Lab03

Henry Hongbin Liu

The written answer for each section is in the R script comment.

2.1 Linear model summary:

Call:

lm(formula = EPI ~ ECO + BDH + MKP + MHP + CDF, data = epi\_data)

Residuals:

Min 1Q Median 3Q Max

-10.6969 -3.3802 0.0762 3.3101 12.2877

Coefficients:

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

(Intercept) 0.34014 2.32418 0.146 0.88388

ECO 1.26193 0.08018 15.738 < 2e-16 \*\*\*

BDH -0.41141 0.06124 -6.718 5.69e-10 \*\*\*

MKP -0.01589 0.01748 -0.909 0.36516

MHP 0.01974 0.02570 0.768 0.44381

CDF 0.05303 0.01606 3.303 0.00124 \*\*

---

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

Residual standard error: 4.788 on 126 degrees of freedom

(48 observations deleted due to missingness)

Multiple R-squared: 0.8351, Adjusted R-squared: 0.8286

F-statistic: 127.6 on 5 and 126 DF, p-value: < 2.2e-16

2.2 Linear Model Summary:

Call:

lm(formula = EPI ~ ECO + BDH + MKP + MHP + CDF, data = eastern\_europe)

Residuals:

Min 1Q Median 3Q Max

-1.6561 -0.8051 -0.2996 0.6681 1.9664

Coefficients:

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

(Intercept) -11.68009 4.80114 -2.433 0.045236 \*

ECO 1.65475 0.21750 7.608 0.000125 \*\*\*

BDH -0.74204 0.17039 -4.355 0.003335 \*\*

MKP -0.05876 0.01887 -3.114 0.016982 \*

MHP 0.03958 0.01614 2.453 0.043938 \*

CDF 0.21533 0.02242 9.605 2.79e-05 \*\*\*

---

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

Residual standard error: 1.405 on 7 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.9873, Adjusted R-squared: 0.9781

F-statistic: 108.4 on 5 and 7 DF, p-value: 1.789e-06

3.1 Matrix:

Confusion Matrix and Statistics

Reference

Prediction Eastern Europe Latin America & Caribbean Southern Asia

Eastern Europe 1 1 0

Latin America & Caribbean 2 6 1

Southern Asia 0 1 0

Overall Statistics

Accuracy : 0.5833

95% CI : (0.2767, 0.8483)

No Information Rate : 0.6667

P-Value [Acc > NIR] : 0.8223

Kappa : 0.0769

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: Eastern Europe Class: Latin America & Caribbean Class: Southern Asia

Sensitivity 0.33333 0.7500 0.00000

Specificity 0.88889 0.2500 0.90909

Pos Pred Value 0.50000 0.6667 0.00000

Neg Pred Value 0.80000 0.3333 0.90909

Prevalence 0.25000 0.6667 0.08333

Detection Rate 0.08333 0.5000 0.00000

Detection Prevalence 0.16667 0.7500 0.08333

Balanced Accuracy 0.61111 0.5000 0.45455

"Accuracy for Model 1: 0.583333333333333”

3.2 Answers:

Confusion Matrix and Statistics

Reference

Prediction Greater Middle East Sub-Saharan Africa

Greater Middle East 2 2

Sub-Saharan Africa 2 4

Accuracy : 0.6

95% CI : (0.2624, 0.8784)

No Information Rate : 0.6

P-Value [Acc > NIR] : 0.6331

Kappa : 0.1667

Mcnemar's Test P-Value : 1.0000

Sensitivity : 0.5000

Specificity : 0.6667

Pos Pred Value : 0.5000

Neg Pred Value : 0.6667

Prevalence : 0.4000

Detection Rate : 0.2000

Detection Prevalence : 0.4000

Balanced Accuracy : 0.5833

'Positive' Class : Greater Middle East

"Accuracy for Model 2: 0.6”

4.1 outcome:

> print(paste("WCSS for Subset 1 (Eastern Europe, Latin America & Caribbean, Southern Asia):", wcss\_subset1))

[1] "WCSS for Subset 1 (Eastern Europe, Latin America & Caribbean, Southern Asia): 47553.4160869565"

> print(paste("WCSS for Subset 2 (Western Europe, Sub-Saharan Africa, Greater Middle East):", wcss\_subset2))

[1] "WCSS for Subset 2 (Western Europe, Sub-Saharan Africa, Greater Middle East): 47924.7299650349"