Information Theory: Practice

M. Drew LaMar
September 26, 2016

> "If all the statisticians in the world were laid head to toe, they wouldn't be able to reach a conclusion."
> - Anonymous

Introduction to Quantitative Biology, Fall 2016

Class announcements

- Reading assignment for Wednesday, September 28
  - Railsback & Grimm: Chapter 1: Models, Agent-Based Models, and the Modeling Cycle
- Lab this week: We will work through Chapter 2 of Railsback & Grimm (Getting Started with NetLogo)
- Blogs and projects start next week (for real)

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

```
g1 <- lm(y ~ 1, data=cement)
g2 <- lm(y ~ x1 + x2, data=cement)
g3 <- lm(y ~ x1 + x2 + x1*x2, data=cement)
g4 <- lm(y ~ x3 + x4, data=cement)
g5 <- lm(y ~ x3 + x4 + x3*x4, data=cement)
```

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

Discuss: What is K for these models?

Answer: 2, 4, 5, 4, and 5, respectively.

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

The maximum likelihood estimate for the variance of the residuals $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \frac{\mathrm{RSS}}{n},$$

where

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2.$$

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

```
sig1 <- sum((fitted(g1) - cement$y)^2)/13
sig2 <- sum((fitted(g2) - cement$y)^2)/13
sig3 <- sum((fitted(g3) - cement$y)^2)/13
sig4 <- sum((fitted(g4) - cement$y)^2)/13
sig5 <- sum((fitted(g5) - cement$y)^2)/13
sig <- c(sig1, sig2, sig3, sig4, sig5)
```

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

You can go from RSS to log-likelihood as follows:

$$\log(\mathcal{L}) = -\frac{n}{2}\log\left(\hat{\sigma}^2\right)$$

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and x_1*x_2	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1*x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and x_3*x_4	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3*x_4)$

You can go from RSS to log-likelihood as follows:

```
loglik <- sapply(sig, function (x)
{-1*(13/2)*log(x)})
```

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

The corrected Akaike Information Criterion (AICc) is given by

$$\mathrm{AICc} = n \log\left(\hat{\sigma}^2\right) + 2K + \frac{2K(K+1)}{n-K-1}.$$

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

```
K <- c(2,4,5,4,5)
(aicc <- sapply(1:5, function (i)
{13*log(sig[i]) + 2*K[i] + (2*K[i]*
(K[i]+1))/(13-K[i]-1)}))
```

```
[1] 74.64443 32.41999 37.82382 46.85258 51.32391
```

```
(rnk <- rank(aicc))
```

```
[1] 5 1 2 3 4
```

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

```
  K        sig     LogLik     AICc Rank
1 2 208.904852 -34.722213 74.64443    5
2 4   4.454191  -9.709995 32.41999    1
3 5   4.397135  -9.626196 37.82382    2
4 4  13.518308 -16.926292 46.85258    3
5 5  12.421406 -16.376238 51.32391    4
```

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

```
AICtable <- data.frame(Manual = aicc, Package =
c(AICc(g1),
  AICc(g2),
  AICc(g3),
  AICc(g4),
  AICc(g5)))
```

Example: Hardening of Portland Cement

Hypotheses/Models

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and x_1*x_2	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1*x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and x_3*x_4	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3*x_4)$

```
    Manual   Package
1 74.64443 111.53683
2 32.41999  69.31239
3 37.82382  74.71622
4 46.85258  83.74499
5 51.32391  88.21631
```

Example: Hardening of Portland Cement

H_1	0 variables	g_1	$E(y) = \beta_0$
H_2	x_1 and x_2	g_2	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2)$
H_3	x_1 and x_2 and $x_1 * x_2$	g_3	$E(y) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_1 * x_2)$
H_4	x_3 and x_4	g_4	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4)$
H_5	x_3 and x_4 and $x_3 * x_4$	g_5	$E(y) = \beta_0 + \beta_1(x_3) + \beta_2(x_4) + \beta_3(x_3 * x_4)$

$$\Delta_i = \text{AICc}_i - \text{AICc}_{min}.$$

```
Model    Manual   Package  Delta_M  Delta_P
    2 32.41999  69.31239  0.00000  0.00000
    3 37.82382  74.71622  5.40383  5.40383
    4 46.85258  83.74499 14.43259 14.43259
    5 51.32391  88.21631 18.90391 18.90391
    1 74.64443 111.53683 42.22443 42.22443
```

Example: Hardening of Portland Cement

Model Likelihoods

$$\mathcal{L}(g_i \mid \text{data}) \propto \exp\left(-\frac{1}{2}\Delta_i\right)$$

Model "Probabilities" (Weights)

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{j=1}^{R}\exp\left(-\frac{1}{2}\Delta_j\right)}$$

This is necessary as AICc *estimates* K-L information. The model probabilities give you a probability each model is the *actual* K-L best model.

Example: Hardening of Portland Cement

```
Model  Manual  Delta_M           w
     2 32.420   0.0000 9.3643e-01
     3 37.824   5.4038 6.2813e-02
     4 46.853  14.4326 6.8782e-04
     5 51.324  18.9039 7.3543e-05
     1 74.644  42.2244 6.3468e-10
```

Model averaging (Multimodel Inference)

$$\mathbf{g} = \sum_{i=1}^{R} w_i g_i.$$

Example: Hardening of Portland Cement

As an example, create an arbitrary input x_1, x_2, x_3, and x_4 and look at the predicted values.

```
X <- data.frame(x1 = 0.14, x2 = 0.40, x3 =
0.52, x4 = 0.05)
W <- AICtable$w
pred1 <- predict(g1, X)
pred2 <- predict(g2, X)
pred3 <- predict(g3, X)
pred4 <- predict(g4, X)
pred5 <- predict(g5, X)
AICtable$pred <- c(pred1, pred2, pred3, pred4,
pred5)

print(AICtable %>% select(Model, w, pred) %>%
mutate('w*pred' = w*pred), digits=4,
row.names=F)
```

Example: Hardening of Portland Cement

```
Model          w   pred     w*pred
    1 6.347e-10  95.42 6.056e-08
    2 9.364e-01  53.05 4.968e+01
    3 6.281e-02  54.14 3.401e+00
    4 6.878e-04 130.62 8.984e-02
    5 7.354e-05 135.29 9.950e-03
```

The predicted value for the *averaged model* is given by

```
W[1]*pred1 + W[2]*pred2 + W[3]*pred3 +
W[4]*pred4 + W[5]*pred5
```

```
       1
53.17581
```