Lab 10

Regression selection and interpretation with airquality data

Load in the airquality dataset, which has daily air quality measurements in New York, from May to September 1973. You can find more information on each variable by typing ?airquality

```
# Load car package
require(car)

# Load airquality data frame from datasets package
airquality <- datasets::airquality

# Subset to only complete cases and name new dataset air
air <- na.omit(airquality)
```

Build a model with just temperature (Temp) predicting ozone.

```
model_temp <- lm(Ozone ~ Temp, data = air)
summary(model_temp)
```

```
##
## Call:
## lm(formula = Ozone ~ Temp, data = air)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.922 -17.459  -0.874  10.444 118.078
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -147.6461    18.7553  -7.872 2.76e-12 ***
## Temp           2.4391     0.2393  10.192  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.92 on 109 degrees of freedom
## Multiple R-squared:  0.488,  Adjusted R-squared:  0.4833
## F-statistic: 103.9 on 1 and 109 DF,  p-value: < 2.2e-16
```

Now, build an additive linear model predicting Ozone from Solar.R, Wind, Temp, Month, and Day.

```
model_full <- lm(Ozone ~ Solar.R + Wind + Temp + Month + Day, data = air)
summary(model_full)
```

```
##
## Call:
## lm(formula = Ozone ~ Solar.R + Wind + Temp + Month + Day, data = air)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.014 -12.284  -3.302   8.454  95.348
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -64.11632   23.48249  -2.730  0.00742 **
## Solar.R       0.05027    0.02342   2.147  0.03411 *
## Wind         -3.31844    0.64451  -5.149 1.23e-06 ***
## Temp          1.89579    0.27389   6.922 3.66e-10 ***
## Month        -3.03996    1.51346  -2.009  0.04714 *
## Day           0.27388    0.22967   1.192  0.23576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.86 on 105 degrees of freedom
## Multiple R-squared:  0.6249, Adjusted R-squared:  0.6071
## F-statistic: 34.99 on 5 and 105 DF,  p-value: < 2.2e-16
```

How did the coefficient on Temp change?

Why do you think this change occured?

In the full model, how do we interpret the coefficient on Temp?

In the original model, how do we interpret the coefficient on Temp?

Regression with standardized coefficients

Let's use the function lm.beta (from the lm.beta package) to get the standardized coefficients of the full model.

```
library(lm.beta) # Install if you don't already have this package
lm.beta(model_full)
```

```
##
## Call:
## lm(formula = Ozone ~ Solar.R + Wind + Temp + Month + Day, data = air)
##
## Standardized Coefficients::
## (Intercept)     Solar.R        Wind        Temp       Month         Day
##  0.00000000  0.13771560 -0.35479273  0.54293794 -0.13460690  0.07166447
```

How do we interpret these coefficients?

How is this helpful to us?

Which variable has the largest effect on Ozone?

Model selection methods

Up to this point, you have learned to do an F test comparing nested models using the anova command in R. Below, we will discuss some other methods for model selection. But first, let's build a bunch of potential models!

```
m1 <- lm(Ozone ~ Temp, data = air)
m2 <- lm(Ozone ~ Wind + Temp, data = air)
m3 <- lm(Ozone ~ Wind + Temp + Solar.R, data = air)
m4 <- lm(Ozone ~ Wind + Temp + Solar.R + Month, data = air)
m5 <- lm(Ozone ~ Wind + Temp + Solar.R + Month + Day, data = air)
```

```r
m6 <- lm(Ozone ~ Wind * Temp + Solar.R + Month + Day, data = air)
m7 <- lm(Ozone ~ Wind * Temp + Wind * Solar.R + Month + Day, data = air)
m8 <- lm(Ozone ~ Wind * Temp + Wind * Solar.R + Temp * Solar.R + Month + Day, data = air)
m9 <- lm(Ozone ~ Wind * Temp * Solar.R + Month + Day, data = air)
m10 <- lm(Ozone ~ Wind * Temp * Solar.R * Month * Day, data = air)
```

You've learned about R squared as a measurement of the proportion of variation in the outcome variable which can be explained by the model. You might say that a model is better if it explains more variation Y. Let's first compare the model R squares of all of these models:

```r
# save the models
models = list(m1, m2, m3, m4, m5, m6, m7, m8, m9, m10)

# how many predictors in each model?
ps = unlist(lapply(models, function(x) {nrow(summary(x)$coef)-1}))
ps
```

```
##  [1]  1  2  3  4  5  6  7  8  9 31
```

```r
# r squares of each model:
rsquares = unlist(lapply(models, function(x) {summary(x)$r.squared}))
rsquares
```

```
##  [1] 0.4879601 0.5813780 0.6058946 0.6198614 0.6249408 0.6826408 0.6901348
##  [8] 0.6971634 0.6986896 0.7383509
```

```r
which.max(rsquares) #which is largest
```

```
## [1] 10
```

What do you notice about the R squares? Do they ever decrease after adding another predictor?

where are the biggest jumps in R squares; where are there negligible differences?

Which model would you choose?

Now, let's compute the adjusted R squares. Adjusted R squared is similar to R squared, except that there is a penalty for adding additional parameters. Therefore, if adding a parameter does not increase the R squared sufficiently enough to "warrant" another loss of degrees freedom, adjusted R squared could decrease:

```r
# r squares of each model:
adj.rsquares = unlist(lapply(models, function(x) {summary(x)$adj.r.squared}))
adj.rsquares
```

```
##  [1] 0.4832625 0.5736257 0.5948449 0.6055166 0.6070808 0.6643316 0.6690760
##  [8] 0.6734116 0.6718402 0.6356785
```

```r
which.max(adj.rsquares) #which is largest?
```

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

How are the adjusted R squares different from or similar to the regular R squares?

Which model would you choose using adjusted R squares?

Log likelihood is another potential criterion we could use. Likelihood is essentially the probability of a set of data given a particular model. Log likelihood is exactly as it sounds: log of the likelihood (note: as likelihood increases, log likelihood also increases). Thus, sometimes people do model selection (or parameter estimation) by trying to maximize likelhood or log likelihood. Let's calculate log likelihood of each model:

```
require(stats) #load stats package

# r squares of each model:
loglikes = unlist(lapply(models, logLik))
loglikes
```

```
##  [1] -508.8876 -497.7080 -494.3586 -492.3560 -491.6094 -482.3381 -481.0118
##  [8] -479.7385 -479.4581 -471.6250
```

```
which.max(loglikes) #which is largest
```

```
## [1] 10
```

Does log likelihood every decrease when we add more parameters?

Which model would you choose using log likelihood?

Similar to how adjusted R squared penalizes the R squared for additional parameters, we can also use criteria that penalize the log likelihood for adding additional parameters. The most commonly used are Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). $AIC = -2 * logL + 2 * p$, $BIC = -2 * logL + p * log(n)$ where LogL is log likelihood, p is the number of parameters, and n is the sample size. By multiplying LogL by a negative number and then adding for each additional parameter, we make it so that smaller AIC or BIC indicates higher log likelihood and/or fewer parameters. BIC should only be used when n is much larger than p; thus, BIC is generally more conservative than AIC (i.e., we penalize each additional parameter more in BIC than AIC).

Let's use BIC and AIC to choose models:

```
#AIC
AICs = unlist(lapply(models, AIC))
AICs
```

```
##  [1] 1023.7751 1003.4160  998.7171  996.7119  997.2188  980.6763  980.0237
##  [8]  979.4769  980.9161 1009.2499
```

```
which.min(AICs) #which is smallest?
```

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

```
#BIC
BICs = unlist(lapply(models, BIC))
BICs
```

```
##  [1] 1031.904 1014.254 1012.265 1012.969 1016.185 1002.353 1004.409
##  [8] 1006.572 1010.721 1098.664
```

```
which.min(BICs) #which is smallest?
```

```
## [1] 6
```

Which model does AIC choose?

Which model does BIC choose?

What about all of the other possible models between 1 and 10 which we didn't compare? We can also use a stepwise procedure to ask R to compare all possible models up to some largest possible model, using a particular criterion (i.e., AIC or BIC). We can work step-wise starting from the smallest model and adding additional terms one at a time, or we can start with the largest possible model and delete terms one at a time. There's also an option to move in "both" directions, adding or subtracting terms as we go.

```
require(MASS)
```

```r
#forward:
summary(stepAIC(m1, list(lower=m1, upper=m10), direction = "forward"))
```

```
## Start:  AIC=706.77
## Ozone ~ Temp
##
##           Df Sum of Sq   RSS    AIC
## + Wind     1   11378.5 50989 686.41
## + Month    1    2824.7 59543 703.63
## + Solar.R  1    2723.1 59644 703.82
## <none>                 62367 706.77
## + Day      1     476.5 61891 707.92
##
## Step:  AIC=686.41
## Ozone ~ Temp + Wind
##
##             Df Sum of Sq   RSS    AIC
## + Wind:Temp  1    6594.8 44394 673.04
## + Solar.R    1    2986.2 48003 681.71
## + Month      1    2734.8 48254 682.29
## <none>                   50989 686.41
## + Day        1     486.6 50502 687.35
##
## Step:  AIC=673.04
## Ozone ~ Temp + Wind + Temp:Wind
##
##           Df Sum of Sq   RSS    AIC
## + Solar.R  1    3618.4 40776 665.60
## + Month    1    2461.0 41933 668.71
## <none>                 44394 673.04
## + Day      1     593.1 43801 673.55
##
## Step:  AIC=665.6
## Ozone ~ Temp + Wind + Solar.R + Temp:Wind
##
##                Df Sum of Sq   RSS    AIC
## + Temp:Solar.R  1   2141.06 38635 661.61
## + Month         1   1375.95 39400 663.79
## + Wind:Solar.R  1    995.72 39780 664.86
## <none>                      40776 665.60
## + Day           1    694.02 40082 665.70
##
## Step:  AIC=661.61
## Ozone ~ Temp + Wind + Solar.R + Temp:Wind + Temp:Solar.R
##
##                Df Sum of Sq   RSS    AIC
## + Month         1    818.67 37816 661.24
## <none>                      38635 661.61
## + Wind:Solar.R  1    429.42 38205 662.37
## + Day           1    379.70 38255 662.52
##
## Step:  AIC=661.24
## Ozone ~ Temp + Wind + Solar.R + Month + Temp:Wind + Temp:Solar.R
##
```

```
##                 Df Sum of Sq   RSS    AIC
## <none>                       37816 661.24
## + Day            1    447.35 37369 661.92
## + Wind:Solar.R   1    440.48 37376 661.94
## + Wind:Month     1    352.27 37464 662.20
## + Temp:Month     1     15.14 37801 663.19
## + Solar.R:Month  1      0.64 37815 663.23
```

```
##
## Call:
## lm(formula = Ozone ~ Temp + Wind + Solar.R + Month + Temp:Wind +
##     Temp:Solar.R, data = air)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.192 -12.754  -2.597   7.762  95.229
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.426e+02  6.388e+01  -2.233  0.02772 *
## Temp          2.748e+00  8.436e-01   3.257  0.00152 **
## Wind          1.126e+01  4.234e+00   2.658  0.00910 **
## Solar.R      -3.088e-01  1.764e-01  -1.750  0.08300 .
## Month        -2.119e+00  1.412e+00  -1.500  0.13652
## Temp:Wind    -1.876e-01  5.394e-02  -3.478  0.00074 ***
## Temp:Solar.R  5.014e-03  2.402e-03   2.087  0.03933 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.07 on 104 degrees of freedom
## Multiple R-squared:  0.6895, Adjusted R-squared:  0.6716
## F-statistic:  38.5 on 6 and 104 DF,  p-value: < 2.2e-16
```

```
#backward:
summary(stepAIC(m10, list(lower=m1, upper=m10), direction = "backward"))
```

```
## Start:  AIC=692.25
## Ozone ~ Wind * Temp * Solar.R * Month * Day
##
##                                Df Sum of Sq   RSS    AIC
## - Wind:Temp:Solar.R:Month:Day  1    112.46 31982 690.64
## <none>                                     31869 692.25
##
## Step:  AIC=690.64
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Temp:Solar.R:Month +
##     Wind:Temp:Day + Wind:Solar.R:Day + Temp:Solar.R:Day + Wind:Month:Day +
##     Temp:Month:Day + Solar.R:Month:Day + Wind:Temp:Solar.R:Month +
##     Wind:Temp:Solar.R:Day + Wind:Temp:Month:Day + Wind:Solar.R:Month:Day +
##     Temp:Solar.R:Month:Day
##
##                            Df Sum of Sq   RSS    AIC
## - Temp:Solar.R:Month:Day    1      1.80 31984 688.64
```

```
## - Wind:Temp:Month:Day      1     49.73 32032 688.81
## - Wind:Temp:Solar.R:Month  1    398.12 32380 690.01
## - Wind:Solar.R:Month:Day   1    472.47 32454 690.26
## - Wind:Temp:Solar.R:Day    1    573.73 32556 690.61
## <none>                                 31982 690.64
##
## Step:  AIC=688.64
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Temp:Solar.R:Month +
##     Wind:Temp:Day + Wind:Solar.R:Day + Temp:Solar.R:Day + Wind:Month:Day +
##     Temp:Month:Day + Solar.R:Month:Day + Wind:Temp:Solar.R:Month +
##     Wind:Temp:Solar.R:Day + Wind:Temp:Month:Day + Wind:Solar.R:Month:Day
##
##                            Df Sum of Sq   RSS    AIC
## - Wind:Temp:Month:Day       1     49.30 32033 686.81
## - Wind:Temp:Solar.R:Month   1    415.27 32399 688.07
## - Wind:Solar.R:Month:Day    1    470.68 32454 688.26
## <none>                                 31984 688.64
## - Wind:Temp:Solar.R:Day     1    593.20 32577 688.68
##
## Step:  AIC=686.81
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Temp:Solar.R:Month +
##     Wind:Temp:Day + Wind:Solar.R:Day + Temp:Solar.R:Day + Wind:Month:Day +
##     Temp:Month:Day + Solar.R:Month:Day + Wind:Temp:Solar.R:Month +
##     Wind:Temp:Solar.R:Day + Wind:Solar.R:Month:Day
##
##                            Df Sum of Sq   RSS    AIC
## - Wind:Temp:Solar.R:Month   1    410.44 32443 686.23
## <none>                                 32033 686.81
## - Wind:Temp:Solar.R:Day     1    603.47 32636 686.89
## - Wind:Solar.R:Month:Day    1    692.95 32726 687.19
## - Temp:Month:Day            1   1584.89 33618 690.17
##
## Step:  AIC=686.23
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Temp:Solar.R:Month +
##     Wind:Temp:Day + Wind:Solar.R:Day + Temp:Solar.R:Day + Wind:Month:Day +
##     Temp:Month:Day + Solar.R:Month:Day + Wind:Temp:Solar.R:Day +
##     Wind:Solar.R:Month:Day
##
##                           Df Sum of Sq   RSS    AIC
## - Temp:Solar.R:Month       1      0.45 32444 684.23
## - Wind:Temp:Solar.R:Day    1    229.02 32672 685.01
## - Wind:Solar.R:Month:Day   1    287.85 32731 685.21
## - Wind:Temp:Month          1    416.22 32860 685.64
## <none>                                32443 686.23
## - Temp:Month:Day           1   1915.92 34359 690.60
```

```
##
## Step:  AIC=684.23
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Wind:Temp:Day + Wind:Solar.R:Day +
##     Temp:Solar.R:Day + Wind:Month:Day + Temp:Month:Day + Solar.R:Month:Day +
##     Wind:Temp:Solar.R:Day + Wind:Solar.R:Month:Day
##
##                           Df Sum of Sq   RSS    AIC
## - Wind:Temp:Solar.R:Day    1    230.09 32674 683.01
## - Wind:Solar.R:Month:Day   1    291.63 32735 683.22
## - Wind:Temp:Month          1    416.39 32860 683.64
## <none>                                 32444 684.23
## - Temp:Month:Day           1   1920.65 34364 688.61
##
## Step:  AIC=683.01
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Wind:Temp:Day + Wind:Solar.R:Day +
##     Temp:Solar.R:Day + Wind:Month:Day + Temp:Month:Day + Solar.R:Month:Day +
##     Wind:Solar.R:Month:Day
##
##                           Df Sum of Sq   RSS    AIC
## - Wind:Temp:Day            1      0.09 32674 681.01
## - Wind:Temp:Solar.R        1      5.45 32679 681.03
## - Wind:Solar.R:Month:Day   1    106.65 32781 681.37
## - Wind:Temp:Month          1    514.89 33189 682.75
## <none>                                 32674 683.01
## - Temp:Solar.R:Day         1    805.71 33480 683.72
## - Temp:Month:Day           1   1853.78 34528 687.14
##
## Step:  AIC=681.01
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Solar.R +
##     Wind:Temp:Month + Wind:Solar.R:Month + Wind:Solar.R:Day +
##     Temp:Solar.R:Day + Wind:Month:Day + Temp:Month:Day + Solar.R:Month:Day +
##     Wind:Solar.R:Month:Day
##
##                           Df Sum of Sq   RSS    AIC
## - Wind:Temp:Solar.R        1      6.92 32681 679.04
## - Wind:Solar.R:Month:Day   1    111.77 32786 679.39
## - Wind:Temp:Month          1    552.61 33227 680.87
## <none>                                 32674 681.01
## - Temp:Solar.R:Day         1    810.15 33484 681.73
## - Temp:Month:Day           1   2204.38 34878 686.26
##
## Step:  AIC=679.04
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Month +
##     Wind:Solar.R:Month + Wind:Solar.R:Day + Temp:Solar.R:Day +
```

```
##     Wind:Month:Day + Temp:Month:Day + Solar.R:Month:Day + Wind:Solar.R:Month:Day
##
##                           Df Sum of Sq   RSS    AIC
## - Wind:Solar.R:Month:Day  1    108.65 32790 677.41
## - Wind:Temp:Month         1    546.60 33228 678.88
## <none>                                32681 679.04
## - Temp:Solar.R:Day        1    815.23 33496 679.77
## - Temp:Month:Day          1   2208.95 34890 684.30
##
## Step:  AIC=677.41
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Month +
##     Wind:Solar.R:Month + Wind:Solar.R:Day + Temp:Solar.R:Day +
##     Wind:Month:Day + Temp:Month:Day + Solar.R:Month:Day
##
##                       Df Sum of Sq   RSS    AIC
## - Wind:Solar.R:Day     1    142.99 32933 675.89
## - Solar.R:Month:Day    1    390.95 33181 676.72
## - Wind:Temp:Month      1    475.83 33265 677.00
## <none>                             32790 677.41
## - Wind:Month:Day       1    721.98 33512 677.82
## - Wind:Solar.R:Month   1    763.06 33553 677.96
## - Temp:Solar.R:Day     1    770.47 33560 677.98
## - Temp:Month:Day       1   2224.16 35014 682.69
##
## Step:  AIC=675.89
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Month +
##     Wind:Solar.R:Month + Temp:Solar.R:Day + Wind:Month:Day +
##     Temp:Month:Day + Solar.R:Month:Day
##
##                       Df Sum of Sq   RSS    AIC
## - Solar.R:Month:Day    1    344.07 33277 675.04
## - Wind:Temp:Month      1    422.85 33355 675.30
## <none>                             32933 675.89
## - Wind:Month:Day       1    624.10 33557 675.97
## - Temp:Solar.R:Day     1    628.30 33561 675.99
## - Wind:Solar.R:Month   1    727.46 33660 676.31
## - Temp:Month:Day       1   2119.32 35052 680.81
##
## Step:  AIC=675.04
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Month +
##     Wind:Solar.R:Month + Temp:Solar.R:Day + Wind:Month:Day +
##     Temp:Month:Day
##
##                       Df Sum of Sq   RSS    AIC
## - Temp:Solar.R:Day     1    330.83 33607 674.14
## - Wind:Temp:Month      1    414.95 33692 674.42
## <none>                             33277 675.04
## - Wind:Solar.R:Month   1    614.80 33891 675.07
```

```
## - Wind:Month:Day      1    696.30 33973 675.34
## - Temp:Month:Day      1   1851.73 35128 679.05
##
## Step:  AIC=674.14
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Solar.R:Day + Month:Day + Wind:Temp:Month +
##     Wind:Solar.R:Month + Wind:Month:Day + Temp:Month:Day
##
##                       Df Sum of Sq   RSS    AIC
## - Solar.R:Day          1    101.25 33709 672.47
## - Wind:Solar.R:Month   1    510.20 34118 673.81
## - Temp:Solar.R         1    534.96 34142 673.89
## - Wind:Month:Day       1    578.45 34186 674.03
## <none>                             33607 674.14
## - Wind:Temp:Month      1    649.00 34256 674.26
## - Temp:Month:Day       1   1910.35 35518 678.28
##
## Step:  AIC=672.47
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Month:Day + Wind:Temp:Month + Wind:Solar.R:Month +
##     Wind:Month:Day + Temp:Month:Day
##
##                       Df Sum of Sq   RSS    AIC
## - Wind:Solar.R:Month   1    488.91 34198 672.07
## - Wind:Month:Day       1    586.45 34295 672.39
## - Wind:Temp:Month      1    599.09 34308 672.43
## <none>                             33709 672.47
## - Temp:Solar.R         1    660.36 34369 672.63
## - Temp:Month:Day       1   1809.45 35518 676.28
##
## Step:  AIC=672.07
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Solar.R:Month +
##     Wind:Day + Temp:Day + Month:Day + Wind:Temp:Month + Wind:Month:Day +
##     Temp:Month:Day
##
##                     Df Sum of Sq   RSS    AIC
## - Solar.R:Month      1     28.52 34226 670.16
## - Wind:Solar.R       1     64.66 34262 670.28
## - Wind:Temp:Month    1    215.21 34413 670.77
## <none>                           34198 672.07
## - Temp:Solar.R       1    773.50 34971 672.55
## - Wind:Month:Day     1   1179.72 35377 673.84
## - Temp:Month:Day     1   1658.96 35857 675.33
##
## Step:  AIC=670.16
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Wind:Solar.R +
##     Temp:Solar.R + Wind:Month + Temp:Month + Wind:Day + Temp:Day +
##     Month:Day + Wind:Temp:Month + Wind:Month:Day + Temp:Month:Day
##
##                     Df Sum of Sq   RSS    AIC
## - Wind:Solar.R       1     62.05 34288 668.37
```

```
## - Wind:Temp:Month  1    242.39 34469 668.95
## <none>                         34226 670.16
## - Temp:Solar.R     1   1105.83 35332 671.69
## - Wind:Month:Day   1   1159.12 35385 671.86
## - Temp:Month:Day   1   1631.58 35858 673.33
##
## Step:  AIC=668.37
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Temp:Solar.R +
##     Wind:Month + Temp:Month + Wind:Day + Temp:Day + Month:Day +
##     Wind:Temp:Month + Wind:Month:Day + Temp:Month:Day
##
##                   Df Sum of Sq   RSS    AIC
## - Wind:Temp:Month  1    356.69 34645 667.51
## <none>                         34288 668.37
## - Wind:Month:Day   1   1142.15 35430 670.00
## - Temp:Solar.R     1   1357.94 35646 670.68
## - Temp:Month:Day   1   1846.01 36134 672.19
##
## Step:  AIC=667.51
## Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp + Temp:Solar.R +
##     Wind:Month + Temp:Month + Wind:Day + Temp:Day + Month:Day +
##     Wind:Month:Day + Temp:Month:Day
##
##                  Df Sum of Sq   RSS    AIC
## <none>                        34645 667.51
## - Wind:Month:Day  1    1309.4 35954 669.63
## - Temp:Solar.R    1    1434.6 36079 670.02
## - Temp:Month:Day  1    1674.1 36319 670.75
## - Wind:Temp       1    4380.9 39026 678.73
##
## Call:
## lm(formula = Ozone ~ Wind + Temp + Solar.R + Month + Day + Wind:Temp +
##     Temp:Solar.R + Wind:Month + Temp:Month + Wind:Day + Temp:Day +
##     Month:Day + Wind:Month:Day + Temp:Month:Day, data = air)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.759 -10.498  -2.030   5.987  88.331
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    304.478188 248.737399   1.224 0.223912
## Wind             2.195286   6.750342   0.325 0.745729
## Temp            -1.352420   2.988939  -0.452 0.651948
## Solar.R         -0.335454   0.198192  -1.693 0.093779 .
## Month          -82.068372  33.688899  -2.436 0.016692 *
## Day            -26.609240  12.892793  -2.064 0.041727 *
## Wind:Temp       -0.247397   0.071007  -3.484 0.000746 ***
## Temp:Solar.R     0.005339   0.002678   1.994 0.049010 *
## Wind:Month       2.098168   0.965338   2.174 0.032202 *
## Temp:Month       0.758650   0.366820   2.068 0.041312 *
## Wind:Day         0.606439   0.361292   1.679 0.096496 .
## Temp:Day         0.276727   0.148298   1.866 0.065091 .
```

```
## Month:Day        4.309260   1.837109   2.346 0.021050 *
## Wind:Month:Day  -0.095871   0.050332  -1.905 0.059800 .
## Temp:Month:Day  -0.044022   0.020439  -2.154 0.033759 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19 on 96 degrees of freedom
## Multiple R-squared:  0.7156, Adjusted R-squared:  0.6741
## F-statistic: 17.25 on 14 and 96 DF,  p-value: < 2.2e-16
```

```
#both:
summary(stepAIC(m1, list(lower=m1, upper=m10), direction = "both"))
```

```
## Start:  AIC=706.77
## Ozone ~ Temp
##
##           Df Sum of Sq   RSS    AIC
## + Wind     1   11378.5 50989 686.41
## + Month    1    2824.7 59543 703.63
## + Solar.R  1    2723.1 59644 703.82
## <none>                 62367 706.77
## + Day      1     476.5 61891 707.92
##
## Step:  AIC=686.41
## Ozone ~ Temp + Wind
##
##             Df Sum of Sq   RSS    AIC
## + Wind:Temp  1    6594.8 44394 673.04
## + Solar.R    1    2986.2 48003 681.71
## + Month      1    2734.8 48254 682.29
## <none>                   50989 686.41
## + Day        1     486.6 50502 687.35
## - Wind       1   11378.5 62367 706.77
##
## Step:  AIC=673.04
## Ozone ~ Temp + Wind + Temp:Wind
##
##             Df Sum of Sq   RSS    AIC
## + Solar.R    1    3618.4 40776 665.60
## + Month      1    2461.0 41933 668.71
## <none>                   44394 673.04
## + Day        1     593.1 43801 673.55
## - Temp:Wind  1    6594.8 50989 686.41
##
## Step:  AIC=665.6
## Ozone ~ Temp + Wind + Solar.R + Temp:Wind
##
##                Df Sum of Sq   RSS    AIC
## + Temp:Solar.R  1    2141.1 38635 661.61
## + Month         1    1375.9 39400 663.79
## + Wind:Solar.R  1     995.7 39780 664.86
## <none>                      40776 665.60
## + Day           1     694.0 40082 665.70
## - Solar.R       1    3618.4 44394 673.04
## - Temp:Wind     1    7227.0 48003 681.71
```

```
##
## Step:  AIC=661.61
## Ozone ~ Temp + Wind + Solar.R + Temp:Wind + Temp:Solar.R
##
##                Df Sum of Sq   RSS    AIC
## + Month         1     818.7 37816 661.24
## <none>                      38635 661.61
## + Wind:Solar.R  1     429.4 38205 662.37
## + Day           1     379.7 38255 662.52
## - Temp:Solar.R  1    2141.1 40776 665.60
## - Temp:Wind     1    4339.8 42975 671.43
##
## Step:  AIC=661.24
## Ozone ~ Temp + Wind + Solar.R + Month + Temp:Wind + Temp:Solar.R
##
##                 Df Sum of Sq   RSS    AIC
## <none>                       37816 661.24
## - Month          1     818.7 38635 661.61
## + Day            1     447.3 37369 661.92
## + Wind:Solar.R   1     440.5 37376 661.94
## + Wind:Month     1     352.3 37464 662.20
## + Temp:Month     1      15.1 37801 663.19
## + Solar.R:Month  1       0.6 37815 663.23
## - Temp:Solar.R   1    1583.8 39400 663.79
## - Temp:Wind      1    4397.5 42214 671.45
##
## Call:
## lm(formula = Ozone ~ Temp + Wind + Solar.R + Month + Temp:Wind +
##     Temp:Solar.R, data = air)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.192 -12.754  -2.597   7.762  95.229
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.426e+02  6.388e+01  -2.233  0.02772 *
## Temp          2.748e+00  8.436e-01   3.257  0.00152 **
## Wind          1.126e+01  4.234e+00   2.658  0.00910 **
## Solar.R      -3.088e-01  1.764e-01  -1.750  0.08300 .
## Month        -2.119e+00  1.412e+00  -1.500  0.13652
## Temp:Wind    -1.876e-01  5.394e-02  -3.478  0.00074 ***
## Temp:Solar.R  5.014e-03  2.402e-03   2.087  0.03933 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.07 on 104 degrees of freedom
## Multiple R-squared:  0.6895, Adjusted R-squared:  0.6716
## F-statistic:  38.5 on 6 and 104 DF,  p-value: < 2.2e-16
```