

#Model Selection Simulation (For Illustration)

Simulate data where the null model is true. Use 500 reps with $n=50$ observations per rep and 10 predictors. For each rep, find the model with the lowest AIC and lowest AICc and count the number of predictors (K).

```
library(MuMIn)
library(plyr)
options(na.action = "na.fail")
set.seed(9513)
Reps <- 500
N <- 50
SimData <- data.frame(Rep = sort(rep(seq(1, Reps), N)),
                      Y = rnorm(Reps*N), X1 = rnorm(Reps*N),
                      X2 = rnorm(Reps*N), X3 = rnorm(Reps*N),
                      X4 = rnorm(Reps*N), X5 = rnorm(Reps*N),
                      X6 = rnorm(Reps*N), X7 = rnorm(Reps*N),
                      X8 = rnorm(Reps*N), X9 = rnorm(Reps*N),
                      X10 = rnorm(Reps*N))

str(SimData)

Results <- ddply(SimData, c("Rep"),
                function(d) {
                  FullModel <- lm(Y ~ ., data = d)
                  Out1 <- dredge(FullModel, rank = "AIC")
                  Out2 <- dredge(FullModel, rank = "AICc")
                  data.frame(AIC.K = (Out1$df[1] - 2),
                             AICc.K = (Out2$df[1] - 2))
                })

str(Results)
table(Results$AIC.K)
table(Results$AICc.K)
par(mfrow = c(1,2))
barplot(table(Results$AIC.K), main = "AIC")
barplot(table(Results$AICc.K), main = "AICc")
sum(Results$AIC.K > 0)/nrow(Results)
sum(Results$AICc.K > 0)/nrow(Results)
```

```
> str(SimData)
'data.frame':      25000 obs. of  12 variables:
 $ Rep: int  1 1 1 1 1 1 1 1 1 1 ...
 $ Y : num  -0.228 0.601 1.6 -0.469 1.661 ...
 $ X1 : num  -1.08 1.82 -1.6 1.54 1.1 ...
 $ X2 : num   0.164 0.363 -1.043 0.734 1.905 ...
 $ X3 : num  -0.542 -1.497 1.238 0.441 0.105 ...
 $ X4 : num   0.1612 0.642 -1.3871 0.0885 -0.5525 ...
```

```

$ X5 : num -1.1777 1.5603 1.9826 0.5774 -0.0969 ...
$ X6 : num -0.798 -0.685 1.761 -0.89 0.468 ...
$ X7 : num -1.659 1.758 -0.721 -0.479 0.136 ...
$ X8 : num -1.7635 -0.3171 0.0831 -0.0638 -0.2011 ...
$ X9 : num 1.231 -0.793 -0.155 1.092 1.521 ...
$ X10: num 0.945 0.455 -1.157 -0.735 0.516 ...
> str(Results)
'data.frame':      500 obs. of  3 variables:
 $ Rep   : int   1 2 3 4 5 6 7 8 9 10 ...
 $ AIC.K : num   1 4 3 6 0 3 1 3 1 3 ...
 $ AICc.K: num   1 3 3 5 0 3 1 2 1 2 ...
> table(Results$AIC.K)

 0    1    2    3    4    5    6    7
70 133 143  81  42  17  12    2
> table(Results$AICc.K)

 0    1    2    3    4    5    6
93 170 141  64  25   6    1
> par(mfrow = c(1,2))
> barplot(table(Results$AIC.K), main = "AIC")
> barplot(table(Results$AICc.K), main = "AICc")
> sum(Results$AIC.K > 0)/nrow(Results)
[1] 0.86
> sum(Results$AICc.K > 0)/nrow(Results)
[1] 0.814

```

