

presented: Sept. 29, 2016

Chapter 13.1
Maegwin Bonar
October 6, 2016

Example. GLM, applied to 2-way ANOVA in Sokal and Rohlf 1995 p332.
Does oxygen consumption depend on salinity, in two species of limpet ?

1. Construct model

Data are: oxygen consumption (microliters per minute) / (mg dry weight) of two species of limpet, at three different salinities.

Response variable: Oxygen consumption

Explanatory variables: are salinity levels and species.

Species. X_{sp} = A.scabra, A.digitalis (two categories, nominal scale)

Salinity. X_{sal} = salinity in three categories: 100%, 75%, 50%

Verbal model: Oxygen consumption depends on salinity and species

Write GLM: $V = \beta_0 + \beta_{sp}X_{sp} + \beta_{sal}X_{sal} + \beta_{sp*sal}X_{sp} * X_{sal}$

2. Execute analysis

```
model1 <- lm(VO2~factor(Sp)+factor(Sal)+factor(Sp)*factor(Sal), data = data)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.1738	1.0932	11.136	4.1e-14 ***
factor(Sp)2	0.1525	1.5460	0.099	0.92189
factor(Sal)75	-4.2838	1.5460	2.771	0.00829 **
factor(Sal)100	-1.6125	1.5460	-1.043	0.30290
factor(Sp)2:factor(Sal)75	-0.7050	2.1863	-0.322	0.74870
factor(Sp)2:factor(Sal)100	-3.2850	2.1863	-1.503	0.14044

Note:

12.17 is the mean VO2 for species 1 at 50% salinity

The other 5 means in the two-way design are calculated from the contrasts.

For example, 12.17 + 0.15 is the mean for the other species at 50% salinity

$$LR = \left(\frac{401.52}{623.41} \right)^{-\frac{48}{2}}$$

$$LR = 38\,512.17$$

The model is 38 512 times more likely than the null!

Now we are going to look at the weight of evidence for each of the response variables by looking at the support for the following models given the data:

$$\text{Model 1: } V = \beta_0 + \beta_{sp}X_{sp} + \beta_{sal}X_{sal} + \beta_{sp*sal}X_{sp}X_{sal}$$

$$\text{Model 2: } V = \beta_0 + \beta_{sp}X_{sp} + \beta_{sal}X_{sal}$$

$$\text{Model 3: } V = \beta_0 + \beta_{sal}X_{sal}$$

$$\text{Model 4: } V = \beta_0 + \beta_{sp}X_{sp}$$

$$\text{Model 5: } V = 1$$

```
model1 <- lm(VO2~factor(Sp)+factor(Sal)+factor(Sp)*factor(Sal), data = data)
```

```
model2 <- lm(VO2~factor(Sp)+factor(Sal), data = data)
```

```
model3 <- lm(VO2~factor(Sal), data = data)
```

```
model4 <- lm(VO2~factor(Sp), data = data)
```

```
model5 <- lm(VO2~1, data = data)
```

These are called nested models.

```
anova(model1)
```

Model1 is the residual (all terms in the model)

```
anova(model2)
```

Model2/Model1 isolates the interaction term.

```
anova(model3)
```

Model3/Model2 isolates the species term

```
anova(model4)
```

Model4/model3 isolates the salinity term.

```
anova(model5)
```

Model5 is the null model.

In R we find the log likelihood for the models:

```
logLik(model1)
```

```
'log Lik.' -119.0865 (df=7)
```

```
> logLik(model2)
```

```
'log Lik.' -120.4756 (df=5)
```

```
> logLik(model3)
```

```
'log Lik.' -121.3963 (df=4)
```

```
> logLik(model4)
```

```
'log Lik.' -128.9958 (df=3)
```

```
> logLik(model5)
```

```
'log Lik.' -129.645 (df=2)
```

AIC is the likelihood penalized for number of parameters

Chapter 13.1
Maegwin Bonar
October 6, 2016

In R we find the AIC for the models:

```
AIC(model1) [1] 252.173
> AIC(model2) [1] 250.9512
> AIC(model3) [1] 250.7926
> AIC(model4) [1] 263.9915
> AIC(model5) [1] 263.29
```

Relative likelihood (table below) is a penalized likelihood ratio
 $-119.09 - (-120.09) = 1.38$ called delta AIC
 Model3 has the lowest AIC, it becomes the reference model

Put the results in a table:

Model	LL	AIC	ΔAIC	Relative likelihood $\exp(-0.5\Delta AIC)$	AIC weight
Sp+Sal+Sp*Sal	-119.09	252.17	1.38	0.5015	0.2065
Sp+Sal	-120.48	250.95	0.16	0.9238	0.3804
Sal	-121.40	250.79	0.00	1.0000	0.4118
Sp	-129.00	263.99	13.20	0.0014	0.0006
Null	-129.65	263.29	12.50	0.0019	0.0008

0
0.16
1.38

We can still see that using our criteria of $\Delta AIC < 3$ for the first three models there are three top models. The model with salinity only seems to be the strongest, though when you add on species and the interaction term, the likelihood is less, but there is not enough weight of evidence to say the model doesn't fit the data.

Why 3?

A delta AIC of 0.5×3 would be $\exp(1.5) = 4.5$ times more likely

We suspect a coattail effect in which the response variables Sp and Sp*Sal don't explain the data but are simply "riding the coattails" of the strong response variable Sal that likely explains the data the best.

We can see from the AIC weights that the model using only salinity is most likely the best model of the set. You could use this as an argument for only salinity explaining oxygen consumption in limpets and it could be enough weight of evidence. *But these models could be too close to choose a best model. Usually ecologists will then choose the most complicated model.*

Sample calculation: AIC weight for model 1:

First calculate the relative likelihood:

$$RL = e^{-0.5 \cdot \Delta AIC}$$

$$RL = e^{-0.5 \cdot 1.38}$$

$$RL = 0.5015$$

RL is taken relative to Model3, the model with the lowest AIC - the model "closest to the truth."

Repeat for the rest of the models and then calculate the sum of all the relative likelihoods: Sum = 2.4285

Then we calculate the weight as follows:

$$w_1 = RL_1 / \sum RL_{1-5}$$

$$w_1 = 0.5015 / 2.4285$$

$$w_1 = 0.2065$$

The Akaike weight w_i represents the probability or “weight of evidence” of model i being the best model in the set (Burnham and Anderson, Wildlife Research 2001 28:111-119).

YOU COULD STOP HERE. However, ecologists will sometimes take a model averaging approach. This is where you average the top models (those where $AIC < 3$) and you take the averaged model as the top model. Furthermore, ecologists will sometimes look at the 95% CI to see which variables within the top averaged model are meaningful.

I feel like the model averaging approach is a departure from the biological question. The point of using a model selection approach is that you come up with a few *biologically relevant* models that could explain the data. However, if you average the top models you are moving away from the biological meaning behind the data and are simply performing extra analysis (a “stats on stats” kind of approach).

✓ Additionally, examining the 95% CI is a departure from the likelihood method and brings us back into the realm of hypothesis testing. This is inappropriate as you should not combine likelihood and hypothesis testing, you should only choose one!

But for the sake of examining how an ecologist might tackle this question, we will continue with on with the model averaging approach. **Now we are making a shift away from likelihood and into the realm of hypothesis testing:**

We model average the 3 “best” ($AIC < 3$) models and look at the parameters within the averaged model:

In R:

```
library(MuMIn)
```

```
model.avg <- model.avg(model1, model2, model3)
```

```
summary(model.avg)[["coefficients"]][,1:2]  
confint(model.avg, level = 0.95)
```

From the summary:

	Estimate	Std. Error	LCL	UCL
(Intercept	12.463	0.912	10.631	14.295
factor(Sal)75	-4.600	1.163	-6.942	-2.258
factor(Sal)100	-3.086	1.261	-5.616	-0.557
factor(Sp)2	-0.892	1.202	-3.301	1.518
factor(Sal 75:factor(Sp)2	-0.705	2.186	-5.117	3.707
factor(Sal)100:factor(Sp)2	-3.285	2.186	-7.697	1.127

When we look at the 95% CI for each parameter we see that the Species and Sp*Sal parameters have CIs that overlap 0. Therefore we can say that there is not enough support to suggest that they explain any of the variance given the data. There is support for salinity only as the factor that best explains O₂ consumption in 2 species of limpets.

Did we get the same answer as you would from hypothesis testing?

From the handout:

Hypotheses for the interaction term.

The research hypothesis H_A is that

$$\beta_{sp \times sal} = 0$$

$$H_A: \beta_{sp \times sal} \neq 0$$

$$H_o: \beta_{sp \times sal} = 0$$

Hypotheses for the species term.

The research hypothesis H_A is that

$$\beta_{sp} = 0$$

$$H_A: \beta_{sp} \neq 0$$

$$H_o: \beta_{sp} = 0$$

Hypotheses for the salinity term.

The research hypothesis H_A is that

$$\beta_{sal} = 0$$

$$H_A: \beta_{sal} \neq 0$$

$$H_o: \beta_{sal} = 0$$

ANOVA Table

Source	Df	SS	MS	F	P
factor(Sp)	1	16.64			
factor(Sal)	2	181.32			
factor(Sp):factor(Sal)	2	23.93			
Residuals	42	401.52			
Total	47	623.41			

Compute and table MS

$$MS_{Sp} = SS_{Sp} / df_{Sp} = 16.64$$

$$MS_{Sal} = SS_{Sal} / df_{Sal} = 90.661$$

$$MS_{Sp*Sal} = SS_{Sp*Sal} / df_{sp*sal} = 11.963$$

$$MS_{res} = SS_{res} / df_{res} = 9.56$$

Compute and table F

Fixed effects for salinity and species, so all variance ratios taken relative to MS_{res}

$$F = MS_{Sp} / MS_{res} = 16.638 / 9.56 = 1.74$$

$$F = MS_{sal} / MS_{res} = 90.661 / 9.56 = 9.48$$

$$F = MS_{sp*sal} / MS_{res} = 11.963 / 9.56 = 1.2$$

Calculate Type I error from F-distribution.

$$F_{2,42} = 1.251 \quad p = 0.297 \text{ interaction}$$

$$F_{1,42} = 1.74 \quad p = 0.194 \text{ species effect}$$

$$F_{2,42} = 9.483 \quad p = 0.0004 \text{ salinity effect}$$

Interaction term: $p = 0.39 > \alpha = 5\%$ accept H_0 : no interaction term

Species: $p = 0.25 > \alpha = 5\%$ accept H_0 : no difference in respiration between species

Salinity: $p < 0.001 < \alpha = 5\%$ reject H_0 accept H_A : respiration depends on salinity

We get the same decision using AIC as we did using hypothesis testing

However, if we don't make a decision between the 3 top models, then we get a different answer from hypothesis testing