

Flipped Assignment 15

Group 5

2022/4/7

Input Data and Function Definition

```
setwd('G:/OneDrive - Texas Tech University/IE 5344 Statistical Data Analysis/Flipped Assignment 15')
data <- read.csv('data-table-B9.csv', header = TRUE)
head(data)
```

```
##      x1 x2  x3   x4   y
## 1 2.14 10 0.34 1.000 28.9
## 2 4.14 10 0.34 1.000 31.0
## 3 8.15 10 0.34 1.000 26.4
## 4 2.14 10 0.34 0.246 27.2
## 5 4.14 10 0.34 0.379 26.1
## 6 8.15 10 0.34 0.474 23.2
```

Part a.

We standardize data because interaction terms are included.

```
data_1 <- scale(data[,1:4], center=TRUE, scale=TRUE)
data_1 <- cbind(data_1, data$y)
data_1 <- as.data.frame(data_1)
colnames(data_1) <- c("x1", "x2", "x3", "x4", "y")
head(data_1)
```

```
##      x1      x2      x3      x4   y
## 1 -1.3287681 -0.3271853 -0.1335207  1.7075209 28.9
## 2 -0.3237834 -0.3271853 -0.1335207  1.7075209 31.0
## 3  1.6912110 -0.3271853 -0.1335207  1.7075209 26.4
## 4 -1.3287681 -0.3271853 -0.1335207 -1.5341614 27.2
## 5 -0.3237834 -0.3271853 -0.1335207 -0.9623528 26.1
## 6  1.6912110 -0.3271853 -0.1335207 -0.5539180 23.2
```

Part b.

```
fit2 <- lm(y ~ 1, data_1)
step(fit2, scope~x1 + x2 + x3 + x4 + x1:x2 + x1:x3+ x1:x4 + x2:x3 + x2:x4 + x3:x4,direction="forward")

## Start:  AIC=269.58
## y ~ 1
##
##      Df Sum of Sq  RSS   AIC
## + x2    1   2839.01 1803.5 212.96
## + x3    1    182.94 4459.6 269.09
```

```
## <none>                4642.5 269.58
## + x4      1      132.59 4509.9 269.79
## + x1      1       9.60 4632.9 271.46
##
## Step: AIC=212.96
## y ~ x2
##
##      Df Sum of Sq  RSS    AIC
## + x3    1  252.408 1551.1 205.61
## + x4    1   75.841 1727.7 212.30
## <none>                1803.5 212.96
## + x1    1   10.367 1793.1 214.60
##
## Step: AIC=205.61
## y ~ x2 + x3
##
##      Df Sum of Sq  RSS    AIC
## + x4    1   68.644 1482.4 204.81
## <none>                1551.1 205.61
## + x1    1   16.911 1534.2 206.93
##
## Step: AIC=204.81
## y ~ x2 + x3 + x4
##
##      Df Sum of Sq  RSS    AIC
## + x2:x4  1   245.680 1236.8 195.57
## + x1    1    49.657 1432.8 204.69
## <none>                1482.4 204.81
## + x3:x4  1    11.880 1470.6 206.31
##
## Step: AIC=195.57
## y ~ x2 + x3 + x4 + x2:x4
##
##      Df Sum of Sq  RSS    AIC
## <none>                1236.8 195.57
## + x1    1    11.262 1225.5 197.01
## + x3:x4  1    10.564 1226.2 197.04
##
## Call:
## lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
##
## Coefficients:
## (Intercept)          x2          x3          x4        x2:x4
##    23.6462     7.3728     2.0010     0.6382    -2.6544
```

The selected model is:

```
fit_b <- lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
summary(fit_b)
```

```
##
## Call:
## lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
##
## Residuals:
```

```
##      Min      1Q Median      3Q      Max
## -9.959 -3.358 -1.131  3.040 11.646
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.6462     0.5930  39.876 < 2e-16 ***
## x2           7.3728     0.6174  11.942 < 2e-16 ***
## x3           2.0010     0.5972   3.351 0.00144 **
## x4           0.6382     0.6106   1.045 0.30039
## x2:x4        -2.6544     0.7888  -3.365 0.00137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.658 on 57 degrees of freedom
## Multiple R-squared:  0.7336, Adjusted R-squared:  0.7149
## F-statistic: 39.24 on 4 and 57 DF,  p-value: 9.297e-16
```

Part c.

```
fit2 <- lm(y ~ ., data_1)
step(fit2, scope~x1 + x2 + x3 + x4 + x1:x2 + x1:x3+ x1:x4 + x2:x3 + x2:x4 + x3:x4,direction="both")

## Start:  AIC=204.7
## y ~ x1 + x2 + x3 + x4
##
##           Df Sum of Sq  RSS   AIC
## + x2:x4    1    207.29 1225.5 197.01
## <none>                        1432.8 204.69
## - x1       1     49.66 1482.4 204.81
## + x1:x2    1     19.57 1413.2 205.84
## + x1:x4    1     14.91 1417.9 206.05
## + x3:x4    1     10.98 1421.8 206.22
## - x4       1    101.39 1534.2 206.93
## - x3       1    255.72 1688.5 212.88
## - x2       1   2841.93 4274.7 270.47
##
## Step:  AIC=197.01
## y ~ x1 + x2 + x3 + x4 + x2:x4
##
##           Df Sum of Sq  RSS   AIC
## - x1       1     11.262 1236.8 195.57
## <none>                        1225.5 197.01
## + x1:x2    1     32.307 1193.2 197.35
## + x1:x4    1     13.435 1212.1 198.32
## + x3:x4    1     10.212 1215.3 198.49
## - x2:x4    1    207.286 1432.8 204.69
## - x3       1    248.430 1473.9 206.45
##
## Step:  AIC=195.57
## y ~ x2 + x3 + x4 + x2:x4
##
##           Df Sum of Sq  RSS   AIC
## <none>                        1236.8 195.57
## + x1       1     11.262 1225.5 197.01
```

```
## + x3:x4 1 10.564 1226.2 197.04
## - x3 1 243.599 1480.4 204.72
## - x2:x4 1 245.680 1482.4 204.81

##
## Call:
## lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
##
## Coefficients:
## (Intercept) x2 x3 x4 x2:x4
## 23.6462 7.3728 2.0010 0.6382 -2.6544
```

The selected model is:

```
fit_c <- lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
summary(fit_c)
```

```
##
## Call:
## lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
##
## Residuals:
## Min 1Q Median 3Q Max
## -9.959 -3.358 -1.131 3.040 11.646
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.6462 0.5930 39.876 < 2e-16 ***
## x2 7.3728 0.6174 11.942 < 2e-16 ***
## x3 2.0010 0.5972 3.351 0.00144 **
## x4 0.6382 0.6106 1.045 0.30039
## x2:x4 -2.6544 0.7888 -3.365 0.00137 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.658 on 57 degrees of freedom
## Multiple R-squared: 0.7336, Adjusted R-squared: 0.7149
## F-statistic: 39.24 on 4 and 57 DF, p-value: 9.297e-16
```

Part d.

```
fit3 <- lm(y ~ . + .*, data_1)
step(fit3, direction="backward")

## Start: AIC=199.73
## y ~ x1 + x2 + x3 + x4 + (x1 + x2 + x3 + x4) * (x1 + x2 + x3 +
## x4)
##
##
## Step: AIC=199.73
## y ~ x1 + x2 + x3 + x4 + x1:x2 + x1:x3 + x1:x4 + x2:x4 + x3:x4
##
##
## Step: AIC=199.73
## y ~ x1 + x2 + x3 + x4 + x1:x2 + x1:x4 + x2:x4 + x3:x4
##
```

```

##           Df Sum of Sq    RSS    AIC
## - x3:x4  1      10.942 1173.4 198.31
## - x1:x4  1      20.682 1183.1 198.82
## <none>                1162.4 199.73
## - x1:x2  1      38.737 1201.2 199.76
## - x2:x4  1     227.751 1390.2 208.82
##
## Step:  AIC=198.31
## y ~ x1 + x2 + x3 + x4 + x1:x2 + x1:x4 + x2:x4
##
##           Df Sum of Sq    RSS    AIC
## - x1:x4  1      19.837 1193.2 197.35
## <none>                1173.4 198.31
## - x1:x2  1      38.709 1212.1 198.32
## - x2:x4  1     228.394 1401.8 207.34
## - x3     1     249.320 1422.7 208.26
##
## Step:  AIC=197.35
## y ~ x1 + x2 + x3 + x4 + x1:x2 + x2:x4
##
##           Df Sum of Sq    RSS    AIC
## - x1:x2  1      32.307 1225.5 197.01
## <none>                1193.2 197.35
## - x2:x4  1     220.026 1413.2 205.84
## - x3     1     252.209 1445.4 207.24
##
## Step:  AIC=197.01
## y ~ x1 + x2 + x3 + x4 + x2:x4
##
##           Df Sum of Sq    RSS    AIC
## - x1     1      11.262 1236.8 195.57
## <none>                1225.5 197.01
## - x2:x4  1     207.286 1432.8 204.69
## - x3     1     248.430 1473.9 206.45
##
## Step:  AIC=195.57
## y ~ x2 + x3 + x4 + x2:x4
##
##           Df Sum of Sq    RSS    AIC
## <none>                1236.8 195.57
## - x3     1     243.60 1480.4 204.72
## - x2:x4  1     245.68 1482.4 204.81
##
## Call:
## lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
##
## Coefficients:
## (Intercept)          x2          x3          x4        x2:x4
##      23.6462      7.3728      2.0010      0.6382     -2.6544

```

The selected model is:

```

fit_d <- lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
summary(fit_d)

```

```
##
## Call:
## lm(formula = y ~ x2 + x3 + x4 + x2:x4, data = data_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.959 -3.358 -1.131  3.040 11.646
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.6462     0.5930  39.876 < 2e-16 ***
## x2           7.3728     0.6174  11.942 < 2e-16 ***
## x3           2.0010     0.5972   3.351 0.00144 **
## x4           0.6382     0.6106   1.045 0.30039
## x2:x4        -2.6544     0.7888  -3.365 0.00137 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.658 on 57 degrees of freedom
## Multiple R-squared:  0.7336, Adjusted R-squared:  0.7149
## F-statistic: 39.24 on 4 and 57 DF,  p-value: 9.297e-16
```

Part. e

We find the same model in parts b, c, and d. We think it is because the selecting criteria are the same, based on AIC, and the scopes are also the same.