NCSU ST 503 Discussion 8

Probem 10.6 Faraway, Julian J. Linear Models with R CRC Press.

Bruce Campbell

10.6 Model Selection with the hipcenter data.

Use the seatpos data with hipcenter as the response.

(a) Fit a model with all eight predictors. Comment on the effect of leg length on the response.

```
##
## Call:
## lm(formula = hipcenter ~ ., data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.827 -22.833  -3.678  25.017  62.337
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 436.43213  166.57162   2.620   0.0138 *
## Age           0.77572    0.57033   1.360   0.1843
## Weight        0.02631    0.33097   0.080   0.9372
## HtShoes      -2.69241    9.75304  -0.276   0.7845
## Ht            0.60134   10.12987   0.059   0.9531
## Seated        0.53375    3.76189   0.142   0.8882
## Arm          -1.32807    3.90020  -0.341   0.7359
## Thigh        -1.14312    2.66002  -0.430   0.6706
## Leg          -6.43905    4.71386  -1.366   0.1824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.72 on 29 degrees of freedom
## Multiple R-squared:  0.6866, Adjusted R-squared:  0.6001
## F-statistic:  7.94 on 8 and 29 DF,  p-value: 1.306e-05
```

We note that leg length is significant at a level of $\alpha = 0.018$ and it has a negative association with the response.

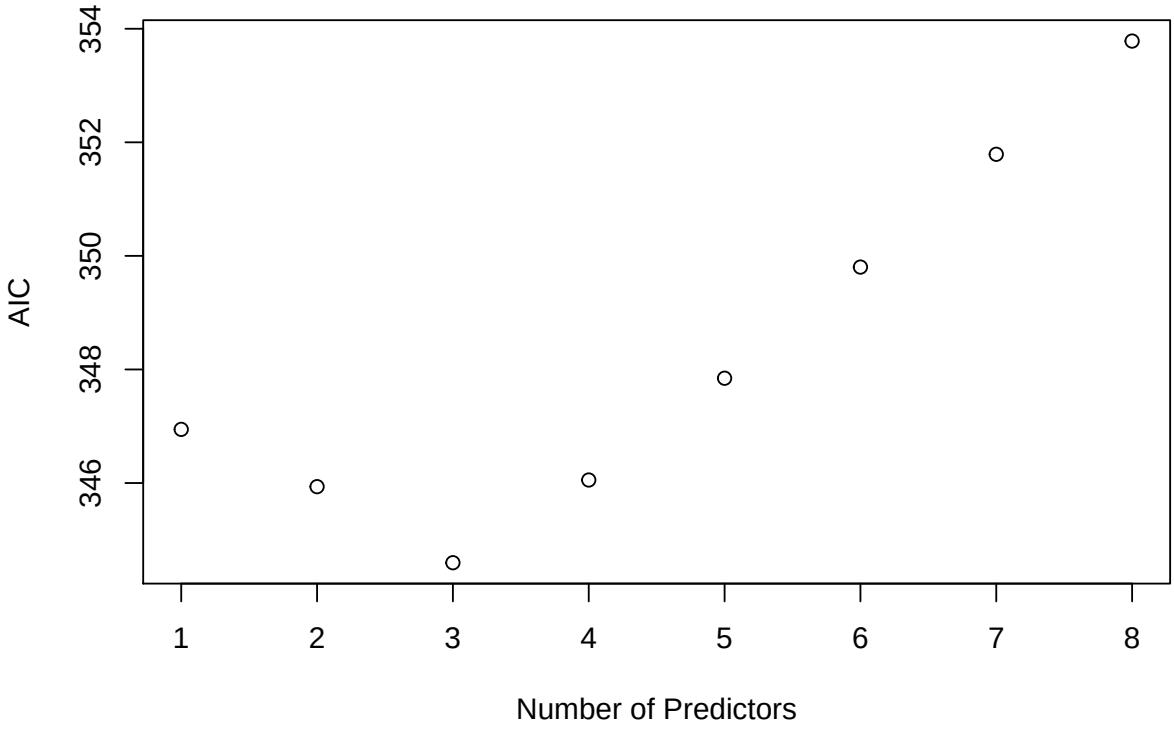
(b) Compute a 95% prediction interval for the mean value of the predictors.

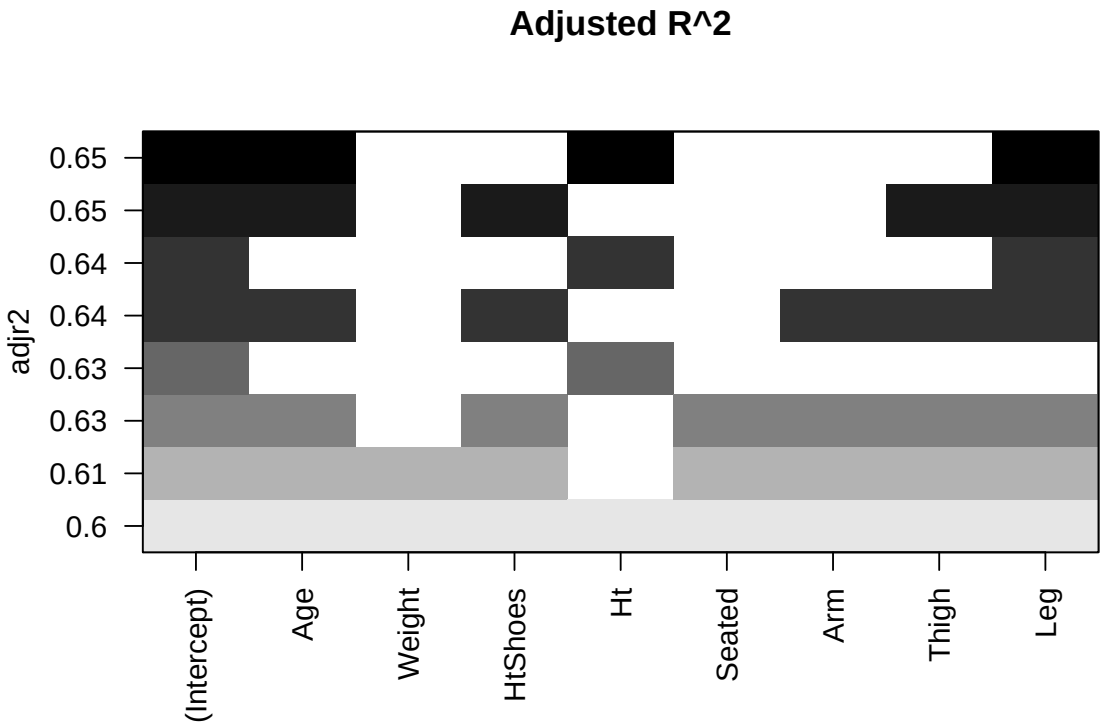
```
##         fit     lwr       upr
## 1 -164.8849 -243.04 -86.72972
```

pi.width
156.3

(c) Use AIC to select a model. Now interpret the effect of leg length and compute the prediction interval. Compare the conclusions from the two models.

```
##   (Intercept)   Age Weight HtShoes    Ht Seated   Arm Thigh   Leg
## 1        TRUE FALSE  FALSE   FALSE  TRUE  FALSE FALSE FALSE FALSE
## 2        TRUE FALSE  FALSE   FALSE  TRUE  FALSE FALSE FALSE  TRUE
## 3        TRUE  TRUE  FALSE   FALSE  TRUE  FALSE FALSE FALSE  TRUE
## 4        TRUE  TRUE  FALSE    TRUE FALSE  FALSE FALSE  TRUE  TRUE
## 5        TRUE  TRUE  FALSE    TRUE FALSE  FALSE  TRUE  TRUE  TRUE
## 6        TRUE  TRUE  FALSE    TRUE FALSE   TRUE  TRUE  TRUE  TRUE
## 7        TRUE  TRUE   TRUE    TRUE FALSE   TRUE  TRUE  TRUE  TRUE
## 8        TRUE  TRUE   TRUE    TRUE  TRUE   TRUE  TRUE  TRUE  TRUE
```


AIC
346
348
350
352
354
1
2
3
4
5
6
7
8
Number of Predictors



We see that $hipcenter \sim +age + ht + Leg$ is the model with the lowest AIC. We also plot the Adjusted R^2 of the models.

```
##
## Call:
## lm(formula = hipcenter ~ Age + Ht + Leg, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -79.715 -22.758  -4.102  21.394  60.576
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 452.1976   100.9482   4.480 8.04e-05 ***
## Age           0.5807     0.3790   1.532   0.1347
## Ht           -2.3254     1.2545  -1.854   0.0725 .
## Leg          -6.7390     4.1050  -1.642   0.1099
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.12 on 34 degrees of freedom
```

```
## Multiple R-squared:  0.6814, Adjusted R-squared:  0.6533
## F-statistic: 24.24 on 3 and 34 DF,  p-value: 1.426e-08
```

Leg now has a p-value of 0.1099

```
##         fit      lwr       upr
## 1 -164.8849 -237.192 -92.57771
```

pi.width
144.6

As expected our prediction interval has decresed in width. Ht is now significant at 0.07 which is a dramatic change. We presume this is due to linear association among the predictors.