

Tutorial 08

ENVX2001 – Applied Statistical Methods

Semester 1

Exercise 1: Backward elimination in R

Data: *Dippers* spreadsheet

Dippers are thrush-sized birds living mainly in the upper reaches of rivers, which feed on benthic invertebrates by probing the river beds with their beaks. The dataset in this exercise contains data from a biological survey which examined the nature of the variables thought to influence the breeding of British dippers.

Twenty-two sites were included in the survey. Some of the variables have been transformed.

The variables measured were:

- Altitude site altitude
- Hardness water hardness
- RiverSlope river-bed slope
- LogCadd the numbers of caddis fly larvae, transformed
- LogStone the numbers of stonefly larvae, transformed
- LogMay the numbers of mayfly larvae, transformed
- LogOther the numbers of all other invertebrates collected, transformed
- Br_Dens the number of breeding pairs of dippers per 10 km of river

In the analyses, the four invertebrate variables were transformed using a $\log(x + 1)$ transformation.

CODE

```
library(readxl)
Dippers <- read_xlsx("data/mlr.xlsx" , sheet = "Dippers")
glimpse(Dippers)
```

OUTPUT

```
Rows: 22
Columns: 8
$ Altitude    <dbl> 259, 198, 251, 184, 145, 145, 198, 160, 251, 159, 160, 145,...
$ Hardness    <dbl> 12.20, 22.00, 26.30, 22.50, 29.50, 39.90, 42.80, 59.60, 69.0...
$ RiverSlope  <dbl> 10.90, 14.70, 6.90, 4.60, 1.91, 5.00, 6.20, 14.30, 4.60, 3.0...
$ Br_Dens     <dbl> 3.60, 4.30, 3.80, 3.40, 3.80, 4.50, 4.30, 5.00, 4.50, 3.40,...
$ LogCadd     <dbl> 2.303, 2.890, 3.784, 4.419, 3.219, 3.932, 3.664, 4.431, 3.7...
$ LogStone    <dbl> 5.242, 4.344, 5.231, 5.242, 3.829, 4.898, 4.357, 6.337, 5.4...
$ LogMay      <dbl> 0.000, 3.401, 5.826, 5.749, 5.509, 5.749, 5.371, 0.000, 4.8...
$ LogOther    <dbl> 1.386, 1.609, 1.386, 1.386, 1.099, 3.045, 1.386, 2.944, 2.5...
```

You may explore the data on your own (hint: look at last week's exercise on histogram and scatterplot matrices).

When ready, perform a backward elimination starting from the full model:

CODE

```
FullMod <- lm(Br_Dens ~ ., data=Dippers)
RedMod <- step(FullMod, direction = "backward")
summary(FullMod)
summary(RedMod)
AIC(FullMod, RedMod)
```

Question 1

Which model is chosen? Why?