DATA2002

Multiple regression and model selection

Garth Tarr

THE UNIVERSITY OF SYDNEY

Interpreting model coefficients

Multiple regression

Model selection

Interpreting model coefficients

How can we interpret the estimated coefficients?

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

- The intercept is the expected value of Y when $x = 0$. I.e. $\mathrm{E}(Y \mid x = 0) = \beta_0$.
- The slope is the amount we expect Y to change by when x increases one unit. I.e. for a one unit increase in x we expect Y to change by β_1 (could be an increase or decrease depending on the sign).

Recall our fitted model

```r
library(tidyverse)
data(environmental, package = "lattice")
environmental = environmental %>%
  mutate(lozone = log(ozone))
lm2 = lm(lozone ~ temperature, data = environmental)
lm2
```

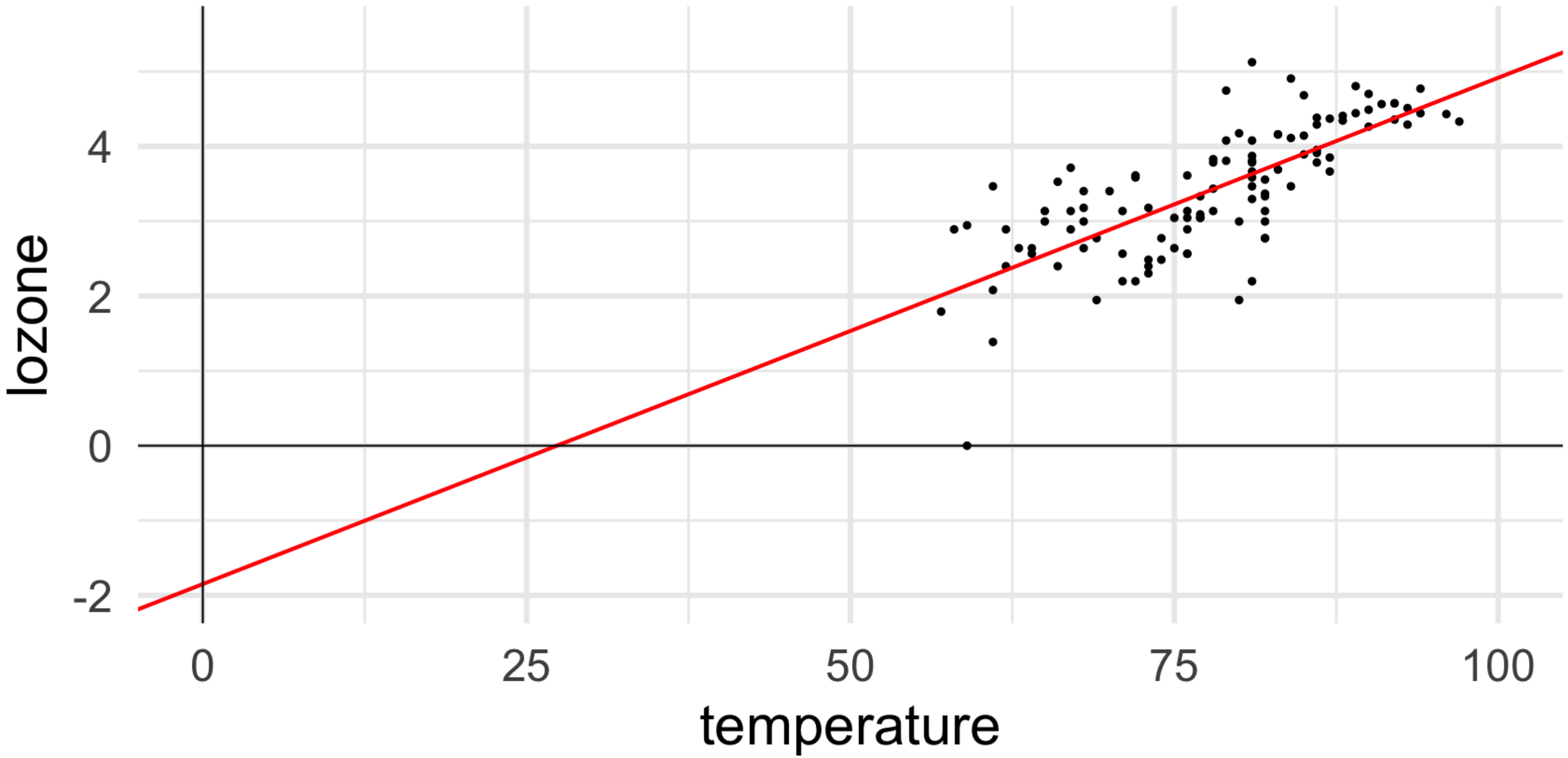
```
##
## Call:
## lm(formula = lozone ~ temperature, data = environmental)
##
## Coefficients:
## (Intercept)  temperature
##    -1.84852      0.06767
```

$$\widehat{\log(\text{ozone})} = -1.84852 + 0.06767 \times \text{temperature}$$

How do we interpret this model?

- Does it make sense to interpret the intercept?

```r
ggplot(environmental, aes(x = temperature, y = lozone)) +
  geom_point() + coord_cartesian(xlim = c(0,100), ylim = c(-2, 5.5)) +
  geom_abline(slope = 0.06767, intercept = -1.84852, colour = "red", size = 1) +
  geom_vline(xintercept = 0) + geom_hline(yintercept = 0) + theme_minimal(base_size = 30)
```



Slope interpretation

$$\widehat{\log(\text{ozone})} = -1.84852 + 0.06767 \times \text{temperature}$$

- Interpreting the slope: a one degree Fahrenheit increase in temperature results in a 0.07 unit **increase** in log ozone, on average.
- A nicer way to interpret this is: a one degree Fahrenheit increase in temperature results in a 7% **increase** in ozone, on average.

Interpreting models with log transformations

Log-linear $\log(Y) = \beta_0 + \beta_1 x$

On average, a one unit increase in x will result in a $\beta_1 \times 100\%$ change in Y.

Linear-log $Y = \beta_0 + \beta_1 \log(x)$

On average, a one percent increase in x will result in a $\beta_1/100$ change in Y.

Log-log $\log(Y) = \beta_0 + \beta_1 \log(x)$

On average, a one percent increase in x will result in a $\beta_1\%$ change in Y.

Multiple regression

Multiple regression

What if we extended our question? Can *radiation*, *temperature* and *wind* be used to predict the log of *ozone*?

$$\log(\text{ozone})_i = \beta_0 + \beta_1 \text{radiation}_i + \beta_2 \text{temperature}_i + \beta_3 \text{wind}_i + \varepsilon_i$$

```
lm3 = lm(lozone ~ radiation + temperature + wind, environmental)
summary(lm3)$coefficients %>% round(4)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2612     0.5534 -0.4719   0.6379
## radiation     0.0025     0.0006  4.5176   0.0000
## temperature   0.0492     0.0061  8.0776   0.0000
## wind         -0.0616     0.0157 -3.9222   0.0002
```

Fitted model:

$$\log(\widehat{\text{ozone}}) = -0.2612 + 0.0025\,\text{radiation} + 0.0492\,\text{temperature} - 0.0616\,\text{wind}$$

Multiple regression

Multiple regression is a natural extension of simple linear regression that incorporates multiple explanatory (or predictor) variables. It has the general form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$.

Often it's convenient to write the model in matrix format,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)'$, $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$ and

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix},$$

where $\boldsymbol{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})$ is the vector of predictors for the ith observation.

Multiple regression

The least squares solution is:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

Interpretation

The estimated coefficients ($\hat{\beta}$'s) are now interpreted as "consitional on" the other variables - each β_i reflects the predicted change in y associated with a one unit increase in x_i, holding the other variables constant (i.e. a marginal effect).

$$\log\widehat{(\text{ozone})} = -0.2612 + 0.0025\,\text{radiation} + 0.0492\,\text{temperature} - 0.0616\,\text{wind}$$

- A one degree Fahrenheit increase in temperature results in a 5% **increase** in ozone on average, holding radiation and wind speed constant.
- A one langley increase solar ratiation results in a 0.3% **increase** in ozone on average, holding radiation and wind constant.
- A 10 langley increase solar ratiation results in a 3% **increase** in ozone on average, holding radiation and wind constant.
- A one mile per hour increase in average wind speed results in a 6% **decrease** in ozone on average, holding radiation and temperature constant.

In-sample performance

The r^2 value has the same interpretation: proportion of total variability in Y explained by the regression model.

Simple linear regression model

```
summary(lm2)$r.squared
```

```
## [1] 0.5547615
```

"Full" model

```
summary(lm3)$r.squared
```

```
## [1] 0.664515
```

```
sjPlot::tab_model(lm2, lm3, digits = 4, show.ci = FALSE)
```

	lozone		**lozone**	
Predictors	*Estimates*	*p*	*Estimates*	*p*
(Intercept)	-1.8485	**<0.001**	-0.2612	0.638
temperature	0.0677	**<0.001**	0.0492	**<0.001**
radiation			0.0025	**<0.001**
wind			-0.0616	**<0.001**
Observations	111		111	
R^2 / R^2 adjusted	0.555 / 0.551		0.665 / 0.655	

Model selection

US crime rate data

- Ehrlich (1973) considered crime data from 47 states of the USA in 1960
- Our aim is to model the crime rate as a function of up to 15 potential explanatory variables
- We will consider **stepwise** search schemes

```
data("UScrime", package = "MASS")
# ?UScrime
dim(UScrime)
```

```
## [1] 47 16
```

Crime data: variables in the data set

Variable	Description
M	percentage of males aged 14-24
So	indicator variable for a southern state
Ed	mean years of schooling
Po1	police expenditure in 1960
Po2	police expenditure in 1959
LF	labour force participation rate
M.F	number of males per 1000 females
Pop	state population
NW	number of nonwhites per 1000 people
U1	unemployment rate of urban males 14-24
U2	unemployment rate of urban males 35-39
GDP	gross domestic product per head
Ineq	income inequality
Prob	probability of imprisonment
Time	average time served in state prisons
y	rate of crimes in a particular category per head of population

Crime data

```
data("UScrime", package = "MASS")
dim(UScrime)
```

```
## [1] 47 16
```

```
n = 47
k = 15
2^k
```

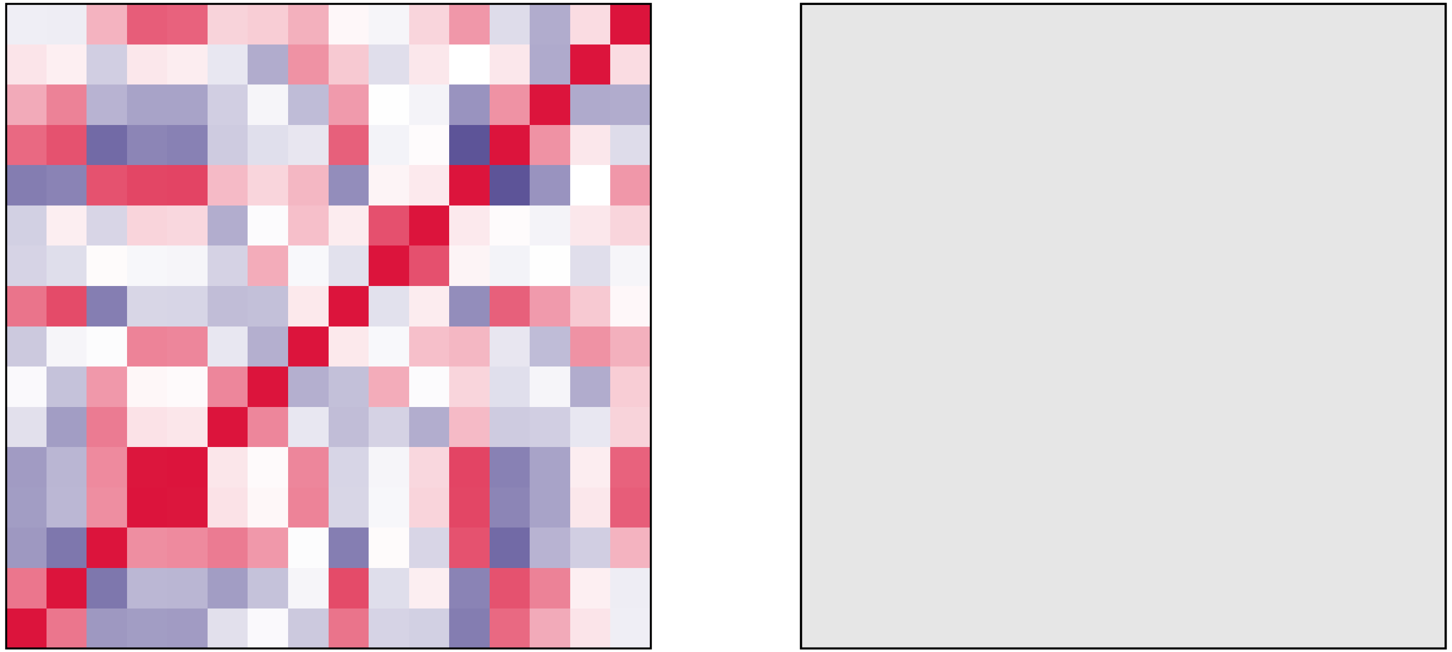
```
## [1] 32768
```

```
cor_mat = cor(UScrime)
melted_cor_mat = cor_mat %>%
  data.frame() %>%
  rownames_to_column(var = "var1") %>%
  gather(key = "var2", value = "cor", -var1)
```

```
ggplot(data = melted_cor_mat,
       aes(x=var1, y=var2, fill=cor)) +
  geom_tile() + theme_minimal(base_size = 30) +
  scale_fill_gradient2(
    low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1,1)) +
  theme(axis.text.x = element_text(angle = 90,
```

Interactive correlation matrix

```
qtlcharts::iplotCorr(UScrime)
```



Crime data: null and full model

```r
M0 = lm(y ~ 1, data = UScrime)  # Null model
M1 = lm(y ~ ., data = UScrime)  # Full model
round(summary(M1)$coef, 3)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5984.288   1628.318  -3.675    0.001
## M               8.783      4.171   2.106    0.043
## So             -3.803    148.755  -0.026    0.980
## Ed             18.832      6.209   3.033    0.005
## Po1            19.280     10.611   1.817    0.079
## Po2           -10.942     11.748  -0.931    0.359
## LF             -0.664      1.470  -0.452    0.655
## M.F             1.741      2.035   0.855    0.399
## Pop            -0.733      1.290  -0.568    0.574
## NW              0.420      0.648   0.649    0.521
## U1             -5.827      4.210  -1.384    0.176
## U2             16.780      8.234   2.038    0.050
## GDP             0.962      1.037   0.928    0.361
## Ineq            7.067      2.272   3.111    0.004
## Prob        -4855.266   2272.375  -2.137    0.041
## Time           -3.479      7.165  -0.486    0.631
```

```r
t(round(broom::glance(M1), 2))
```

```
##                     [,1]
## r.squared           0.80
## adj.r.squared       0.71
## sigma             209.06
## statistic           8.43
## p.value             0.00
## df                 15.00
## logLik           -308.01
## AIC               650.03
## BIC               681.48
## deviance      1354945.77
## df.residual        31.00
## nobs               47.00
```

The `drop1` and `update` command in R

- For a response variable `Y` and explanatory variables `x1, ..., xk` stored in the data frame `dat` consider
- `M1 = lm(Y ~ ., data = dat)`
- The function `drop1(M1, test = "F")` returns a number of information criteria for all explanatory variables used in `M1`
- In particular, this includes the p-values of the F-test for $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$ for all $j = 1, \ldots, k$
- To efficiently delete a variable from regression model `M1`, say `x1`, the `update()` function can be used:
- `M2 = update(M1, . ~ . - x1)`
- The full stops in the `update()` formula `.~.` stand for "whatever was in the corresponding position in the old formula"

Crime data: `drop1` and `update`

Start with the full model, `M1`

```r
drop1(M1, test = "F")
```

```
## Single term deletions
##
## Model:
## y ~ M + So + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 +
##     GDP + Ineq + Prob + Time
##        Df Sum of Sq     RSS AIC F value Pr(>F)
## <none>              1354946 515
## M       1    193770 1548716 519    4.43 0.0434 *
## So      1        29 1354974 513    0.00 0.9798
## Ed      1    402117 1757063 525    9.20 0.0049 **
## Po1     1    144306 1499252 517    3.30 0.0789 .
## Po2     1     37919 1392865 514    0.87 0.3588
## LF      1      8917 1363862 513    0.20 0.6547
## M.F     1     31967 1386913 514    0.73 0.3990
## Pop     1     14122 1369068 513    0.32 0.5738
## NW      1     18395 1373341 513    0.42 0.5213
## U1      1     83722 1438668 515    1.92 0.1762
## U2      1    181536 1536482 519    4.15 0.0502 .
## GDP     1     37613 1392558 514    0.86 0.3608
## Ineq    1    423031 1777977 525    9.68 0.0040 **
## Prob    1    199538 1554484 519    4.57 0.0406 *
## Time    1     10304 1365250 513    0.24 0.6307
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
M2 = update(M1, . ~ . - So)
drop1(M2, test = "F")
```

```
## Single term deletions
##
## Model:
## y ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 + GDP +
##     Ineq + Prob + Time
##        Df Sum of Sq     RSS AIC F value Pr(>F)
## <none>              1354974 513
## M       1    195084 1550059 517    4.61 0.0395 *
## Ed      1    403140 1758114 523    9.52 0.0042 **
## Po1     1    144302 1499277 515    3.41 0.0742 .
## Po2     1     37954 1392929 512    0.90 0.3509
## LF      1     10878 1365852 511    0.26 0.6157
## M.F     1     32449 1387423 512    0.77 0.3879
## Pop     1     14127 1369101 511    0.33 0.5676
## NW      1     21626 1376600 511    0.51 0.4800
## U1      1     96420 1451395 514    2.28 0.1411
## U2      1    189859 1544834 517    4.48 0.0421 *
## GDP     1     39223 1394197 512    0.93 0.3430
## Ineq    1    488834 1843808 525   11.54 0.0018 **
## Prob    1    204463 1559437 517    4.83 0.0353 *
## Time    1     10341 1365315 511    0.24 0.6246
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
M3 = update(M2, . ~ . - Time)
```

Crime data: `drop1` and `update`

```r
drop1(M3, test = "F")
```

```
## Single term deletions
##
## Model:
## y ~ M + Ed + Po1 + Po2 + LF + M.F + Pop + NW + U1 + U2 + GDP +
##     Ineq + Prob
##        Df Sum of Sq     RSS AIC F value Pr(>F)
## <none>              1365315 511
## M       1    186110 1551425 515    4.50 0.0415 *
## Ed      1    409448 1774763 521    9.90 0.0035 **
## Po1     1    134137 1499452 513    3.24 0.0809 .
## Po2     1     28932 1394247 510    0.70 0.4090
## LF      1     10533 1375848 509    0.25 0.6172
## M.F     1     41784 1407099 510    1.01 0.3222
## Pop     1     21846 1387161 510    0.53 0.4726
## NW      1     15482 1380797 510    0.37 0.5449
## U1      1     91420 1456735 512    2.21 0.1466
## U2      1    184143 1549458 515    4.45 0.0426 *
## GDP     1     36070 1401385 510    0.87 0.3572
## Ineq    1    502909 1868224 524   12.16 0.0014 **
## Prob    1    237493 1602808 517    5.74 0.0224 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
M4 = update(M3, . ~ . - LF)
```

```r
drop1(M4, test = "F")
```

```
## Single term deletions
##
## Model:
## y ~ M + Ed + Po1 + Po2 + M.F + Pop + NW + U1 + U2 + GDP + Ineq +
##     Prob
##        Df Sum of Sq     RSS AIC F value Pr(>F)
## <none>              1375848 509
## M       1    217716 1593564 514    5.38 0.0265 *
## Ed      1    413254 1789103 520   10.21 0.0030 **
## Po1     1    123896 1499744 511    3.06 0.0892 .
## Po2     1     21418 1397266 508    0.53 0.4719
## M.F     1     31252 1407100 508    0.77 0.3857
## Pop     1     27803 1403651 508    0.69 0.4129
## NW      1     11675 1387523 508    0.29 0.5947
## U1      1     80954 1456802 510    2.00 0.1663
## U2      1    190746 1566594 513    4.71 0.0370 *
## GDP     1     35035 1410883 509    0.87 0.3587
## Ineq    1    500944 1876792 522   12.38 0.0013 **
## Prob    1    226971 1602819 515    5.61 0.0237 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
M5 = update(M4, . ~ . - NW)
## drop1(M5, test = "F")
```

Backward variable selection

1. Start with model containing all possible explanatory variables
2. For each variable in turn, investigate effect of removing variable from current model
3. Remove the least informative variable, unless this variable is nonetheless supplying significant information about the response
4. Go to step 2. Stop only if all variables in the current model are important

The Akaike information criterion

- **AIC** is the most widely known and used model selection method (of course this does not imply it is the best/recommended method to use)
- The **AIC** was introduced by Hirotugu Akaike in his seminal 1973 paper and for our linear regression models, is defined as,

$$\text{AIC} = n \log\left(\frac{\text{Residual sum of squares}}{n}\right) + 2p$$

- The smaller the **AIC** the better the model
- Models that differ by less than one or two **AIC** values can be regarded as somewhat equally well fitting

Crime data: backward search using AIC

```
step.back.aic = step(M1,
                     direction = "backward",
                     trace = FALSE)
round(summary(step.back.aic)$coef,3)
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.101   1194.611  -5.379    0.000
## M               9.332      3.350   2.786    0.008
## Ed             18.012      5.275   3.414    0.002
## Po1            10.265      1.552   6.613    0.000
## M.F             2.234      1.360   1.642    0.109
## U1             -6.087      3.339  -1.823    0.076
## U2             18.735      7.248   2.585    0.014
## Ineq            6.133      1.396   4.394    0.000
## Prob        -3796.032   1490.646  -2.547    0.015
```

Use `trace = TRUE` to see the sequence of models considered by the procedure.

```
step.back.aic %>%
  broom::glance() %>%
  round(2) %>% t()
```

```
##                     [,1]
## r.squared           0.79
## adj.r.squared       0.74
## sigma             195.55
## statistic          17.74
## p.value             0.00
## df                  8.00
## logLik           -309.66
## AIC               639.32
## BIC               657.82
## deviance      1453067.77
## df.residual        38.00
## nobs               47.00
```

Crime data: `drop1`

We could try to see if we can drop any variables from the backward stepwise using AIC model:

```
drop1(step.back.aic, test = "F")
```

```
## Single term deletions
##
## Model:
## y ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob
##        Df Sum of Sq     RSS AIC F value  Pr(>F)
## <none>              1453068 504
## M       1    296790 1749858 511    7.76  0.0083 **
## Ed      1    445788 1898855 515   11.66  0.0015 **
## Po1     1   1672038 3125105 538   43.73 8.3e-08 ***
## M.F     1    103159 1556227 505    2.70  0.1087
## U1      1    127044 1580112 506    3.32  0.0762 .
## U2      1    255443 1708511 510    6.68  0.0137 *
## Ineq    1    738244 2191312 521   19.31 8.6e-05 ***
## Prob    1    247978 1701046 509    6.49  0.0151 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Forward variable selection

1. Start with model containing no explanatory variables, i.e. `lm(y ~ 1, data = dat)`.
2. For each variable in turn, investigate effect of adding variable from current model
3. Add the most informative/significant variable, unless this variable is not supplying significant information about the response
4. Go to step 2. Stop only if none of the non-included variables is important

Crime data: forward search using AIC

```r
M0 = lm(y ~ 1, data = UScrime)  # Null model
M1 = lm(y ~ ., data = UScrime)  # Full model
step.fwd.aic = step(M0, scope = list(lower = M0, upper = M1), direction = "forward", trace = FALSE)
summary(step.fwd.aic)
```

```
## 
## Call:
## lm(formula = y ~ Po1 + Ineq + Ed + M + Prob + U2, data = UScrime)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -470.7  -78.4  -19.7  133.1  556.2 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -5040.50     899.84   -5.60  1.7e-06 ***
## Po1            11.50       1.38    8.36  2.6e-10 ***
## Ineq            6.77       1.39    4.85  1.9e-05 ***
## Ed             19.65       4.48    4.39  8.1e-05 ***
## M              10.50       3.33    3.15   0.0031 ** 
## Prob        -3801.84    1528.10   -2.49   0.0171 *  
## U2              8.94       4.09    2.18   0.0348 *  
## ---
```

The `add1()` function

- For a response variable `Y` and explanatory variables `x1`, ..., `xk` stored in the data frame `dat` consider
  ```
  M0 = lm(Y ~ 1, data = dat)
  ```
- The R function
  ```
  add1(M0, scope = ~ x1+x2+...+xk, data=dat, test="F")
  ```
  returns a number of information criteria for all variables specified after the option `scope= ~` to model the response variable
- Alternatively you can avoid listing all the variable names by coding up a full model
  ```
  Mf = lm(Y ~ ., data = dat)
  ```
  and then using
  ```
  add1(M0, scope = Mf, data = dat, test = "F")
  ```
- In particular, this includes the p-values of the F-test for $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$ for all $j \notin \alpha_{\mathrm{M0}}$

Crime data: `add1()`

```
add1(step.fwd.aic, test = "F", scope = M1)
```

```
## Single term additions
##
## Model:
## y ~ Po1 + Ineq + Ed + M + Prob + U2
##        Df Sum of Sq     RSS AIC F value Pr(>F)
## <none>                1611057 505
## So      1     17958 1593098 506    0.44   0.51
## Po2     1     25017 1586040 506    0.62   0.44
## LF      1     13179 1597878 506    0.32   0.57
## M.F     1     30945 1580112 506    0.76   0.39
## Pop     1     51320 1559737 505    1.28   0.26
## NW      1       359 1610698 507    0.01   0.93
## U1      1     54830 1556227 505    1.37   0.25
## GDP     1     59910 1551147 505    1.51   0.23
## Time    1      7159 1603898 507    0.17   0.68
```

Comparing forward and backwards stepwise

```
sjPlot::tab_model(
  step.fwd.aic, step.back.aic,
  show.ci = FALSE,
  show.aic = TRUE,
  dv.labels = c("Forward model",
                "Backward model")
)
```

	Forward model		Backward model	
Predictors	*Estimates*	*p*	*Estimates*	*p*
(Intercept)	-5040.50	**<0.001**	-6426.10	**<0.001**
Po1	11.50	**<0.001**	10.27	**<0.001**
Ineq	6.77	**<0.001**	6.13	**<0.001**
Ed	19.65	**<0.001**	18.01	**0.002**
M	10.50	**0.003**	9.33	**0.008**
Prob	-3801.84	**0.017**	-3796.03	**0.015**
U2	8.94	**0.035**	18.73	**0.014**
M.F			2.23	0.109
U1			-6.09	0.076
Observations	47		47	
R^2 / R^2 adjusted	0.766 / 0.731		0.789 / 0.744	
AIC	640.166		639.315	

Exhaustive searches

Exhaustive searching the full set of possible models is totally feasible for $p < 100$.

```r
library(leaps)
exh = regsubsets(y~., data = UScrime, nvmax = 15)
summary(exh)$outmat
```

```
##           M   So  Ed  Po1 Po2 LF  M.F Pop NW  U1  U2  GDP Ineq Prob Time
## 1  ( 1 )  " " " " " " "*" " " " " " " " " " " " " " " " " " "  " "  " " 
## 2  ( 1 )  " " " " " " "*" " " " " " " " " " " " " " " " " "*"  " "  " " 
## 3  ( 1 )  " " " " "*" "*" " " " " " " " " " " " " " " " " "*"  " "  " " 
## 4  ( 1 )  "*" " " "*" "*" " " " " " " " " " " " " " " " " "*"  " "  " " 
## 5  ( 1 )  "*" " " "*" "*" " " " " " " " " " " " " " " " " "*"  "*"  " " 
## 6  ( 1 )  "*" " " "*" "*" " " " " " " " " " " " " "*" " " "*"  "*"  " " 
## 7  ( 1 )  "*" " " "*" "*" " " " " " " " " " " " " "*" "*" "*"  "*"  " " 
## 8  ( 1 )  "*" " " "*" "*" " " " " "*" " " " " "*" "*" " " "*"  "*"  " " 
## 9  ( 1 )  "*" " " "*" "*" " " " " "*" " " " " "*" "*" "*" "*"  "*"  " " 
## 10  ( 1 ) "*" " " "*" "*" " " " " "*" "*" " " "*" "*" "*" "*"  "*"  " " 
## 11  ( 1 ) "*" " " "*" "*" "*" " " "*" "*" " " "*" "*" "*" "*"  "*"  " " 
## 12  ( 1 ) "*" " " "*" "*" "*" " " "*" "*" "*" "*" "*" "*" "*"  "*"  " " 
## 13  ( 1 ) "*" " " "*" "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"  "*"  " " 
## 14  ( 1 ) "*" " " "*" "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"  "*"  "*" 
## 15  ( 1 ) "*" "*" "*" "*" "*" "*" "*" "*" "*" "*" "*" "*" "*"  "*"  "*" 
```

Review: stepwise variable selection

1. Start with **some** model, typically null model (with no explanatory variables) or full model (with all variables)
2. For each variable in the current model, **investigate** effect of **removing** it
3. **Remove** the least informative variable, unless this variable is nonetheless supplying significant information about the response
4. For each variable not in the current model, **investigate** effect of **including** it
5. **Include** the most statistically significant variable not currently in model (unless no significant variable exists)
6. Go to step 2. Stop only if no change in steps 2-5

References

Akaike, H. (1973). "Information theory and and extension of the maximum likelihood principle". In: *Second International Symposium on Information Theory*. Ed. by B. Petrov and F. Csaki. Academiai Kiado, pp. 267-281.

Baumer, B. S., D. T. Kaplan, and N. J. Horton (2017). *Modern Data Science with R*. Boca Raton: Chapman and Hall/CRC. URL: https://mdsr-book.github.io/index.html.

Ehrlich, I. (1973). "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation". In: *Journal of Political Economy* 81.3, pp. 521-565. URL: http://www.jstor.org/stable/1831025.

Tarr, G., S. Müller, and A. Welsh (2018). "mplot: An R Package for Graphical Model Stability and Variable Selection Procedures". In: *Journal of Statistical Software* 83.9, pp. 1-28. DOI: 10.18637/jss.v083.i09.