

The backward selection algorithm with P-values

opposite of forward selection

1. Begin with the most complicated model to be considered.
full model with all predictors
2. For every term currently included in the model, calculate a P-value for the removal of that term.
3. If the largest P-value is greater than the threshold p_{out} (usually chosen to be 0.05), remove that term to the model.
4. Iterate (2), (3) until the model contains only significant terms.

Example 4.2

Consider again the marks data in Example 4.1.

Now fit a multiple linear regression to the data using backward selection.

We will use $p_{out} = 0.05$.

Example 4.2 Solution

- ① `full <- lm(E ~ OQ + A1 + A2 + A3 + A4 + A5 + A6
 , data=marks)`
- ② `drop1(full, test = "F")`

```
## Single term deletions
##
## Model:
## E ~ OQ + A1 + A2 + A3 + A4 + A5 + A6
##      Df Sum of Sq  RSS   AIC F value    Pr(>F)
## <none>                 10.280 -1169.1
## OQ      1    0.64466 10.925 -1150.4 20.7558 7.346e-06 ***
## A1      1    0.00357 10.284 -1170.9  0.1149 0.73487
## A2      1    0.01594 10.296 -1170.5  0.5132 0.47427
## A3      1    0.18896 10.470 -1164.9  6.0838 0.01415 *
## A4      1    0.06291 10.343 -1169.0  2.0255 0.15562
## A5      1    0.02283 10.303 -1170.3  0.7352 0.39182
## A6      1    1.06880 11.349 -1137.5 34.4120 1.081e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

drop A1 from our model

- 1) Start with full model (i.e. with all 7 predictors)
- 2) Fit 7 linear models, each with a different predictor removed
- 3) For each model, perform an F -test to compare it with the null model
- 4) Find the model with the largest (non-significant) P-value and remove the corresponding predictor from our model

Example 4.2 Solution

```
③ bs1 <- update(full, .~. -A1)  
④ drop1(bs1, test = "F")
```

```
## Single term deletions  
##  
## Model:  
## E ~ OQ + A2 + A3 + A4 + A5 + A6  
##           Df Sum of Sq    RSS      AIC F value    Pr(>F)  
## <none>                 10.284 -1170.9  
## OQ           1    0.65095  10.935 -1152.1  21.0145 6.464e-06 ***  
## A2           1    0.01280  10.297 -1172.5   0.4133 0.52075  
## A3           1    0.20414  10.488 -1166.3   6.5901 0.01069 *  
## A4           1    0.06108  10.345 -1170.9   1.9720 0.16117  
## A5           1    0.02263  10.307 -1172.2   0.7306 0.39330  
## A6           1    1.08120  11.365 -1139.0  34.9042 8.579e-09 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

drop A2 from
our model

Example 4.2 Solution

```
bs2 <- update(bs1, .~. -A2)  
drop1(bs2, test = "F")
```

```
## Single term deletions  
##  
## Model:  
## E ~ OQ + A3 + A4 + A5 + A6  
##      Df Sum of Sq    RSS      AIC F value    Pr(>F)  
## <none>                 10.297 -1172.5  
## OQ      1    0.64366 10.941 -1154.0 20.8160 7.118e-06 ***  
## A3      1    0.19343 10.490 -1168.2  6.2554 0.01286 *  
## A4      1    0.05453 10.351 -1172.7  1.7634 0.18511  
## A5      1    0.02353 10.320 -1173.8  0.7609 0.38366  
## A6      1    1.07782 11.375 -1140.8 34.8565 8.750e-09 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

drop A5

Example 4.2 Solution

```
bs3 <- update(bs2, .~. -A5)  
drop1(bs3, test = "F")
```

```
## Single term deletions  
##  
## Model:  
## E ~ OQ + A3 + A4 + A6  
##      Df Sum of Sq    RSS    AIC F value    Pr(>F)  
## <none>            10.320 -1173.8  
## OQ      1    0.70510 11.025 -1153.3 22.8190 2.670e-06 ***  
## A3      1    0.21622 10.537 -1168.7  6.9974 0.008549 **  
## A4      1    0.07088 10.391 -1173.4  2.2939 0.130826  
## A6      1    1.53773 11.858 -1128.7 49.7654 1.004e-11 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

drop A4

Example 4.2 Solution

```
bs4 <- update(bs3, .~. -A4)  
drop1(bs4, test = "F")
```

```
## Single term deletions  
##  
## Model:  
## E ~ OQ + A3 + A6  
##      Df Sum of Sq    RSS    AIC F value    Pr(>F)  
## <none>            10.391 -1173.4  
## OQ      1    0.90140 11.293 -1147.2   29.060 1.327e-07 ***  
## A3      1    0.33372 10.725 -1164.7   10.759 0.001147 **  
## A6      1    2.42664 12.818 -1104.3   78.231 < 2.2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

all predictors have
a significant P-value
We can stop here.

Example 4.2 Solution

Summary(bs4)

```
##  
## Call:  
## lm(formula = E ~ OQ + A3 + A6, data = stats_marks)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.81856 -0.06018  0.02859  0.09063  0.60694   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  0.13219    0.03273   4.039 6.65e-05 ***  
## OQ           0.20085    0.03726   5.391 1.33e-07 ***  
## A3           0.14387    0.04386   3.280  0.00115 **   
## A6           0.36301    0.04104   8.845 < 2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.1761 on 335 degrees of freedom  
## Multiple R-squared:  0.5088, Adjusted R-squared:  0.5044   
## F-statistic: 115.7 on 3 and 335 DF,  p-value: < 2.2e-16
```