Natan Alper 2/19/2020

Business Intelligence & Consumer Insights- Professor Kovtun

HW #3

**1 (a)**

> summary(lm(y~.,data=data2[,-1]))$adj

[1] 0.9999479

> summary(lm(y~.,data=data2[,-2]))$adj

[1] 0.9999971

> summary(lm(y~.,data=data2[,-3]))$adj

[1] 0.9999991

> summary(lm(y~.,data=data2[,-4]))$adj

[1] 0.9394085

> summary(lm(y~.,data=data2[,-5]))$adj # removing 5

[1] 0.9999991

> summary(lm(y~.,data=data2[,c(-1,-5)]))$adj

[1] 0.999946

> summary(lm(y~.,data=data2[,c(-2,-5)]))$adj

[1] 0.9996597

> summary(lm(y~.,data=data2[,c(-3,-5)]))$adj # removing 3 and 5

[1] 0.9999991

> summary(lm(y~.,data=data2[,c(-4,-5)]))$adj

[1] 0.9351295

> summary(lm(y~.,data=data2[,c(-1,-3,-5)]))$adj

[1] 0.9999473

> summary(lm(y~.,data=data2[,c(-2,-3,-5)]))$adj

[1] 0.9996288

> summary(lm(y~.,data=data2[,c(-4,-3,-5)]))$adj

[1] 0.9366124

> # removing 3 and 5 yield best results

Final Estimated Model: Y = 5.587 - 5.940\*x1 + 3.981\*x2 + 1.999\*x1sq

**1 (b)**

> step(lm(y~.,data=data2))

Start: AIC=144.22

y ~ x1 + x2 + x3 + x1sq + x2sq

Df Sum of Sq RSS AIC

- x2sq 1 41 1132 143.72

<none> 1091 144.22

- x3 1 72 1163 144.78

- x2 1 2662 3753 191.65

- x1 1 66128 67218 307.07

- x1sq 1 78150017 78151107 589.41

Step: AIC=143.72

y ~ x1 + x2 + x3 + x1sq

Df Sum of Sq RSS AIC

<none> 1132 143.72

- x3 1 64 1196 143.91

- x1 1 70562 71694 307.65

- x2 1 450325 451457 381.25

- x1sq 1 86059590 86060722 591.27

Call:

lm(formula = y ~ x1 + x2 + x3 + x1sq, data = data2)

Coefficients:

(Intercept) x1 x2 x3 x1sq

-2.92238 -5.93824 3.96569 0.04524 1.99928

**1 (c)**

Adjusted R^2 closely estimated the coefficients and alpha

AIC closely estimated the coefficients but alpha was different

**4 (a)**

Var1 = -0.08356 - 0.04044\*var2 + 0.03321\*var3 + 0.02263\*var2^2

**4 (b)**

preds <- -0.08356 - 0.04044\*var2 + 0.03321\*var3 + 0.02263\*var2^2

**4 (c)**

1.009078

**4 (d)**

0.7992628

**4 (e)**

> step(lm(var4~.,data=NewData))

Start: AIC=15.25

var4 ~ var1 + var2 + var3 + var5 + var6

Step: AIC=15.25

var4 ~ var1 + var2 + var3 + var5

Df Sum of Sq RSS AIC

- var1 1 0.12207 1005.4 13.373

- var3 1 0.12502 1005.4 13.376

- var5 1 0.81243 1006.1 14.059

- var2 1 1.76970 1007.0 15.010

<none> 1005.3 15.252

Step: AIC=13.37

var4 ~ var2 + var3 + var5

Df Sum of Sq RSS AIC

- var3 1 0.11736 1005.5 11.490

- var5 1 0.79667 1006.2 12.165

- var2 1 1.74046 1007.1 13.103

<none> 1005.4 13.373

Step: AIC=11.49

var4 ~ var2 + var5

Df Sum of Sq RSS AIC

- var5 1 0.82194 1006.3 10.307

- var2 1 1.78737 1007.3 11.266

<none> 1005.5 11.490

Step: AIC=10.31

var4 ~ var2

Df Sum of Sq RSS AIC

- var2 1 1.8055 1008.1 10.099

<none> 1006.3 10.307

Step: AIC=10.1

var4 ~ 1

Call:

lm(formula = var4 ~ 1, data = NewData)

Coefficients:

(Intercept)

0.0117