1a)

HDI LE2013 MEANYRSCH EYRSCH GNI2013 HDI2012

HDI 1.0000000 0.9014085 0.8976834 0.8950789 0.7250635 0.9998629

LE2013 0.9014085 1.0000000 0.7287353 0.7556001 0.6079015 0.9016022

MEANYRSCH 0.8976834 0.7287353 1.0000000 0.7997989 0.5596667 0.8971927

EYRSCH 0.8950789 0.7556001 0.7997989 1.0000000 0.5932825 0.8945572

GNI2013 0.7250635 0.6079015 0.5596667 0.5932825 1.0000000 0.7262048

HDI2012 0.9998629 0.9016022 0.8971927 0.8945572 0.7262048 1.0000000

HDI 2012 and LE2013 are the two variables that are most highly correlated with our y value that is HDI, with correlations of .99986290 and .90140854 respectively.

b)

It is noticeable that HDI is extremely correlated to HDI2012 and relatively correlated to LE2013, MEANYRSCH, and EYRSCH.

c)

Yes, they are coincide with each other, since the correlations of each pair of variable could be reflected by a scatter plot where we could see how plots appears to fit in a line.

d) In our linear model, the residual standard error came out to be 0.002595. This demonstrates the typical size for an error in the model. The coefficient of determination, R-sqr, measures the proportion of the variability that is explained by the regression line. In our model, 0.9997 of the variability in the y values that are explained by the x values. This is a very large number, which is good for the model. R-sqr adjusted is the proportion of variation explained by the regression, having taken in account for the penalty on the 5 explanatory variables in the model.

e) The variable HDI2012 is the only significant variable in this model. We know this because the p-value is 2e-16. Since this is smaller than 0.05, we can confirm that this is a significant variable.

f)**LE2013 is the 2nd variable, has 3.281e-05 as its estimate coefficient—the slope, for every 1 more year added to the LE2013, HDI in 2013 would increase by 3.281e-05in index, holding the other X’s variables constant.**

g) The residual standard error, s, of model2 is 0.002589. This is extremely similar to that of model 1. The coefficient of determination, R2, is 0.9997. Again, this did not change from our first model. R2 adjusted remained at 0.9997 as well.

h) The values in part g are very similar to the values in part d. The p value of LE2013 was quite high at 0.604, meaning it was too high to be a significant variable in the model. This implies that there should not be much of a change if we take it out because of its insignificance.

i) MEANYRSCH has 1.298e^04 as its estimate coefficient—the slope, for every 1 more year added to the MEANYRSCH, HDI in 2013 would increase by 1.298e^04 in index, holding the other X’s variables constant.

j) I pick LE2013 as the explanatory variable conditioned by the other 4 variables and HDI as the Y variable. The partial correlation coefficient between HDI and LE2013 is 0.03861117. This is a low partial correlation coefficient. This is so different from the original correlation coefficient between them because the partial correlation coefficient takes into account the effect of other X variables.

k) 

This graph matches the partial correlation of 0.03861117 because there is hardly any positive correlation. We see a horizontal-fit line at best instead of a line with a steep slope. This is extremely different from the scatterplot of the two variables in the matrix. This is because without all the other explanatory variable, these two variables are not much correlated, meaning that other variables has play a big part at predicting Y hat.

l)

Both R2 and R2 adjusted without observation 9 remained roughly the same as the complete dataset, since very small changes could not be counted as notable,for example, residual standard error of the original model(withoutLE2013) and the one under removal is .002589 and .002594,respectively,whose difference is very small.

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

> attach(fds)

> cor(cbind(HDI, LE2013, MEANYRSCH, EYRSCH, GNI2013, HDI2012))

HDI LE2013 MEANYRSCH EYRSCH GNI2013 HDI2012

HDI 1.0000000 0.9014085 0.8976834 0.8950789 0.7250635 0.9998629

LE2013 0.9014085 1.0000000 0.7287353 0.7556001 0.6079015 0.9016022

MEANYRSCH 0.8976834 0.7287353 1.0000000 0.7997989 0.5596667 0.8971927

EYRSCH 0.8950789 0.7556001 0.7997989 1.0000000 0.5932825 0.8945572

GNI2013 0.7250635 0.6079015 0.5596667 0.5932825 1.0000000 0.7262048

HDI2012 0.9998629 0.9016022 0.8971927 0.8945572 0.7262048 1.0000000

> varb = data.frame(HDI, LE2013, MEANYRSCH, EYRSCH, GNI2013, HDI2012)

> pairs(varb, upper.panel=NULL)

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

> summary(model1)

Call:

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

Residuals:

Min 1Q Median 3Q Max

-0.0267570 -0.0008281 0.0000985 0.0011206 0.0050706

Coefficients:

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

(Intercept) 2.812e-03 2.135e-03 1.317 0.190

LE2013 3.281e-05 6.311e-05 0.520 0.604

MEANYRSCH 1.800e-04 1.768e-04 1.018 0.310

EYRSCH 1.916e-04 1.659e-04 1.155 0.250

GNI2013 -7.769e-09 1.763e-08 -0.441 0.660

HDI2012 9.898e-01 8.006e-03 123.641 <2e-16 \*\*\*

---

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

Residual standard error: 0.002595 on 181 degrees of freedom

Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997

F-statistic: 1.346e+05 on 5 and 181 DF, p-value: < 2.2e-16

> model2=lm(HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012)

> summary(model2)

Call:

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

Residuals:

Min 1Q Median 3Q Max

-0.0268061 -0.0008438 0.0000876 0.0011594 0.0051917

Coefficients:

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

(Intercept) 3.680e-03 1.329e-03 2.769 0.0062 \*\*

MEANYRSCH 1.298e-04 1.479e-04 0.878 0.3812

EYRSCH 1.579e-04 1.524e-04 1.036 0.3017

GNI2013 -1.156e-08 1.601e-08 -0.722 0.4713

HDI2012 9.933e-01 4.551e-03 218.253 <2e-16 \*\*\*

---

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

Residual standard error: 0.002589 on 182 degrees of freedom

Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997

F-statistic: 1.689e+05 on 4 and 182 DF, p-value: < 2.2e-16

e2 = residuals(model2)

model3=lm(LE2013 ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012) e3 = residuals(model3)

>plot(e3, e2)

> cor(e3, e2)

[1] 0.03861117

> model4=lm(HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012, subset=-c(9))

> summary(model4)

Call:

lm(formula = HDI ~ MEANYRSCH + EYRSCH + GNI2013 + HDI2012, subset = -c(9))

Residuals:

Min 1Q Median 3Q Max

-0.0268016 -0.0008505 0.0000908 0.0011685 0.0051847

Coefficients:

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

(Intercept) 3.648e-03 1.332e-03 2.738 0.0068 \*\*

MEANYRSCH 1.313e-04 1.482e-04 0.886 0.3767

EYRSCH 1.592e-04 1.527e-04 1.042 0.2987

GNI2013 -1.342e-08 1.636e-08 -0.820 0.4132

HDI2012 9.933e-01 4.560e-03 217.840 <2e-16 \*\*\*

---

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

Residual standard error: 0.002594 on 181 degrees of freedom

Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997

F-statistic: 1.665e+05 on 4 and 181 DF, p-value: < 2.2e-16

>