6.3 Model Selection: Information Criteria-based procedures

The idea behind these methods is to score each model according to an information criterion and then use a searching algorithm to find the optimal model. In this case, model selection criteria/scores often takes the following form:

$$\text{Training Error} + \text{Complexity Penalty}$$

In the context of linear regression models, *complexity of a model increases with the number of predictor variables* (i.e., p_γ).

So, why don't we just use R^2 or RSS?
The main reason is that R^2 always increases when we introduce variables in the model, while RSS always reduces. Therefore, they do not penalize for introducing unnecessary variables in the model.

6.3.1 Information-Based Criteria

Akaike Information Criterion & Bayesian Information Criterion

These are the famous AIC/BIC criterian that are defined as

$$AIC = -2 \times loglik_\gamma + 2p_\gamma$$
$$BIC = -2 \times loglik_\gamma + \log(n)p_\gamma$$

where p_γ is the number of predictors included in model γ.

For the linear regression model, the first term computes:

$$-2 \times loglik_\gamma = n\log\frac{RSS_\gamma}{n},$$

which means that

$$AIC = n\log\frac{RSS_\gamma}{n} + 2p_\gamma$$
$$BIC = n\log\frac{RSS_\gamma}{n} + \log(n)p_\gamma$$

The **lower** the AIC/BIC the better. Note that when n is large, *adding an additional predictor costs a lot more in BIC than AIC*. So, AIC tends to pick a bigger model than BIC.

Adjusted-R^2 for model γ

$$R^2_a = 1 - \frac{RSS/(n-p_\gamma-1)}{TSS/(n-1)}$$
$$= 1 - (1-R^2)\left(\frac{n-1}{n-p_\gamma-1}\right)$$
$$R^2_a = 1 - \frac{\hat{\sigma}^2_\gamma}{\hat{\sigma}^2_0}$$

The **higher** the R^2_a the better.

Mallow's C_p

$$C_p = \frac{RSS_\gamma}{\hat{\sigma}^2} + 2p_\gamma - n$$

where $\hat{\sigma}^2$ is the estimate of the error variance from the full model. Mallow's C_p behaves very similar to AIC and the **lower** the C_p the better.

6.3.2 Searching Algorithms

Leap and Bounds

This is an algorithm that returns the global optimal solution among **all possible models**, but it only works for less than 50 variables.

The idea is that it finds m different models (default m in `R` is 8) of up to size p with the smallest RSS among all the models of the same size. Then evaluate the score on the p models and report the optimal one. It is important to note that the algorithm does not need to visit every model. For example, if we know that $RSS(X1, X2) < RSS(X3, X4, X5, X6)$, then we do not need to visit any size 2 or 3 *sub-models* of set $(X3, X4, X5, X6)$, which can be **leaped** over.

Birthweight Example

The library we need to run the *leaps and bound* algorithm in `R` is the `leaps` library, and the function we are going to use is the `regsubsets` function.

```
library(leaps)
regsubsets_selection=regsubsets(Birthweight~., data = birthweight2)
```

We save all the information from the `summary` output that we will need to extract in the `rs` object.

```
rs = summary(regsubsets_selection)
```

The `rs` object above contains all the different criteria for all the models that have been evaluated. The default maximum size of models is **8** and this can change with the `nvmax` option.

Below we extract the R^2, Adjusted R^2, AIC, BIC and C_p-Mallows criteria for the 8 models that have been evaluated:

Extract the models R^2

```
rs$rsq
```

```
## [1] 0.5282869 0.6911212 0.7240428 0.7449807 0.7557968 0.7649775 0.7744383
## [8] 0.7816033
```

The best model is the **8th** model. As expected, the more variables the higher the R^2.
Extract the models Adjusted R^2

```
rs$adjr2
```

```
## [1] 0.5164941 0.6752813 0.7022567 0.7174110 0.7218797 0.7246880 0.7279992
## [8] 0.7286586
```

The best model is the **7th** model.
Extract the models Cp Mallows

```
rs$cp
```

```
## [1] 24.199118  4.728123  2.387147  1.626325  2.200126  2.989580  3.742096
## [8]  4.797343
```

The best model is the **4th** model.

Extract the models BIC

```
rs$bic
```

```
## [1] -24.08280 -38.12886 -39.12472 -38.70112 -36.78368 -34.65543 -32.64344
## [8] -30.26154
```

The best model is the 3rd model.

If we want to identify which variables are included in each model, then we use the `rs$which` command, where the rows correspond to the model and the columns to the variable. `TRUE` means that the variable is included in the model and `FALSE` means otherwise.

```
rs$which
```

```
##   (Intercept) Length Headcirc Gestation smoker  mage mnocig mheight mppwt  fag
## 1        TRUE   TRUE    FALSE     FALSE  FALSE FALSE  FALSE   FALSE FALSE FALS
## 2        TRUE  FALSE     TRUE      TRUE  FALSE FALSE  FALSE   FALSE FALSE FALS
## 3        TRUE  FALSE     TRUE      TRUE   TRUE FALSE  FALSE   FALSE FALSE FALS
## 4        TRUE  FALSE     TRUE      TRUE   TRUE FALSE  FALSE   FALSE  TRUE FALS
## 5        TRUE   TRUE     TRUE      TRUE   TRUE FALSE  FALSE   FALSE  TRUE FALS
## 6        TRUE   TRUE     TRUE      TRUE   TRUE FALSE  FALSE   FALSE  TRUE FALS
## 7        TRUE   TRUE     TRUE      TRUE   TRUE  TRUE  FALSE   FALSE  TRUE FALS
## 8        TRUE   TRUE     TRUE      TRUE   TRUE  TRUE  FALSE   FALSE  TRUE FALS
##   fedyrs fnocig fheight lowbwt
## 1  FALSE  FALSE   FALSE  FALSE
## 2  FALSE  FALSE   FALSE  FALSE
## 3  FALSE  FALSE   FALSE  FALSE
## 4  FALSE  FALSE   FALSE  FALSE
## 5  FALSE  FALSE   FALSE  FALSE
## 6  FALSE  FALSE    TRUE  FALSE
## 7  FALSE  FALSE    TRUE  FALSE
## 8  FALSE   TRUE    TRUE  FALSE
```

For example, Model 1 only contains `Length` and the `intercept`.

If we want to compute the AIC and BIC criteria by "hand", we have

```
n=dim(birthweight2)[1]
msize = 1:8

AIC = n*log(rs$rss/n) + 2*msize;
which.min(AIC)
```

```
## [1] 4
```

```
BIC = n*log(rs$rss/n) + msize*log(n);
which.min(BIC)
```

```
## [1] 3
```

`leaps` does not return the AIC, but the BIC. The `leaps` BIC differs from what we compute "by hand", but the difference is a **constant**, which means that in the end we reach the same conclusion with both methods. So, the two BIC formulae (ours and the one used by `leaps`) are the same *up to a constant*.

```
names(rs)
```

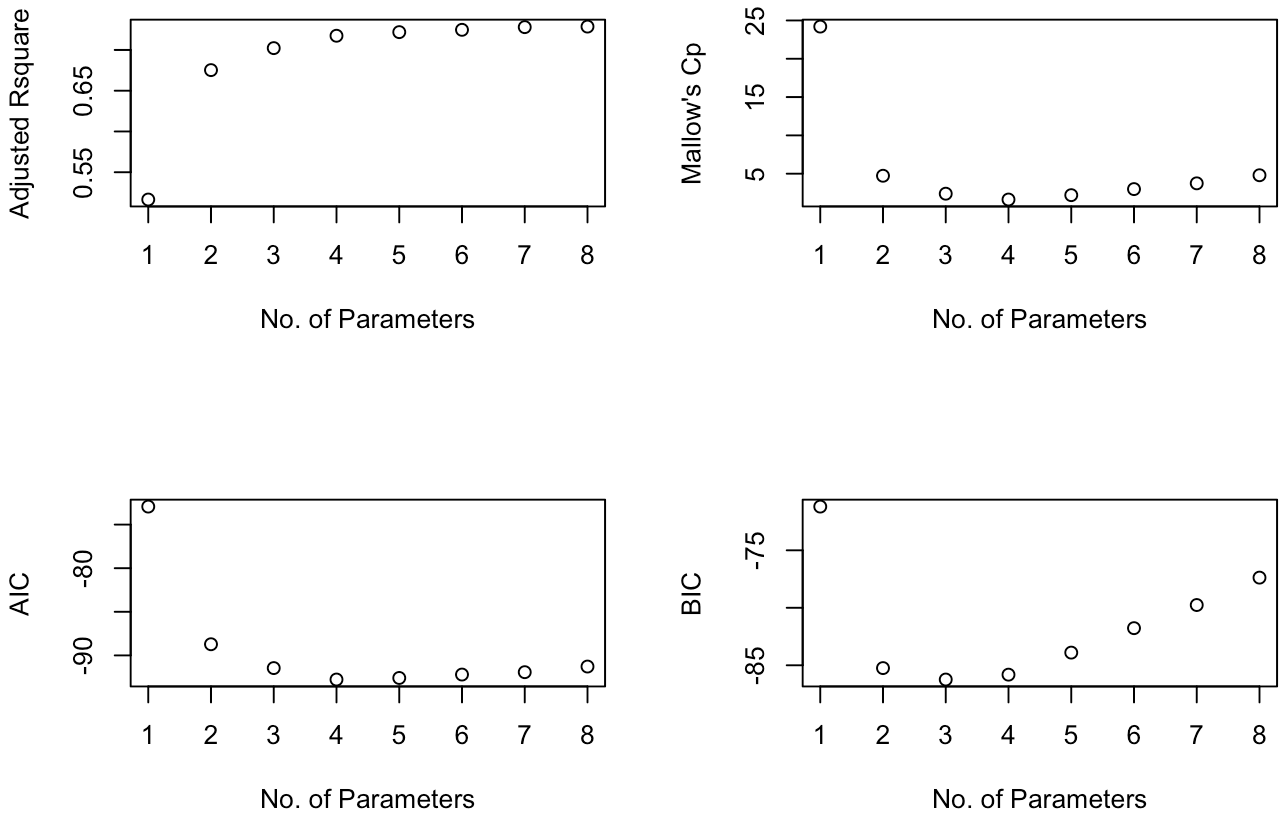
```
## [1] "which"  "rsq"    "rss"    "adjr2"  "cp"     "bic"    "outmat" "obj"
```

```
cbind(rs$bic, BIC, rs$bic-BIC )
```

```
##                      BIC
## [1,] -24.08280 -71.19838 47.11558
## [2,] -38.12886 -85.24443 47.11558
## [3,] -39.12472 -86.24030 47.11558
## [4,] -38.70112 -85.81670 47.11558
## [5,] -36.78368 -83.89926 47.11558
## [6,] -34.65543 -81.77100 47.11558
## [7,] -32.64344 -79.75901 47.11558
## [8,] -30.26154 -77.37712 47.11558
```

We can also visualize the different criteria values as a function of the number of parameters:

```
par(mfrow=c(2,2))
plot(msize, rs$adjr2, xlab="No. of Parameters", ylab = "Adjusted Rsquare");
plot(msize, rs$cp, xlab="No. of Parameters", ylab = "Mallow's Cp");
plot(msize, AIC, xlab="No. of Parameters", ylab = "AIC");
plot(msize, BIC, xlab="No. of Parameters", ylab = "BIC");
```



Again, as you can see each criterion selects a different model.

Greedy algorithms

These are algorithms that add/remove variables based on the score given by a criterion such as AIC/BIC. They can move :

- *Backwards*: start with the full model and sequentially delete predictors until the score does not improve.
- *Forward*: start with the null model and sequentially add predictors until the score does not improve.

- *Stepwise*: consider both deleting and adding one predictor at each stage.

These algorithms are computationally efficient, but they only return a *locally optimal* solution (which might be good enough in practice).

Remark: Note that these algorithms are not testing-based as the ones we discussed before, but criterion-based (although they *also* move forward or backward to find the best model).

Birthweight Example

To do stepwise model selection using the AIC criterion, we use the `step` function in `R` with `direction="both"` and the argument of the function is the full model. If we want backward elimination or forward selection, then we use `direction="backward"` or `direction="forward"` .

```
step(lm(Birthweight~., data = birthweight2), direction="both") # using the AIC cr
```

```
## Start:  AIC=-80.46
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mnocig + mheight + mppwt + fage + fedyrs + fnocig + fheight +
##     lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - mnocig     1   0.00010 3.1752 -82.456
## - fedyrs     1   0.00031 3.1754 -82.454
## - fage       1   0.00870 3.1838 -82.343
## - mheight    1   0.01743 3.1925 -82.228
## - mppwt      1   0.05993 3.2350 -81.672
## - lowbwt     1   0.06318 3.2383 -81.630
## - Length     1   0.06670 3.2418 -81.585
## - mage       1   0.07731 3.2524 -81.447
## - fnocig     1   0.10947 3.2846 -81.034
## <none>                   3.1751 -80.458
## - fheight    1   0.17819 3.3533 -80.164
## - smoker     1   0.18340 3.3585 -80.099
## - Gestation  1   0.91516 4.0903 -71.820
## - Headcirc   1   0.92558 4.1007 -71.714
##
## Step:  AIC=-82.46
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mheight + mppwt + fage + fedyrs + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - fedyrs     1   0.00041 3.1756 -84.451
## - fage       1   0.00860 3.1838 -84.343
## - mheight    1   0.01848 3.1937 -84.213
## - mppwt      1   0.05983 3.2350 -83.672
## - lowbwt     1   0.06577 3.2410 -83.595
## - Length     1   0.06712 3.2423 -83.578
## - mage       1   0.07863 3.2538 -83.429
## - fnocig     1   0.11169 3.2869 -83.004
## <none>                   3.1752 -82.456
```

```
## - fheight    1   0.18358 3.3588 -82.096
## + mnocig     1   0.00010 3.1751 -80.458
## - smoker     1   0.34671 3.5219 -80.104
## - Headcirc   1   0.94148 4.1167 -73.550
## - Gestation  1   0.94459 4.1198 -73.518
##
## Step:  AIC=-84.45
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mheight + mppwt + fage + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - fage       1   0.00830 3.1839 -86.341
## - mheight    1   0.01808 3.1937 -86.213
## - mppwt      1   0.06157 3.2372 -85.645
## - Length     1   0.06674 3.2424 -85.578
## - lowbwt     1   0.06758 3.2432 -85.567
## - mage       1   0.08746 3.2631 -85.310
## - fnocig     1   0.12082 3.2964 -84.883
## <none>                   3.1756 -84.451
## - fheight    1   0.19225 3.3679 -83.982
## + fedyrs     1   0.00041 3.1752 -82.456
## + mnocig     1   0.00019 3.1754 -82.454
## - smoker     1   0.34635 3.5220 -82.103
## - Gestation  1   0.94517 4.1208 -75.508
## - Headcirc   1   0.94827 4.1239 -75.477
##
## Step:  AIC=-86.34
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mheight + mppwt + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - mheight    1   0.01185 3.1958 -88.185
## - Length     1   0.06383 3.2477 -87.508
## - mppwt      1   0.07546 3.2594 -87.358
## - lowbwt     1   0.07892 3.2628 -87.313
## - mage       1   0.13607 3.3200 -86.584
```

```
## - fnocig     1   0.14475 3.3287 -86.474
## <none>                   3.1839 -86.341
## - fheight    1   0.21987 3.4038 -85.537
## + fage       1   0.00830 3.1756 -84.451
## + fedyrs     1   0.00011 3.1838 -84.343
## + mnocig     1   0.00001 3.1839 -84.342
## - smoker     1   0.33811 3.5220 -84.103
## - Gestation  1   0.97558 4.1595 -77.116
## - Headcirc   1   1.15887 4.3428 -75.304
##
## Step:  AIC=-88.19
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mppwt + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - lowbwt     1   0.06976 3.2655 -89.278
## - Length     1   0.10403 3.2998 -88.840
## - fnocig     1   0.13440 3.3302 -88.455
## - mage       1   0.15337 3.3491 -88.217
## <none>                   3.1958 -88.185
## - fheight    1   0.20808 3.4038 -87.536
## - mppwt      1   0.21053 3.4063 -87.506
## + mheight    1   0.01185 3.1839 -86.341
## + fage       1   0.00207 3.1937 -86.213
## + mnocig     1   0.00062 3.1951 -86.194
## + fedyrs     1   0.00000 3.1958 -86.185
## - smoker     1   0.32822 3.5240 -86.079
## - Gestation  1   0.97939 4.1752 -78.958
## - Headcirc   1   1.20221 4.3980 -76.774
##
## Step:  AIC=-89.28
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mppwt + fnocig + fheight
##
##             Df Sum of Sq    RSS     AIC
## - fnocig     1   0.10713 3.3727 -89.923
```

```
## - mage       1   0.15757 3.4231 -89.299
## <none>                   3.2655 -89.278
## - Length     1   0.16181 3.4273 -89.247
## + lowbwt     1   0.06976 3.1958 -88.185
## - mppwt      1   0.27247 3.5380 -87.913
## - fheight    1   0.29079 3.5563 -87.696
## + fage       1   0.01174 3.2538 -87.430
## + mnocig     1   0.00318 3.2623 -87.319
## + mheight    1   0.00269 3.2628 -87.313
## + fedyrs     1   0.00081 3.2647 -87.289
## - smoker     1   0.36308 3.6286 -86.850
## - Headcirc   1   1.27887 4.5444 -77.399
## - Gestation  1   1.30321 4.5687 -77.174
##
## Step:  AIC=-89.92
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mppwt + fheight
##
##             Df Sum of Sq    RSS     AIC
## - mage       1   0.14146 3.5141 -90.197
## <none>                   3.3727 -89.923
## - Length     1   0.20359 3.5762 -89.461
## - fheight    1   0.20816 3.5808 -89.407
## + fnocig     1   0.10713 3.2655 -89.278
## - smoker     1   0.26688 3.6395 -88.724
## - mppwt      1   0.27458 3.6472 -88.635
## + lowbwt     1   0.04249 3.3302 -88.455
## + fage       1   0.03471 3.3379 -88.357
## + fedyrs     1   0.00904 3.3636 -88.035
## + mnocig     1   0.00185 3.3708 -87.946
## + mheight    1   0.00000 3.3727 -87.923
## - Gestation  1   1.20036 4.5730 -79.135
## - Headcirc   1   1.28264 4.6553 -78.386
##
## Step:  AIC=-90.2
## Birthweight ~ Length + Headcirc + Gestation + smoker + mppwt +
```

```
##     fheight
##
##             Df Sum of Sq    RSS     AIC
## - fheight    1   0.13727 3.6514 -90.588
## <none>                   3.5141 -90.197
## + mage       1   0.14146 3.3727 -89.923
## - mppwt      1   0.19592 3.7100 -89.918
## - Length     1   0.19771 3.7118 -89.898
## + fnocig     1   0.09102 3.4231 -89.299
## + fedyrs     1   0.06440 3.4497 -88.974
## + lowbwt     1   0.04733 3.4668 -88.767
## + fage       1   0.03082 3.4833 -88.567
## + mheight    1   0.00396 3.5102 -88.244
## + mnocig     1   0.00204 3.5121 -88.221
## - smoker     1   0.41553 3.9296 -87.503
## - Headcirc   1   1.20067 4.7148 -79.853
## - Gestation  1   1.23437 4.7485 -79.553
##
## Step:  AIC=-90.59
## Birthweight ~ Length + Headcirc + Gestation + smoker + mppwt
##
##             Df Sum of Sq    RSS     AIC
## - Length     1   0.16173 3.8131 -90.767
## <none>                   3.6514 -90.588
## - mppwt      1   0.19142 3.8428 -90.442
## + fheight    1   0.13727 3.5141 -90.197
## + lowbwt     1   0.10447 3.5469 -89.807
## + mage       1   0.07058 3.5808 -89.407
## + fedyrs     1   0.06400 3.5874 -89.330
## + fnocig     1   0.02668 3.6247 -88.896
## + mheight    1   0.00172 3.6497 -88.607
## + fage       1   0.00083 3.6506 -88.597
## + mnocig     1   0.00005 3.6513 -88.588
## - smoker     1   0.49161 4.1430 -87.283
## - Gestation  1   1.17230 4.8237 -80.893
## - Headcirc   1   1.27513 4.9265 -80.008
```

```
##
## Step:  AIC=-90.77
## Birthweight ~ Headcirc + Gestation + smoker + mppwt
##
##             Df Sum of Sq    RSS     AIC
## <none>                   3.8131 -90.767
## + Length     1   0.16173 3.6514 -90.588
## + lowbwt     1   0.15336 3.6598 -90.491
## + fheight    1   0.10129 3.7118 -89.898
## + fedyrs     1   0.08455 3.7286 -89.709
## + mage       1   0.07388 3.7392 -89.589
## - mppwt      1   0.31307 4.1262 -89.453
## + fnocig     1   0.05263 3.7605 -89.351
## + mheight    1   0.01068 3.8024 -88.885
## + fage       1   0.00508 3.8080 -88.823
## + mnocig     1   0.00006 3.8131 -88.768
## - smoker     1   0.54119 4.3543 -87.193
## - Headcirc   1   1.89720 5.7103 -75.807
## - Gestation  1   2.88819 6.7013 -69.085
```

```
##
## Call:
## lm(formula = Birthweight ~ Headcirc + Gestation + smoker + mppwt,
##     data = birthweight2)
##
## Coefficients:
## (Intercept)     Headcirc    Gestation       smoker        mppwt
##    -5.19590      0.10208      0.11115     -0.23171      0.01291
```

If we want to use the BIC criterion, then we select `k=long(n)` .

```
step(lm(Birthweight~., data = birthweight2), direction="both", k=log(n)) # using
```

```
## Start:  AIC=-56.13
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mnocig + mheight + mppwt + fage + fedyrs + fnocig + fheight +
##     lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - mnocig     1   0.00010 3.1752 -59.867
## - fedyrs     1   0.00031 3.1754 -59.864
## - fage       1   0.00870 3.1838 -59.753
## - mheight    1   0.01743 3.1925 -59.638
## - mppwt      1   0.05993 3.2350 -59.083
## - lowbwt     1   0.06318 3.2383 -59.041
## - Length     1   0.06670 3.2418 -58.995
## - mage       1   0.07731 3.2524 -58.858
## - fnocig     1   0.10947 3.2846 -58.444
## - fheight    1   0.17819 3.3533 -57.575
## - smoker     1   0.18340 3.3585 -57.510
## <none>                   3.1751 -56.130
## - Gestation  1   0.91516 4.0903 -49.231
## - Headcirc   1   0.92558 4.1007 -49.124
##
## Step:  AIC=-59.87
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mheight + mppwt + fage + fedyrs + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - fedyrs     1   0.00041 3.1756 -63.599
## - fage       1   0.00860 3.1838 -63.491
## - mheight    1   0.01848 3.1937 -63.361
## - mppwt      1   0.05983 3.2350 -62.820
## - lowbwt     1   0.06577 3.2410 -62.743
## - Length     1   0.06712 3.2423 -62.726
## - mage       1   0.07863 3.2538 -62.577
## - fnocig     1   0.11169 3.2869 -62.152
## - fheight    1   0.18358 3.3588 -61.244
```

```
## <none>                  3.1752 -59.867
## - smoker     1   0.34671 3.5219 -59.252
## + mnocig     1   0.00010 3.1751 -56.130
## - Headcirc   1   0.94148 4.1167 -52.698
## - Gestation  1   0.94459 4.1198 -52.666
##
## Step:  AIC=-63.6
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mheight + mppwt + fage + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - fage       1   0.00830 3.1839 -67.227
## - mheight    1   0.01808 3.1937 -67.098
## - mppwt      1   0.06157 3.2372 -66.530
## - Length     1   0.06674 3.2424 -66.463
## - lowbwt     1   0.06758 3.2432 -66.452
## - mage       1   0.08746 3.2631 -66.196
## - fnocig     1   0.12082 3.2964 -65.768
## - fheight    1   0.19225 3.3679 -64.868
## <none>                  3.1756 -63.599
## - smoker     1   0.34635 3.5220 -62.989
## + fedyrs     1   0.00041 3.1752 -59.867
## + mnocig     1   0.00019 3.1754 -59.864
## - Gestation  1   0.94517 4.1208 -56.394
## - Headcirc   1   0.94827 4.1239 -56.362
##
## Step:  AIC=-67.23
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mheight + mppwt + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - mheight    1   0.01185 3.1958 -70.809
## - Length     1   0.06383 3.2477 -70.131
## - mppwt      1   0.07546 3.2594 -69.981
## - lowbwt     1   0.07892 3.2628 -69.936
## - mage       1   0.13607 3.3200 -69.207
```

```
## - fnocig     1   0.14475 3.3287 -69.097
## - fheight    1   0.21987 3.4038 -68.160
## <none>                   3.1839 -67.227
## - smoker     1   0.33811 3.5220 -66.726
## + fage       1   0.00830 3.1756 -63.599
## + fedyrs     1   0.00011 3.1838 -63.491
## + mnocig     1   0.00001 3.1839 -63.489
## - Gestation  1   0.97558 4.1595 -59.739
## - Headcirc   1   1.15887 4.3428 -57.928
##
## Step:  AIC=-70.81
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mppwt + fnocig + fheight + lowbwt
##
##             Df Sum of Sq    RSS     AIC
## - lowbwt     1   0.06976 3.2655 -73.639
## - Length     1   0.10403 3.2998 -73.201
## - fnocig     1   0.13440 3.3302 -72.816
## - mage       1   0.15337 3.3491 -72.578
## - fheight    1   0.20808 3.4038 -71.897
## - mppwt      1   0.21053 3.4063 -71.867
## <none>                   3.1958 -70.809
## - smoker     1   0.32822 3.5240 -70.440
## + mheight    1   0.01185 3.1839 -67.227
## + fage       1   0.00207 3.1937 -67.098
## + mnocig     1   0.00062 3.1951 -67.079
## + fedyrs     1   0.00000 3.1958 -67.071
## - Gestation  1   0.97939 4.1752 -63.319
## - Headcirc   1   1.20221 4.3980 -61.135
##
## Step:  AIC=-73.64
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mppwt + fnocig + fheight
##
##             Df Sum of Sq    RSS     AIC
## - fnocig     1   0.10713 3.3727 -76.021
```

```
## - mage       1   0.15757 3.4231 -75.398
## - Length     1   0.16181 3.4273 -75.346
## - mppwt      1   0.27247 3.5380 -74.011
## - fheight    1   0.29079 3.5563 -73.794
## <none>                   3.2655 -73.639
## - smoker     1   0.36308 3.6286 -72.949
## + lowbwt     1   0.06976 3.1958 -70.809
## + fage       1   0.01174 3.2538 -70.053
## + mnocig     1   0.00318 3.2623 -69.943
## + mheight    1   0.00269 3.2628 -69.936
## + fedyrs     1   0.00081 3.2647 -69.912
## - Headcirc   1   1.27887 4.5444 -63.497
## - Gestation  1   1.30321 4.5687 -63.273
##
## Step:  AIC=-76.02
## Birthweight ~ Length + Headcirc + Gestation + smoker + mage +
##     mppwt + fheight
##
##             Df Sum of Sq    RSS     AIC
## - mage       1   0.14146 3.5141 -78.033
## - Length     1   0.20359 3.5762 -77.297
## - fheight    1   0.20816 3.5808 -77.244
## - smoker     1   0.26688 3.6395 -76.560
## - mppwt      1   0.27458 3.6472 -76.472
## <none>                   3.3727 -76.021
## + fnocig     1   0.10713 3.2655 -73.639
## + lowbwt     1   0.04249 3.3302 -72.816
## + fage       1   0.03471 3.3379 -72.718
## + fedyrs     1   0.00904 3.3636 -72.396
## + mnocig     1   0.00185 3.3708 -72.307
## + mheight    1   0.00000 3.3727 -72.284
## - Gestation  1   1.20036 4.5730 -66.971
## - Headcirc   1   1.28264 4.6553 -66.222
##
## Step:  AIC=-78.03
## Birthweight ~ Length + Headcirc + Gestation + smoker + mppwt +
```

```
##     fheight
##
##             Df Sum of Sq    RSS     AIC
## - fheight    1   0.13727 3.6514 -80.162
## - mppwt      1   0.19592 3.7100 -79.492
## - Length     1   0.19771 3.7118 -79.472
## <none>                   3.5141 -78.033
## - smoker     1   0.41553 3.9296 -77.077
## + mage       1   0.14146 3.3727 -76.021
## + fnocig     1   0.09102 3.4231 -75.398
## + fedyrs     1   0.06440 3.4497 -75.072
## + lowbwt     1   0.04733 3.4668 -74.865
## + fage       1   0.03082 3.4833 -74.666
## + mheight    1   0.00396 3.5102 -74.343
## + mnocig     1   0.00204 3.5121 -74.320
## - Headcirc   1   1.20067 4.7148 -69.427
## - Gestation  1   1.23437 4.7485 -69.127
##
## Step:  AIC=-80.16
## Birthweight ~ Length + Headcirc + Gestation + smoker + mppwt
##
##             Df Sum of Sq    RSS     AIC
## - Length     1   0.16173 3.8131 -82.079
## - mppwt      1   0.19142 3.8428 -81.753
## <none>                   3.6514 -80.162
## - smoker     1   0.49161 4.1430 -78.594
## + fheight    1   0.13727 3.5141 -78.033
## + lowbwt     1   0.10447 3.5469 -77.643
## + mage       1   0.07058 3.5808 -77.244
## + fedyrs     1   0.06400 3.5874 -77.167
## + fnocig     1   0.02668 3.6247 -76.732
## + mheight    1   0.00172 3.6497 -76.444
## + fage       1   0.00083 3.6506 -76.433
## + mnocig     1   0.00005 3.6513 -76.425
## - Gestation  1   1.17230 4.8237 -72.205
## - Headcirc   1   1.27513 4.9265 -71.319
```

```
##
## Step:  AIC=-82.08
## Birthweight ~ Headcirc + Gestation + smoker + mppwt
##
##             Df Sum of Sq    RSS     AIC
## - mppwt      1   0.31307 4.1262 -82.503
## <none>                   3.8131 -82.079
## - smoker     1   0.54119 4.3543 -80.243
## + Length     1   0.16173 3.6514 -80.162
## + lowbwt     1   0.15336 3.6598 -80.065
## + fheight    1   0.10129 3.7118 -79.472
## + fedyrs     1   0.08455 3.7286 -79.283
## + mage       1   0.07388 3.7392 -79.163
## + fnocig     1   0.05263 3.7605 -78.925
## + mheight    1   0.01068 3.8024 -78.459
## + fage       1   0.00508 3.8080 -78.397
## + mnocig     1   0.00006 3.8131 -78.342
## - Headcirc   1   1.89720 5.7103 -68.856
## - Gestation  1   2.88819 6.7013 -62.135
##
## Step:  AIC=-82.5
## Birthweight ~ Headcirc + Gestation + smoker
##
##             Df Sum of Sq    RSS     AIC
## <none>                   4.1262 -82.503
## + mppwt      1    0.3131 3.8131 -82.079
## + Length     1    0.2834 3.8428 -81.753
## - smoker     1    0.4923 4.6184 -81.507
## + lowbwt     1    0.2597 3.8665 -81.495
## + mheight    1    0.1914 3.9348 -80.760
## + fheight    1    0.0851 4.0411 -79.640
## + fnocig     1    0.0726 4.0536 -79.511
## + fedyrs     1    0.0441 4.0821 -79.216
## + mage       1    0.0155 4.1107 -78.923
## + mnocig     1    0.0117 4.1145 -78.884
## + fage       1    0.0006 4.1256 -78.771
```

```
## - Headcirc   1    2.4032 6.5294 -66.964
## - Gestation  1    3.2652 7.3914 -61.756


##
## Call:
## lm(formula = Birthweight ~ Headcirc + Gestation + smoker, data = birthweight2)
##
## Coefficients:
## (Intercept)     Headcirc    Gestation       smoker
##     -5.0137       0.1117       0.1168      -0.2205
```