Information Criteria for Model Selection

Andrew Zieffler

March 8, 2016

The Data

```
mn = read.csv(file = "/Users/andrewz/Documents/GitHub/EPsy-8252/data/mnSchools.csv")
head(mn)
```

id	name	gradRate	public	sat	tuition
1	Augsburg College	65.2	0	1030	39294
3	Bethany Lutheran College	52.6	0	1065	30480
4	Bethel University, Saint Paul, MN	73.3	0	1145	39400
5	Carleton College	92.6	0	1400	54265
6	College of Saint Benedict	81.1	0	1185	43198
7	Concordia College at Moorhead	69.4	0	1145	36590

Fit the Candidate Models

```
lm.a = lm(gradRate ~ public + sat + tuition, data = mn)
lm.b = lm(gradRate ~ public + sat + tuition + public:sat, data = mn)
lm.c = lm(gradRate ~ public + sat + tuition + public:tuition, data = mn)
lm.d = lm(gradRate ~ public + sat + tuition + public:sat + public:tuition, data = mn)
```

Compute AICc

```
# Compute AICc for Model A
n = 33
k = 5
a = -2 * logLik(lm.a)[[1]] + 2 * k * n / (n - k - 1)


# Compute AICc for Model B
n = 33
k = 6
b = -2 * logLik(lm.b)[[1]] + 2 * k * n / (n - k - 1)


# Compute AICc for Model C
n = 33
k = 6
c = -2 * logLik(lm.c)[[1]] + 2 * k * n / (n - k - 1)
```

```r
# Compute AICc for Model D
n = 33
k = 7
d = -2 * logLik(lm.d)[[1]] + 2 * k * n / (n - k - 1)
```

In summary, here are the four models, and their AICc values.

Model	AICc
Model A	226
Model C	229
Model B	229
Model D	232

Model A has the lowest AICc value. Given the data, and the candidate set of models, Model A is estimated to be closest to full reality.

Difference Between Each Candidate Model and the Best Model

Since the minimum AICc value is for Model A, we compute the difference between each model's AICc value and Model A's AICc value, Δ_i.

```r
# Compute dalta values
delta_a = a - a
delta_b = b - a
delta_c = c - a
delta_d = d - a
```

Model	Delta
Model A	0.00
Model C	2.74
Model B	2.90
Model D	5.96

Using the criteria in Figure 2 on p. 25 of Burnham, Anderson, & Huyvaert (2011):

- Model A has a lot of empirical support
- Model C and B have less empirical support than Model A, but, nonetheless still a fair amount of empirical support
- Model D has considerably less empirical support than Model A

Akaike Weights: Relative Likelihoods for Each Candidate Model

```r
w_a = exp(-1/2 * delta_a)
w_b = exp(-1/2 * delta_b)
w_c = exp(-1/2 * delta_c)
w_d = exp(-1/2 * delta_d)
```

Model	Weight
Model A	1.000
Model C	0.254
Model B	0.235
Model D	0.051

Relative to the other three candidate models, Model A has the highest likelihood. Model C is slightly more likely than Model B, and Model D is not a likely candidate relative to the others in the set.

Evidence Ratios

By themselves, the weights (relative likelihoods) do no give us really any more information than we already had by examining the Δ_i values. One way to make use of these values is to compute the evidence ratio for each model. This allows a comparison to the best model.

```
er_a = w_a / w_a
er_b = w_a / w_b
er_c = w_a / w_c
er_d = w_a / w_d
```

Model	ER
Model A	1.00
Model C	3.94
Model B	4.26
Model D	19.70

- Given the data, and the set of candidate models, the empirical support for Model A is 4.26 times that for Model B.
- Given the data, and the set of candidate models, the empirical support for Model A is 3.94 times that for Model C.
- Given the data, and the set of candidate models, the empirical support for Model A is 19.70 times that for Model D.

Model Probabilities

We can also use the model weights to compute the model probabilities.

```
mp_a = w_a / (w_a + w_b + w_c + w_d)
mp_b = w_b / (w_a + w_b + w_c + w_d)
mp_c = w_c / (w_a + w_b + w_c + w_d)
mp_d = w_d / (w_a + w_b + w_c + w_d)
```

Model	Prob
Model A	0.650
Model C	0.165
Model B	0.152
Model D	0.033

- Given the data, and the set of candidate models, the probability of Model A is 0.650.
- Given the data, and the set of candidate models, the probability of Model B is 0.152.
- Given the data, and the set of candidate models, the probability of Model C is 0.165.
- Given the data, and the set of candidate models, the probability of Model D is 0.033.

Use Pre-Existing Functions

We will use the `aictab()` function from the **AICcmodavg** package to compute these values directly from the `lm()` fitted models. This function takes a list of models in the candidate set (it actually has to be an R list). The optional argument `modnames=` is a vector of model names associated with the models in the candidate set.

```r
library(AICcmodavg)

# AICc Table for Model Selection
myAIC = aictab(
  cand.set = list(lm.a, lm.b, lm.c, lm.d),
  modnames = c("Model A", "Model B", "Model C", "Model D")
  )
myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt
1	Model A	5	226	0.00	1.000	0.650	-107	0.650
3	Model C	6	229	2.74	0.254	0.165	-107	0.815
2	Model B	6	229	2.90	0.235	0.152	-107	0.967
4	Model D	7	232	5.96	0.051	0.033	-107	1.000

```r
# Evidence Ratios
evidence(myAIC, model.high = "Model A", model.low = "Model B")
```

```
##
## Evidence ratio between models 'Model A' and 'Model B':
## 4.26
```

```r
evidence(myAIC, model.high = "Model A", model.low = "Model C")
```

```
##
## Evidence ratio between models 'Model A' and 'Model C':
## 3.94
```

```r
evidence(myAIC, model.high = "Model A", model.low = "Model D")
```

```
##
## Evidence ratio between models 'Model A' and 'Model D':
## 19.7
```

Note that you still would have to compute the model probabilities from this output. Since the table output was assigned to the object `myAIC` we can access the columns and automate these computations.

```r
# Examine object
str(myAIC)
```

```
## Classes 'aictab' and 'data.frame':   4 obs. of  8 variables:
##  $ Modnames  : Factor w/ 4 levels "Model A","Model B",..: 1 3 2 4
##  $ K         : num  5 6 6 7
##  $ AICc      : num  226 229 229 232
##  $ Delta_AICc: num  0 2.74 2.9 5.96
##  $ ModelLik  : num  1 0.2537 0.2346 0.0508
##  $ AICcWt    : num  0.65 0.165 0.152 0.033
##  $ LL        : num  -107 -107 -107 -107
##  $ Cum.Wt    : num  0.65 0.815 0.967 1
```

Since `myAIC` is a data frame, we can add columns, use indexing, submit results to `ggplot()`, etc. Here we compute the evidence ratios and the model probabilities and add them as columns into the data frame.

```r
# Create new columns
myAIC$ER = max(myAIC$AICcWt) / myAIC$AICcWt
myAIC$Prob = myAIC$AICcWt / sum(myAIC$AICcWt)
myAIC
```

	Modnames	K	AICc	Delta_AICc	ModelLik	AICcWt	LL	Cum.Wt	ER	Prob
1	Model A	5	226	0.00	1.000	0.650	-107	0.650	1.00	0.650
3	Model C	6	229	2.74	0.254	0.165	-107	0.815	3.94	0.165
2	Model B	6	229	2.90	0.235	0.152	-107	0.967	4.26	0.152
4	Model D	7	232	5.96	0.051	0.033	-107	1.000	19.70	0.033

We can also customize the output to only include information we are interested in. For example, we could remove columns, re-order columns, or change column names.

```r
myAIC = myAIC[ , c("Modnames", "K", "AICc", "Delta_AICc", "AICcWt", "ER", "Prob")]
names(myAIC) = c("", "K", "AICc", "Delta", "Weight", "Evidence Ratio", "Model Probability")
```

	K	AICc	Delta	Weight	Evidence Ratio	Model Probability
Model A	5	226	0.00	0.650	1.00	0.650
Model C	6	229	2.74	0.165	3.94	0.165
Model B	6	229	2.90	0.152	4.26	0.152
Model D	7	232	5.96	0.033	19.70	0.033

All of this evidence points to Model A as the best candidate model given the data. Model B and C have some tenability, but are not nearly as supported as Model A. Finally, Model D seems almost completely unsupported.

AICc Function

Lastly, as a sidenote, this package also includes the function `AICc()` which computes the AICc value for a fitted model.

```r
AICc(lm.a)
```

```
## [1] 226
```