451 1 7.854327

CHINRANK 1.065370 1 1.032168

DL 22.112312 3 1.675351

# Any vif value that exceeds 10 denotes severe multicollinearity

#Thus,HDI2012 and DL has a very high multicolinearity effect at the model and thus are the

#variance inflation factors

Coefficients:

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

(Intercept) 4.739e-03 1.897e-03 2.498 0.01341 \*

LE2013 9.088e-06 3.003e-05 0.303 0.76254

MEANYRSCH 1.271e-04 8.329e-05 1.526 0.12877

EYRSCH 1.511e-04 7.893e-05 1.914 0.05725 .

GNI2013 2.522e-08 1.160e-08 2.174 0.03104 \*

HDI2012 9.903e-01 4.431e-03 223.487 < 2e-16 \*\*\*

CHINRANK 9.586e-04 7.679e-05 12.483 < 2e-16 \*\*\*

DLlow 6.970e-05 6.547e-04 0.106 0.91533

DLmedium 9.404e-04 3.351e-04 2.806 0.00559 \*\*

DLvery high -7.485e-04 3.796e-04 -1.972 0.05021 .

---

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

Residual standard error: 0.001143 on 174 degrees of freedom

Multiple R-squared: 0.9999, Adjusted R-squared: 0.9999

F-statistic: 3.713e+05 on 9 and 174 DF, p-value: < 2.2e-16

1.e1)

Ho:Beta0=Beta1=Beta2=…Beta9=0;

H1:at least one Beta!=0;

#model2 is significant, because the s= 0.001143 on 174 degrees of freedom,

#Rsqrt is .9999,and Rsqrt-adj is .9999,so there almost no changes here and the p-value

#is almost 0

1.e2)

Ho:Beta2=0;

H1:Beta2!=0;

#Since our p-value = .12877 >.05.We can’t reject Ho and conclude MEANYRSCH is not significant in the model.

1.f.1)

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



> #We want to see random scatter plot here to rule out model inadequacy. Yet,it did show model inadequacy because it spread horizontally above -.005,where shows a pattern.

> plot(residuals(model2) ~ fitted.values(model2), main="Residuals vs.Fitted Value")



# We want to see random scatter plot here to rule out model inadequacy.

# No it does not directly show model inadequency,since it spread through the whole

> # picture.and it spreads through positive and negative values around 0

> #Thus,it look way better than the model1,as it spreads through the plot table

g)

plot(residuals(model1) ~ LE2013, main="Residuals vs.LE2013")





> We want to see random scatter plot here to rule out model inadequacy.

#But we get a vertical spread which show it affects little in the residuals as LE2013 values spread.

> plot(residuals(model1) ~ MEANYRSCH, main="Residuals vs.MEANYRSCH")



> #It results almost the same as the previous one.

h.

> extractAIC(model1)

[1] 10.000 -2264.278

> extractAIC(model2)

[1] 8.000 -2484.174

> extractAIC(model3)

[1] 6.000 -2482.826

#The smaller,the better,thus,I prefer model2,

#As expected,model3 is less favorable, because it has a heavy muti-collinearity.

i.1)

attach(model3)



#overall,we found the data scatter pretty evenly in the picture

varb = data.frame(HDI2012,CHINRANK, DL,HDI)

pairs(varb,upper.panel=NULL)



#It appears all variables don’t suffer from the limited sampling region error. Obviously, HDI2012 don’t since it has X‘s spread through the whole plot. Yet,both CHINRANK and DL are categorical variable in integers and they have values in each category, so there are no limited sampling region effect on them, either.

i.2)

#I don’t think this model suffers from the omitted variables,because the F-p value is almost 0 and both R sqr and R sqr adj are almost one. Thus,in the predicted model,the X variables explains well on Y variables.