**1a)**

NATION HDI LE2013 MEANYRSCH

Afghanistan : 1 Min. :0.3370 Min. :45.60 Min. : 1.300

Albania : 1 1st Qu.:0.5625 1st Qu.:64.95 1st Qu.: 5.450

Algeria : 1 Median :0.7170 Median :73.20 Median : 8.500

Andorra : 1 Mean :0.6856 Mean :70.58 Mean : 7.902

Angola : 1 3rd Qu.:0.8110 3rd Qu.:76.70 3rd Qu.:10.300

Antigua and Barbuda: 1 Max. :0.9440 Max. :83.60 Max. :12.900

(Other) :181

EYRSCH GNI2013 HDI2012 CHINRANK

Min. : 4.10 10,339 : 2 Min. :0.3330 Min. :-5.00000

1st Qu.:11.10 1,011 : 1 1st Qu.:0.5585 1st Qu.: 0.00000

Median :12.90 1,090 : 1 Median :0.7150 Median : 0.00000

Mean :12.81 1,129 : 1 Mean :0.6837 Mean :-0.04278

3rd Qu.:15.10 1,142 : 1 3rd Qu.:0.8100 3rd Qu.: 0.00000

Max. :19.90 1,147 : 1 Max. :0.9430 Max. : 4.00000

(Other):180

DL

high :53

low :43

medium :42

very high:49

MEANYRSCH EYRSCH

Min. : 1.300 Min. : 4.10

1st Qu.: 5.450 1st Qu.:11.10

Median : 8.500 Median :12.90

Mean : 7.902 Mean :12.81

3rd Qu.:10.300 3rd Qu.:15.10

Max. :12.900 Max. :19.90

In 2012,the MEANYRSCH is skewed left,since the median is larger than the mean,

Similarly, EYRSCH is also skewed left, yet almost appears normal since the data is almost equal.

**1b)**



The data is left skewed.



The data appears almost normal.





The data almost converges at the midrange, yet diverges at both sides.



The EYRSCH data seems converge quite well

**1c)**



The original histogram of EYRSCH appears more normally distributed than the

one in the square root, which appears left skewed.

**1d)**

****

While both graphs are not normal, the natural log transformation of MEANYRSCH (bottom graph) is more skewed than the boxplot of the variable itself.

**1e)**





The histogram of HDI is skewed left

The data in the qqnorm is divergent especially on the ends and in the center.

**1f)**



For high DL, HDL is normally distributed,since its mean seems equal its median

For Low DL, HDL is left-skewed,since its mean is smaller than its median

For Medium DL,HDL is almost normal(a bit right skewed),since its mean is a tiny bit larger than its median

For very high DL,HDL is left-skewed, since its mean is smaller than its median

To sum up, by categories ,they are not all normally distributed.

**SUBSET**

**1g)**





The MEANYRSCH is skewed left while the EYRSCH appears approximately normal



The two qq plots for MEANYRSCH and EYRSCH in the subset are comparable to those in the full data set, with divergences appearing on both ends of the graphs.



**2a)**

MEANYRSCH EYRSCH HDI

MEANYRSCH 1.0000000 0.7997989 0.8976834

EYRSCH 0.7997989 1.0000000 0.8950789

HDI 0.8976834 0.8950789 1.0000000

Cor (MEANYRSCH,EYRSCH) is .7997989

Cor (MEANYRSCH,HDI) is .8976834

Cor (EYRSCH,HDI) is .8950789

The strongest correlation value is between MEANYRSCH and HDI.

The weakest correlation is between MEANYRSCH and EYRSCH.

**2b)**



Yes, this graph seems to coincide with the correlation of 0.7997989



This seems to correspond to the correlation of .8976834 because it is significantly better than the previous pair.



By comparison with plot(MEANYRSCH,HDI), it is not as good while the correlation value is almost the same

**2c)**

Call:

lm(formula = HDI ~ EYRSCH)

Coefficients:

(Intercept) EYRSCH

0.05618 0.04915

Call:

lm(formula = HDI ~ EYRSCH)

Residuals:

Min 1Q Median 3Q Max

-0.198922 -0.044069 0.009931 0.043908 0.198798

Coefficients:

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

(Intercept) 0.05618 0.02361 2.379 0.0184 \*

EYRSCH 0.04915 0.00180 27.302 <2e-16 \*\*\*

---

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

Residual standard error: 0.06978 on 185 degrees of freedom

Multiple R-squared: 0.8012, Adjusted R-squared: 0.8001

F-statistic: 745.4 on 1 and 185 DF, p-value: < 2.2e-16

R-sqr=.8012,80.12% of the variation in HDI is being explained by EYRSCH ,this is pretty high and good enough.

Via t-stat, 0.0018 is the standard error of .04915

df=187-(1+1)=185

qt(.975,185)

[1] 1.97287

Since our t-stat + 27.302 is significantly larger than 1.97287, Beta is not equal to 0, this leads us to believe that EYRSCH is a significant predictor of HDI.

HDI=.05618+.04915EYRSCH

Call:

lm(formula = EYRSCH ~ HDI)

Coefficients:

(Intercept) HDI

1.631 16.302

> model2=lm(EYRSCH~HDI)

> summary(model2)

Call:

lm(formula = EYRSCH ~ HDI)

Residuals:

Min 1Q Median 3Q Max

-3.7415 -0.7308 -0.1609 0.7383 3.3208

Coefficients:

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

(Intercept) 1.6307 0.4198 3.885 0.000143 \*\*\*

HDI 16.3015 0.5971 27.302 < 2e-16 \*\*\*

---

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

Residual standard error: 1.271 on 185 degrees of freedom

Multiple R-squared: 0.8012, Adjusted R-squared: 0.8001

F-statistic: 745.4 on 1 and 185 DF, p-value: < 2.2e-16

R-sqr=.8012,80.12% of the variation in EYRSCH is being explained by HDI, which is the same as where Y is HDI, when X is EYRSCH

Next, the residual std Error in Y by EYRSCH is higher at 1.271, df is still in 185

In terms of p-value, both Y by HDI and Y by EYRSCH are the same.

Thus,Y by HDI works better

Ho:beta = 0; H1: beta!= 0;

For EYRSCH=1.6307+16.3015HDI, df is 185

qt(.975,185)

[1] 1.97287

Since our t-stat + 27.302 is significantly larger than 1.97287,

Ho:Beta is not equal 0,this leads us to believe that HDI is a significant predictor of EYRSCH

It is the same as the previous Y by HDI case in the p-value approach

p-value < 2.2e-16 which is smaller than 0.000143

Thus, HDI as a predictor of EYRSCH does work fine.

**2g)**

**> newdata=data.frame(HDI=5)**

**> predict(model2, newdata, interval="confidence", level=.99)**

0.5 % 99.5 %

(Intercept) 0.5381014 2.723232

HDI 14.7475343 17.855483

The confidence Interval for X=HDI's slope is (14.74753,17.85548)

###CI= if the EYRSCH has 5 as its value, we have 99% confidence that the predicted HDI would have an average of score at (14.74753, 17.85548) .

##PI=have an individual score at (14.74753, 17.85548) .

**2h)**

1.6307+16.3015\*5

[1] 83.1382

Y hat is 83.1382, the residual value = Y - 83.1382

**2i)**

> newdata=data.frame(HDI=5)

> predict(model2, newdata, interval="prediction", level=.95)

fit lwr upr

1 83.13821 77.46831 88.80811

The prediction Interval for X=HDI's slope is (77.46831, 88.80811)

The interval is (77.24855,89.02795).

##PI=If a country has 5 as its value in HDI, we have 95% confidence that the predicted EYRSCH for that country would have a score at (77.46831, 88.80811) .

**2j)**

Ho: B = 2

H1: B ≠ 2

(16.3015 – 2)/0.5971

[1] 23.9516

T-score(WRONG!)

Since T-score 23.9516 is smaller than critical t-value of HDI: 27.302,we reject Ho:B =2; 2 is significantly different from the original slope of 16.3015.