

Chapter 12 - Strategies for Variable Selection

April 23, 2018

Setting: In previous chapters we had a response variable and a small set of potential explanatory variables that either:

- directly related to the research question (e.g., Sex when studying Alcohol metabolism in Chapter 11); or
- were known potential confounding variables (e.g., Body size when studying Brain size in Chapter 9).

In Chapter 12 we consider many potential explanatory variables without *a priori* information to suggest any over the others. We only want a small subset of the possible explanatory variables in the model. This may be because you want to apply Occam's Razor, or maybe because some of the explanatory variables are **collinear**.

How do you determine which explanatory variables to include in the model?

We have already seen how to use R^2 and extra sum of squares tests to help answer this question. Chapter 12 offers a few more tools, called **variable selection techniques**.

12.2.1 Objectives

CAUTIONS:

- Variable selection techniques are *not* going to uncover some important law of nature.
- Results from “automatic” variable selection techniques can be misleading (e.g., the forward selection and backward elimination techniques that your book presents).
- Regardless of what variable selection techniques suggest, always include variables in your model that allow you to address *well-defined questions of interest*.
- It is tempting to think that finding *the* important set of explanatory variables *is* a question of interest, but the actual set chosen by variable selection techniques is just ONE OF MANY equally good sets.
- “All models are wrong, some are useful” - George Box.

When analyzing data via MLR, you typically have 1 of the following 3 objectives:

Objective I: Adjusting

- **Goal:** examine the effect of a few explanatory variables on the response **after accounting** for other, potentially confounding, explanatory variables. There is little or no interest in the confounding variables themselves.
- For example, we studied how Brain size was affected by Gestation length and Litter size **after accounting for** Body size in Chapter 9. Body size is a potential confounding variable that has a known positive association with Brain size. It was clear in that problem to include all 3 variables: the research question required that Gestation and Litter be included; and due to possible confounding, Body was included.
- **Approach:** Find a subset of the possible confounding variables using a variable selection technique (e.g., Body size) and then add the explanatory variables of interest (e.g., Gestation and Litter) into that model.
- Although a particular confounding explanatory variable may not be included in the final model, it has still been adjusted for, since it was given a chance to be in the final model.

- The coefficients for variables chosen through variable selection are *not* specifically interpreted **and** no specific meaning is attached to the particular set of variables chosen.

Objective II: Explaining

- **Goal:** answer a vague question, such as

Which variables are important in explaining the response?

or

What is the relationship between the response and explanatory variables?

These questions are all too common. If you find yourself stating a question of interest in such general terms, be prepared, you are on a fishing expedition.

- The analyst is often tempted to attach meaning to the set of variables chosen by variable selection techniques and/or interpret coefficients but the explanatory variables chosen are NOT necessarily SPECIAL.
- For explanation purposes, err on the side of using simpler models compared to more complicated ones.
- The inclusion/exclusion of variables can be strongly influenced by the dependencies between and among them (i.e., **multicollinearity**).
- Interpreting the coefficients when explanatory variables are correlated can be difficult. For example, a coefficient may have an opposite sign of that suggested by the scatterplot.
- Regression coefficients are technically interpreted as the effect of one variable while holding all other variables fixed. But, if two explanatory variables are highly correlated it may not make much sense to talk of increasing one while holding the other one constant. For example, it may be useful to think of a subpopulation of subjects who weigh the same but have different heights and to think about the association between the response and heights for that subpopulation.

Objective III: Predicting

- If the purpose of the model is prediction, then no interpretation of the individual coefficients is needed.
- Variable selection techniques are useful for selecting a few models that should do well at predicting either the mean response or a future response. Choose one with most easily available explanatory variables.
- For prediction purposes, you may err on the side of using more complicated models compared to simpler ones.

Explanation and **prediction** are two **very different** model building objectives. Given a large set of possible explanatory variables, the best model for one is rarely the best model for the other.

12.2.2 Increase in variance (loss of precision)

The variability in estimating a regression coefficient of interest or in predicting a mean or future response may be increased if too many explanatory variables are included in the model. Correlations among those variables, multicollinearity, may further increase variance.

The **Variance Inflation Factor (VIF)** can be used to assess the degree of multicollinearity:

$$VIF = \frac{1}{(1 - R_X^2)}$$

where R_X^2 is the proportion of the variation in an explanatory variable X that is explained by its linear relationship to other explanatory variables in the model.

\sqrt{VIF} directly tells how much the SE of a regression coefficient is increased due to multicollinearity because

$$SE(\hat{\beta}_j) = \frac{\hat{\sigma}}{SD(X_j)\sqrt{n-1}} \times \sqrt{VIF}$$

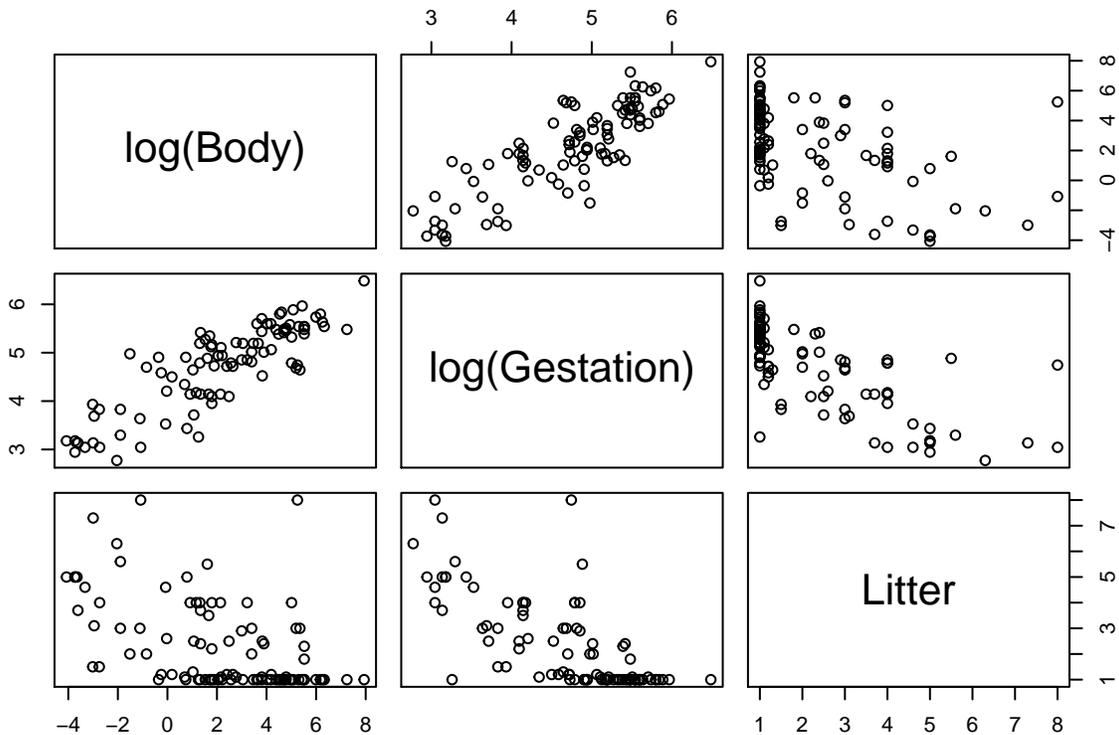
- If X has no association with the other explanatory variables, then $R_X^2 \approx 0$ and $VIF \approx 1$, which is the smallest possible value for VIF .
- Larger VIFs indicate greater multicollinearity. $VIF = 9$ for $\hat{\beta}_j$ indicates that $SE(\hat{\beta}_j)$ is 3 times larger than if X_j was independent of the other predictors; $VIF = 100$ (LARGE!) for $\hat{\beta}_j$ indicates that $SE(\hat{\beta}_j)$ is 10 times larger than if X_j was independent of the other predictors.
- Models with interactions tend to show a high VIF for the main effects and interactions because of expected multicollinearity of the main effects with the interactions terms.
- In R, calculate VIF using `vif()` in package `car`.

EXAMPLE: Consider Brain size as a function of Body size, Gestation length and Litter size from Chapter 9 notes. Consider some diagnostics of multicollinearity of the explanatory variables.

```
library(Sleuth3)
library(car)

# Best model from Chapter 9
m.red = lm(log(Brain) ~ log(Body) + log(Gestation) + Litter, data=case0902)

# Visually assess correlations
pairs(log(Body) ~ log(Gestation) + Litter, data=case0902)
```



```

# Here are the pairwise correlations. Does multicollinearity appear to be an issue?
#cor(log(case0902[,-1])) # shows correlations with log-Brain
cor(log(case0902[,-c(1,2)]))

##           Body  Gestation  Litter
## Body      1.0000000  0.8455317 -0.5457434
## Gestation  0.8455317  1.0000000 -0.7583399
## Litter     -0.5457434 -0.7583399  1.0000000

# Calculate the variance inflation factor for each explanatory variable in the model.
vif(m.red)

##      log(Body) log(Gestation)      Litter
##      3.793184      5.914154      2.290389

# Note the large VIFs of the interaction term
m.full = lm(log(Brain) ~ log(Body)*log(Gestation) + Litter, data=case0902)
vif(m.full)

##           log(Body)           log(Gestation)           Litter
##           30.893840           6.287916           2.354973
## log(Body):log(Gestation)
##           32.336053

```

- (a) From the scatterplots, which explanatory variables appear to be correlated?

- (b) From the pairwise correlations, which explanatory variables appear to be correlated?

- (c) From VIFs, which explanatory variables appear to be correlated?

- (d) Does multicollinearity appear to be an issue?

12.2.3 A Strategy for Dealing with Many Explanatory Variables

1. Identify the key objective (see section 12.2.1).
2. Decide which variables are sensitive to the objective. This is not a statistical decision!
3. Examine graphical displays (e.g., boxplots, scatterplots, matrix plots, and/or Trellis plots).
4. Perform transformations as necessary.
5. Fit a rich tentative model, examine residual plots, consider outliers and other unusual data, consider influential data, consider VIFs.

6. Find a model that includes a subset of explanatory variables that are sensitive to questions of interest and that is “almost as good” as the rich model. If there are not many explanatory variables, this can be done “by hand” as we have done up until now. If there are a lot of explanatory variables, then you may need to use a computer assisted variable selection technique (see section 12.3-12.4).
7. Proceed with the analysis, using the selected explanatory variables.

12.3-12.4 Sequential and all Subsets Variable-Selection Techniques

Your book introduces 4 automatic variable selection techniques:

1. forward selection,
2. backward elimination,
3. step-wise regression (which combines forward selection and backward elimination), and
4. a comparison of all possible subsets of models based on the explanatory variables. In R, use `regsubsets()` in package `leaps`. R returns the “best” model with highest R^2 with 1 explanatory variable; “best” model with 2 explanatory variables; ... “best” model with all explanatory variables.

The first 3 methods can produce misleading results, although some fields of science (e.g., engineering) still like to use them.

EXAMPLE: Consider Brain size as a function of Body size, Gestation length and Litter size from Chapter 9 notes.

```
library(leaps)

## Warning: package 'leaps' was built under R version 3.4.3
m.ss = regsubsets(log(Brain) ~ log(Body) + log(Gestation) + Litter, data=case0902)
summary(m.ss)$outmat
```

```
##           log(Body) log(Gestation) Litter
## 1  ( 1 ) "*"           " "           " "
## 2  ( 1 ) "*"           "*"           " "
## 3  ( 1 ) "*"           "*"           "*"

```

- (a) What is the best single predictor of Brain size?

- (b) What is the best two-predictor model of Brain size?

12.4.2 Aikake (AIC) and Bayesian (BIC) information criteria

We have already used R^2 as a way to compare models and select a model as being “best”. As the Residual sum of squares (SSR) decreases, then

$$R^2 = 1 - \frac{SSR}{SST}$$

always increases. In other words, using the R^2 to pick a model among a set of models always chooses the one with the smallest SSR .

We would like a model selection tool that penalizes a model for including too many variables. However, for models with normal errors like we have been fitting in this class (ANOVA and regression), R^2 is still probably the most common way to compare models.

The extra sum of squares test is one approach for seeing whether the change in R^2 when comparing a simple model to a more complicated model is worth the added complexity.

AIC and BIC are also popular as model comparison tools, especially in the ecological sciences. Unlike R^2 , AIC and BIC do penalize a model for including too many terms that are not useful. For models with normal errors like we have been fitting in this class (ANOVA and regression), like R^2 , AIC and BIC are also functions of the Residual sum of squares¹ (compare with Display 12.8).

$$AIC = n \ln \left(\frac{SSR}{n} \right) + 2(p + 1)$$

$$BIC = n \ln \left(\frac{SSR}{n} \right) + \ln(n)(p + 1)$$

- n is the sample size
- $p + 1$ is the number of coefficients in the model $(\beta_0, \beta_1, \dots, \beta_p)$
- SSR is the residual sum of squares.
- For AIC and BIC, the first term to the right of the equals sign is a measure of the benefit and the second term is the penalty.
- The smaller the AIC or BIC the “better” the model.
- Both AIC and BIC evaluate the cost and benefit of adding variables. Mathematically they are quite similar differing only in their penalty terms, but they are based on very different theoretical considerations.
- AIC is derived using *information theory*. The difference in AICs between two models is the expected difference in information loss when using the models to represent the true process that generated the data.
- BIC is derived using *Bayesian* statistical thinking (see sections 12.5).

Using AIC and BIC for model selection:

- AIC and BIC are relative measures. The model with the smallest AIC or BIC in a set of models is considered “best”. This is why a rich full model is generally included - in order to compare other simpler models to it.
- It is assumed that the rich model fits the data well. You can check this assumption via residual plots, etc.
- Calculate AIC in R, using the formula above, using `extractAIC()`. You can also use `AIC()`, but this value is off by a constant from `extractAIC()`. However, the difference of two models using `extractAIC()` is the same as the difference between two models using `AIC()`. Because only differences in AICs matter when selecting a model among many models, then in practice it does not really matter which R function you use.
- Calculate BIC in R, using the formula above, using `extractAIC(...,k=log(n))`. You can also use `BIC()`, but this value is off by a constant from `extractAIC(...,k=log(n))`. However, the difference of two models using `extractAIC(...,k=log(n))` is the same as the difference between two models using `BIC()`. Because only differences in AICs matter when selecting a model among many models, then in practice it does not really matter which R function you use.

¹for other types of models (e.g., Poisson and logistic regression), AIC and BIC are based on the *likelihood* function that you will learn about in STAT422

EXAMPLE: Consider Brain size as a function of Body size, Gestation length and Litter size from Chapter 9 notes.

```
library(Sleuth3)
m.red = lm(log(Brain) ~ log(Body) + log(Gestation) + Litter,data=case0902)
m.full = lm(log(Brain) ~ log(Body)*log(Gestation) + Litter,data=case0902)
```

```
AIC(m.red,m.full)
```

```
##      df      AIC
## m.red  5 135.6430
## m.full  6 135.5989
```

```
BIC(m.red,m.full)
```

```
##      df      BIC
## m.red  5 148.4648
## m.full  6 150.9850
```

```
anova(m.red,m.full)
```

```
## Analysis of Variance Table
##
## Model 1: log(Brain) ~ log(Body) + log(Gestation) + Litter
## Model 2: log(Brain) ~ log(Body) * log(Gestation) + Litter
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      92  20.806
## 2      91  20.368  1   0.43834 1.9584 0.1651
```

QUESTIONS:

- Objective I: Adjusting:** For Brain size as a function of Body size, Gestation length and Litter size from Chapter 9 notes.
 - Which of AIC, BIC or extra sum of squares test prefer the simpler reduced model with no interaction?

- Objective II: Explaining** Consider Exercise 20 on page 378. It is known that the Total number of species on an island is related to the island's Area. What other variables are also related to Total number of species after accounting for Area? The data are available via the Sleuth3 package in the data frame ex1220. Follow the steps outlined in section 12.2.3.

- In step #2, include all 6 explanatory variables.
- In step #3, a matrix plot is crucial (as usual) to determine transforms.

```
# `Look' at data
summary(ex1220)
```

```
##           Island      Total      Native      Area
## Baltra      : 1  Min.   :  2.00  Min.   :  1.00  Min.   :  0.010
## Bartolome   : 1  1st Qu.: 13.00  1st Qu.:  8.25  1st Qu.:  0.398
## Caldwell    : 1  Median : 42.00  Median :18.00  Median :  2.590
## Champion    : 1  Mean   : 85.23  Mean   :26.50  Mean   : 261.742
## Coamano     : 1  3rd Qu.: 96.00  3rd Qu.:32.25  3rd Qu.: 59.237
## Daphne Major: 1  Max.   :444.00  Max.   :95.00  Max.   :4669.320
## (Other)     :24
##           Elev      DistNear      DistSc      AreaNear
## Min.      : 25.0  Min.      : 0.20  Min.      :  0.00  Min.      :  0.03
```

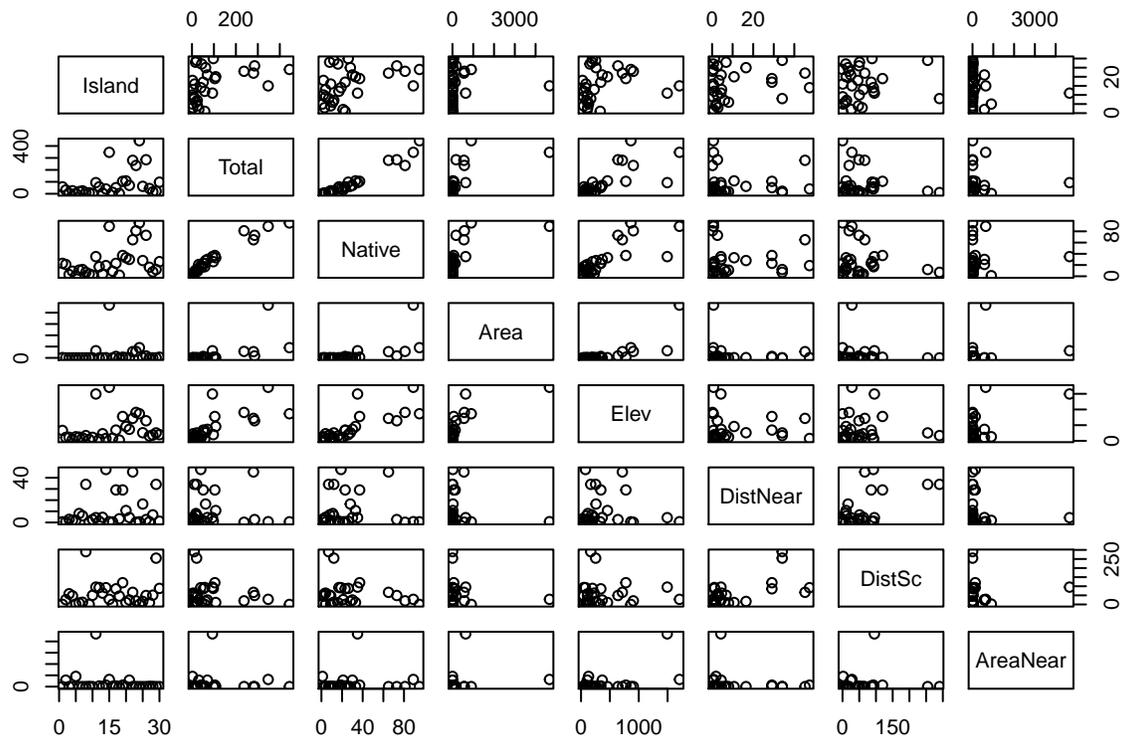
```
## 1st Qu.: 109.8 1st Qu.: 0.80 1st Qu.: 11.03 1st Qu.: 0.52
## Median : 192.0 Median : 3.05 Median : 46.65 Median : 2.59
## Mean : 368.4 Mean : 10.06 Mean : 56.98 Mean : 261.10
## 3rd Qu.: 435.2 3rd Qu.: 10.03 3rd Qu.: 81.08 3rd Qu.: 59.24
## Max. : 1707.0 Max. : 47.40 Max. : 290.20 Max. : 4669.32
##
```

```
dim(ex1220)
```

```
## [1] 30 8
```

```
# Matrix plot of raw data
```

```
pairs(ex1220)
```

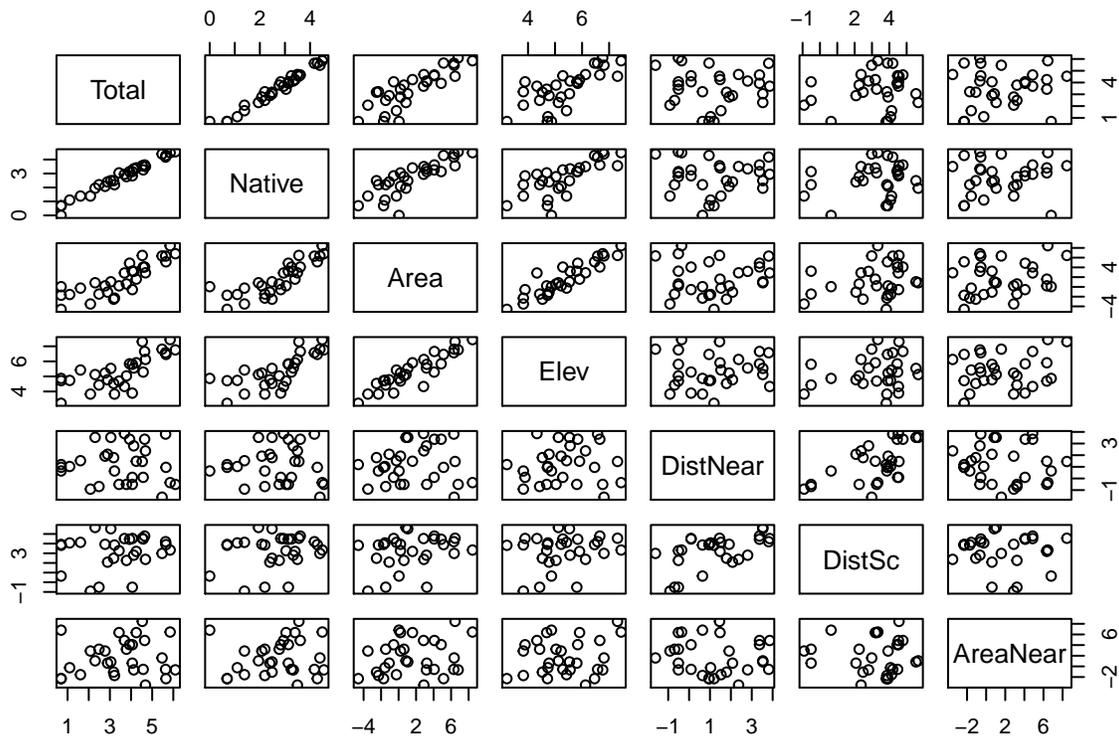


COMMENT: The matrix plot suggests that we consider log-transforms of the explanatory and response variables.

(c) In step #4, a quick way to look at log-transforms of all the variables is `pairs(log(ex1220[, -1]))`

```
# Matrix plot of log-transformed variables
```

```
pairs(log(ex1220[, -1]))
```



COMMENT: The matrix plot of the log-transformed variables suggests pairwise linear relationships of log-Total with the log-transformed explanatory variables. At this point (before we fit a model and inspect residual plots) no further transforms are needed.

(d) In step #5, consider the model, `m.int`, with all 2-way interactions among the 6 explanatory variables as the rich tentative model using the syntax (from p.15 of Chapter 11 notes)

```
# Rich tentative model with all 2-way interactions
m.int = lm(log(Total) ~ (log(Native) + log(Area) + log(Elev) + log(DistNear)
                    + log(I(DistSc+.01)) + log(AreaNear))^2,data=ex1220)
summary(m.int)
```

```
##
## Call:
## lm(formula = log(Total) ~ (log(Native) + log(Area) + log(Elev) +
##   log(DistNear) + log(I(DistSc + 0.01)) + log(AreaNear))^2,
##   data = ex1220)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.219375 -0.066358 -0.000591  0.060158  0.309110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -8.337914    7.091525  -1.176   0.273
## log(Native)    2.132993    1.361625   1.567   0.156
## log(Area)     -0.660999    0.442163  -1.495   0.173
```

```

## log(Elev)                1.837101    1.509881    1.217    0.258
## log(DistNear)           -1.030304    0.642900   -1.603    0.148
## log(I(DistSc + 0.01))   1.720880    1.086667    1.584    0.152
## log(AreaNear)           -0.008274    0.476814   -0.017    0.987
## log(Native):log(Area)    0.009938    0.164503    0.060    0.953
## log(Native):log(Elev)   -0.238663    0.289006   -0.826    0.433
## log(Native):log(DistNear) 0.209799    0.242608    0.865    0.412
## log(Native):log(I(DistSc + 0.01)) 0.049366    0.161010    0.307    0.767
## log(Native):log(AreaNear) -0.037856    0.044321   -0.854    0.418
## log(Area):log(Elev)     0.079158    0.129056    0.613    0.557
## log(Area):log(DistNear) -0.056264    0.066724   -0.843    0.424
## log(Area):log(I(DistSc + 0.01)) 0.076832    0.061465    1.250    0.247
## log(Area):log(AreaNear) -0.005419    0.026154   -0.207    0.841
## log(Elev):log(DistNear)  0.057093    0.125048    0.457    0.660
## log(Elev):log(I(DistSc + 0.01)) -0.364164    0.251018   -1.451    0.185
## log(Elev):log(AreaNear)  0.019426    0.079667    0.244    0.813
## log(DistNear):log(I(DistSc + 0.01)) 0.041258    0.041085    1.004    0.345
## log(DistNear):log(AreaNear) 0.009510    0.025911    0.367    0.723
## log(I(DistSc + 0.01)):log(AreaNear) -0.001851    0.038652   -0.048    0.963
##
## Residual standard error: 0.2225 on 8 degrees of freedom
## Multiple R-squared:  0.9944, Adjusted R-squared:  0.9798
## F-statistic: 67.82 on 21 and 8 DF,  p-value: 7.92e-07

```

COMMENT: A small value of 0.01 is added to `DistSc` because it has a 0, and `log(0)` is undefined, causing `lm()` to err out. There are better ways to deal with zeros: the simplest is to determine the limit of detection (LOD) of the process used to measure `DistSc` and substitute in LOD or LOD/2 for the zero. More sophisticated approaches, necessary when there are many zeros, are also available in R (see p. 3 of Chapter 3 notes).

Now let's check the assumptions of the model as well as the case influence statistics.

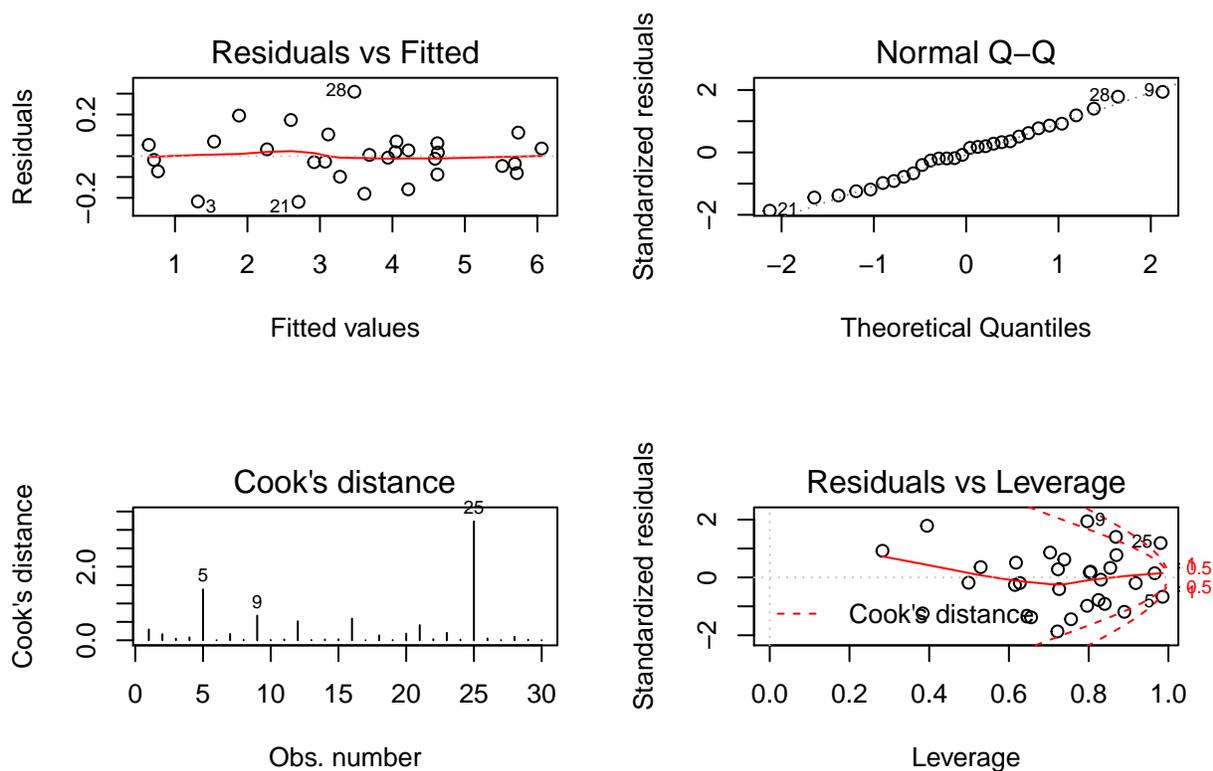
```

# Check assumptions of rich tentative model, inc. case influence statistics
par(mfrow=c(2,2))
plot(m.int,which=c(1,2,4,5))

```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



COMMENT: Linearity and normality and constant variance look OK. With more time we would scrutinize cases 5 and 25 (large Cook's Distance and large leverage) and case 28 (large residual).

- (e) In step #6, in the interest of time, assume that there are no influential data to consider. Instead, focus your time on a refined model with all 6 explanatory variables and no interactions, `m.add`. Perform an extra sum of squares test, compare AICs (`AIC(m.add,m.int)`) and BICs (`BIC(m.add,m.int)`). Which measures prefer `m.add` as a better model?

```
# Model with no interactions
m.add = lm(log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear) + log(I(DistSc+.01)) + log(AreaNear))

# Check VIFs.
vif(m.add)

##          log(Native)          log(Area)          log(Elev)
##          3.478413          9.140996          5.803418
##          log(DistNear) log(I(DistSc + 0.01))          log(AreaNear)
##          1.534952          1.482812          1.126777

anova(m.add,m.int)

## Analysis of Variance Table
##
## Model 1: log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear) +
##          log(I(DistSc + 0.01)) + log(AreaNear)
## Model 2: log(Total) ~ (log(Native) + log(Area) + log(Elev) + log(DistNear) +
##          log(I(DistSc + 0.01)) + log(AreaNear))^2
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
```

```
## 1      23 1.70039
## 2       8 0.39595 15      1.3044 1.7571 0.2126
```

```
AIC(m.add,m.int)
```

```
##      df      AIC
## m.add 8 15.026117
## m.int 23 1.306018
```

```
BIC(m.add,m.int)
```

```
##      df      BIC
## m.add 8 26.23570
## m.int 23 33.53356
```

ANSWER: BIC (=26.2) and extra sum of squares test ($p=0.2126$) prefer the simpler reduced model (m.add).

- (e) Still in step #6, perform an all subsets regression with m.add as the full rich model (i.e., m.add has all 6 explanatory variables and no interactions).

```
library(leaps)
m.ss = regsubsets(log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear) + log(I(DistSc+.01)))
summary(m.ss)$outmat
```

```
##      log(Native) log(Area) log(Elev) log(DistNear)
## 1 ( 1 ) "*"      " "      " "      " "
## 2 ( 1 ) "*"      " "      " "      " "
## 3 ( 1 ) "*"      "*"      "*"      " "
## 4 ( 1 ) "*"      "*"      "*"      " "
## 5 ( 1 ) "*"      "*"      "*"      "*"
## 6 ( 1 ) "*"      "*"      "*"      "*"
##      log(I(DistSc + 0.01)) log(AreaNear)
## 1 ( 1 ) " "                " "
## 2 ( 1 ) "*"                " "
## 3 ( 1 ) " "                " "
## 4 ( 1 ) "*"                " "
## 5 ( 1 ) "*"                " "
## 6 ( 1 ) "*"                "*"

```

COMMENT: R shows that the “best” 6 predictor (7 parameter) model corresponds to m.add:

$$\mu(\ln(Total)|\dots) = \beta_0 + \beta_1 \ln(Native) + \beta_2 \ln(Area) + \beta_3 \ln(Elev) + \beta_4 \ln(DistNear) + \beta_5 \ln(DistSc) + \beta_6 \ln(AreaNear).$$

The “best” 5 predictor (6 parameter model) (fit as model m5 below) is:

$$\mu(\ln(Total)|\dots) = \beta_0 + \beta_1 \ln(Native) + \beta_2 \ln(Area) + \beta_3 \ln(Elev) + \beta_4 \ln(DistNear) + \beta_5 \ln(DistSc).$$

The “best” 4 predictor (5 parameter) model (fit as model m4 below) is:

$$\mu(\ln(Total)|\dots) = \beta_0 + \beta_1 \ln(Native) + \beta_2 \ln(Area) + \beta_3 \ln(Elev) + \beta_4 \ln(DistSc).$$

The “best” 3 predictor (4 parameter) model (fit as model m3 below) is:

$$\mu(\ln(Total)|\dots) = \beta_0 + \beta_1 \ln(Native) + \beta_2 \ln(Area) + \beta_3 \ln(Elev).$$

- (f) Compare the “best model” with 3 explanatory variables (m3) to the “best model” with 4 explanatory variables (m4) to the “best model” with 5 explanatory variables (m5) to the “best model” with 6 explanatory variables (m.add). Which model is “best” according to AIC? Which is “best” according to BIC?

```

# Fit the models in R and store them in objects m6, m5, m4, m3
# so that the models can be compared using anova(), AIC(), BIC()
m6=m.add
m5 = lm(log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear)
      + log(I(DistSc+.01)),data=ex1220)
m4 = lm(log(Total) ~ log(Native) + log(Area) + log(Elev) + log(I(DistSc+.01)),
      data=ex1220)
m3 = lm(log(Total) ~ log(Native) + log(Area) + log(Elev),data=ex1220)

# F-test comparing the simplest to the most complicated
anova(m3,m6)

```

```

## Analysis of Variance Table
##
## Model 1: log(Total) ~ log(Native) + log(Area) + log(Elev)
## Model 2: log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear) +
##   log(I(DistSc + 0.01)) + log(AreaNear)
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      26 1.8393
## 2      23 1.7004  3  0.13889 0.6262 0.6054

```

```

# F-tests sequentially comparing each model to the next most complicated
anova(m3,m4,m5,m6)

```

```

## Analysis of Variance Table
##
## Model 1: log(Total) ~ log(Native) + log(Area) + log(Elev)
## Model 2: log(Total) ~ log(Native) + log(Area) + log(Elev) + log(I(DistSc +
##   0.01))
## Model 3: log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear) +
##   log(I(DistSc + 0.01))
## Model 4: log(Total) ~ log(Native) + log(Area) + log(Elev) + log(DistNear) +
##   log(I(DistSc + 0.01)) + log(AreaNear)
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      26 1.8393
## 2      25 1.7140  1  0.125248 1.6941 0.2059
## 3      24 1.7081  1  0.005935 0.0803 0.7795
## 4      23 1.7004  1  0.007705 0.1042 0.7497

```

```

AIC(m3,m4,m5,m6)

```

```

##   df      AIC
## m3  5 11.38159
## m4  6 11.26581
## m5  7 13.16176
## m6  8 15.02612

```

```

BIC(m3,m4,m5,m6)

```

```

##   df      BIC
## m3  5 18.38757
## m4  6 19.67299
## m5  7 22.97014
## m6  8 26.23570

```

ANSWER: BIC (=18.4) and extra sum of squares test ($p \geq 0.2059$) prefer the simpler reduced model (m3).

$$\mu(\ln(Total)|\dots) = \beta_0 + \beta_1 \ln(Native) + \beta_2 \ln(Area) + \beta_3 \ln(Elev).$$