QUIZ

Fabian Abrego

4/27/2020

#Step Wise Model Selection

##Question

For this question, we will be using the `teengam` data set included in the `Faraway` library.

(a) Set `gamble` as your response variable and using stepwise model selection, derive the best model. Do not use the `step` function for this. Use the `extarctAIC` function for AIC values instead of the `AIC` function.

```
gamble = faraway::teengamb
modelstart = lm(gamble ~ 1, data = gamble)


#First step
extractAIC(modelstart) #325.3354 Starting AIC
```

```
## [1]    1.0000 325.3354
```

```
#Add a variable
extractAIC(lm(gamble ~ sex, data = gamble)) #318.7866 -  Improvement
```

```
## [1]    2.0000 318.7866
```

```
extractAIC(lm(gamble ~ status, data = gamble)) #327.2157 - No Improvement
```

```
## [1]    2.0000 327.2157
```

```
extractAIC(lm(gamble ~ income, data = gamble)) #304.3356 - Greatest Improvement - Add to model
```

```
## [1]    2.0000 304.3356
```

```
extractAIC(lm(gamble ~ verbal, data = gamble)) #325.0025 - Improvement
```

```
## [1]    2.0000 325.0025
```

```
#Second Step
#Current AIC = 304.3356

#Add a variable
extractAIC(lm(gamble ~ income + sex, data = gamble)) #296.6268 - Add to model
```

```
## [1]    3.0000 296.6268
```

```
extractAIC(lm(gamble ~ income + status, data = gamble)) #305.112 - No Improvement
```

```
## [1]    3.000 305.112
```

```
extractAIC(lm(gamble ~ income + verbal, data = gamble)) #305.3536 - No Improvement
```

```
## [1]    3.0000 305.3536
```

```
#Remove a variable
extractAIC(modelstart) #325.3354 - No Improvement
```

```
## [1]    1.0000 325.3354
```

```
#Third step
#Current AIC = 296.6268

#Add a variable
extractAIC(lm(gamble ~ income + sex + status, data = gamble)) #298.2086 - No Improvement
```

```
## [1]    4.0000 298.2086
```

```
extractAIC(lm(gamble ~ income + sex + verbal, data = gamble)) #296.2145 - Add to model
```

```
## [1]    4.0000 296.2145
```

```
#Remove a variable
extractAIC(lm(gamble ~ income, data = gamble)) #304.3356 - No Improvement
```

```
## [1]    2.0000 304.3356
```

```
extractAIC(lm(gamble ~ sex, data = gamble)) #318.7866 - No Improvement
```

```
## [1]    2.0000 318.7866
```

```
#Fourth Step
#Current AIC = 296.2145

#Add a variable
extractAIC(lm(gamble ~ income + sex + verbal + status, data = gamble)) #298.1758 - No Improvement
```

```
## [1]    5.0000 298.1758
```

```
#Remove a variable
extractAIC(lm(gamble ~ income + sex, data = gamble)) #296.6268 - No Improvement
```

```
## [1]    3.0000 296.6268
```

```
extractAIC(lm(gamble ~ income + verbal, data = gamble)) #305.3536 - No Improvement
```

```
## [1]    3.0000 305.3536
```

```
extractAIC(lm(gamble ~ sex + verbal, data = gamble)) #316.6443 - No Improvement
```

```
## [1]    3.0000 316.6443
```

```
Final_Model = lm(gamble ~ income + sex + verbal, data = gamble)
```

(b) Use the `step` function to derive the best model from stepwise selection. Are they the same?

```
#Using Step Function
model_step = step(modelstart, scope = gamble ~ sex + status + income + verbal, direction = "both")
```

```
## Start:  AIC=325.34
## gamble ~ 1
##
##           Df Sum of Sq   RSS    AIC
## + income   1   17680.9 28009 304.34
## + sex      1    7598.4 38091 318.79
## + verbal   1    2212.5 43477 325.00
```

```
## <none>                      45689 325.34
## + status   1      116.2 45573 327.22
##
## Step:  AIC=304.34
## gamble ~ income
##
##           Df Sum of Sq   RSS    AIC
## + sex      1    5227.3 22781 296.63
## <none>                 28009 304.34
## + status   1     719.8 27289 305.11
## + verbal   1     579.1 27429 305.35
## - income   1   17680.9 45689 325.34
##
## Step:  AIC=296.63
## gamble ~ income + sex
##
##           Df Sum of Sq   RSS    AIC
## + verbal   1    1139.8 21642 296.21
## <none>                 22781 296.63
## + status   1     201.8 22580 298.21
## - sex      1    5227.3 28009 304.34
## - income   1   15309.8 38091 318.79
##
## Step:  AIC=296.21
## gamble ~ income + sex + verbal
##
##           Df Sum of Sq   RSS    AIC
## <none>                 21642 296.21
## - verbal   1    1139.8 22781 296.63
## + status   1      17.8 21624 298.18
## - sex      1    5787.9 27429 305.35
## - income   1   13236.1 34878 316.64
```
```r
#They are equal
summary(model_step)
```
```
##
## Call:
## lm(formula = gamble ~ income + sex + verbal, data = gamble)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.639 -11.765  -1.594   9.305  93.867
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.1390    14.7686   1.634   0.1095
## income        4.8981     0.9551   5.128 6.64e-06 ***
## sex         -22.9602     6.7706  -3.391   0.0015 **
## verbal       -2.7468     1.8253  -1.505   0.1397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.43 on 43 degrees of freedom
## Multiple R-squared:  0.5263, Adjusted R-squared:  0.4933
```

3

```
## F-statistic: 15.93 on 3 and 43 DF,  p-value: 4.148e-07
```

```
summary(Final_Model)
```

```
##
## Call:
## lm(formula = gamble ~ income + sex + verbal, data = gamble)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -50.639 -11.765  -1.594   9.305  93.867
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.1390    14.7686   1.634   0.1095
## income        4.8981     0.9551   5.128 6.64e-06 ***
## sex         -22.9602     6.7706  -3.391   0.0015 **
## verbal       -2.7468     1.8253  -1.505   0.1397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.43 on 43 degrees of freedom
## Multiple R-squared:  0.5263, Adjusted R-squared:  0.4933
## F-statistic: 15.93 on 3 and 43 DF,  p-value: 4.148e-07
```