Mixed Effects Models for Fish Growth

Sanford Weisberg, sandy@stat.umn.edu

January 20, 2012

Much like tree rings on trees, many fish preserve a record of their growth history in annular rings on fish scales and other bony parts. The number of rings provides the age of the fish at capture, and the increment between the rings is a measure of annual growth. We use data from four samples (two years, two gear types) of smallmouth bass in West Bearskin Lake, Minnesota.

```
> loc <- "http://www.stat.umn.edu/~sandy/SMBassWBlong.txt"
> data <- read.table(loc, header=TRUE)[, c(3, 4, 5, 6, 9, 10, 11, 12)]
> library(car)
> some(data, n=5)
```

```
      gear yearcap fish agecap age   inc yearclass year
118      E    1990   78      2   2 1.784      1988 1989
408      E    1990  186      3   1 1.581      1987 1987
608      E    1988  236      5   4 1.655      1983 1986
1099     T    1989  329      6   1 1.196      1983 1983
1192     T    1989  344      6   4 1.127      1983 1986
```

The variables `year` and `age` are, respectively, the year of capture and age at capture. The `yearclass` is the year in which the fish was born, and `gear` is the gear used for fishing, either T for trapnetting done in the spring or E for electrofishing done in the fall. Each row of the data file has information about one increment, so for example a six year old fish would contribute six lines to the data file. There were 430 fish in the study. The tables below give counts of the number of increments.

```
> library(xtable)
> xtable(xtabs(~ yearclass + age, data), digits=0)
```

	1	2	3	4	5	6	7
1983	177	177	177	177	177	119	48
1984	32	32	32	32	24	14	0
1985	1	1	1	0	0	0	0
1986	18	18	16	7	0	0	0
1987	150	131	84	0	0	0	0
1988	39	23	0	0	0	0	0
1989	3	0	0	0	0	0	0

The `xtable` function produces LaTeX output, and `xtabs` tabulates the data.

Weisberg (1993, *Can. J. Fish. Aquat. Sci.*, 50, 1229) proposed a fixed-effects linear model for growth based on annular increments. We define y_{cka} to be the a-th annular increment for the k-th fish from year-class c. Then the basic fixed-effects model can be described by the equation

$$y_{cka} = \iota_a + \eta_{c+a-1} + \gamma g_{ck} + e_{cka} \tag{1}$$

where ι_a is the *intrinsic growth* or annular increment for a fish in the a-th year of life, η_ℓ is the *environmental effect* for year $\ell = c + a - 1$, which is the year in which a fish in year class c will be of age a. The e_{cka} were assumed to be independent errors with mean zero and common variance σ^2. Model (1) *does not* include an intercept, which allows an estimate to be obtained for ι_a for each a. γ is a parameter to be estimated and g_{ck} is a dummy variable equal to one if gear type used to catch the k-th fish in year-class c is T and equal to zero if the gear type is E. This model is easily fit with `lm`:

```
> data$age <- factor(data$age)
> data$year <- factor(data$year)
> m1 <- lm(inc~ 0 + age + year + gear, data=data)
```

The fixed-effects growth model has a number of defects. First, the estimates of the ι_a and the η_ℓ are *not unique*. For any constant m, adding m to each of the $\hat{\iota}_a$ and simultaneously subtracting m from all the $\hat{\eta}_\ell$ would give a new set of estimates that gives the same fitted values as those produced above by R. R produces unique estimates by setting $\hat{\eta}_\ell = 0$ for ℓ equal to the earliest year in the data; SAS would produce unique estimates by setting to zero the environmental effect with ℓ equal to the last year in the data, and other programs might do something else entirely. The result of this is to make the estimates depend on the years of sampling, and so comparison of estimates between locations or over disconnected time periods may be impossible. However, in comparisons between years or between ages within a single study the arbitrary constant m will cancel so these comparisons are meaningful.

Second, the assumption that the increments are *independently distributed, with common constant variance* was originally made for computational convenience, but can hardly be expected to reflect the reality of biological growth. All fish within a year are likely to be correlated, as environmental effects are likely to impact all fish in that year. Similarly, increments *on the same fish* will in most populations be (positively) correlated. Fish that grow poorly in year ℓ would typically be more likely to grow poorly in year $\ell + 1$, while relatively larger fish in a year-class are more likely to continue to exhibit more growth in succeeding years.

A mixed-effects model

This section follows Weisberg, Spangler and Richmond (*Can. J. Fish. Aquat. Sci.*, 67(2): 269-277 (2010) http://dx.doi.org/10.1139/F09-181). Each fish contributes more than one measurement, and so a random effect between fish should be expected. Similarly, treating years as a random effect, at least as a baseline model, can make sense in problems with data from many years.

We now assume that the random environmental effects h_ℓ are random draws from a normal distribution with mean zero and variance σ_h^2. Similarly, let f_{ck} be the random effect with zero mean and variance σ_f^2 for the k-th fish of year-class c. The mixed model is

$$y_{cka} = \iota_a + h_{c+a-1} + \gamma g_{ck} + f_{ck} + e_{cka} \tag{2}$$

The R fit is

```
> library(lme4)
> print(n2 <- lmer(inc ~ 0 + age + gear + (1|year) + (1|fish),data), corr=FALSE)

Linear mixed model fit by REML
Formula: inc ~ 0 + age + gear + (1 | year) + (1 | fish)
   Data: data
  AIC  BIC logLik deviance REMLdev
 1164 1224   -571     1104    1142
Random effects:
 Groups   Name        Variance Std.Dev.
 fish     (Intercept) 0.0502   0.224
 year     (Intercept) 0.0268   0.164
 Residual             0.0825   0.287
Number of obs: 1710, groups: fish, 420; year, 7

Fixed effects:
      Estimate Std. Error t value
age1    1.2724     0.0669   19.02
age2    1.3082     0.0666   19.64
age3    1.2592     0.0686   18.34
age4    1.2110     0.0720   16.82
age5    1.1370     0.0727   15.64
age6    1.2239     0.0751   16.30
age7    1.1642     0.0872   13.35
gearT   0.1252     0.0289    4.34
```

This standard output gives: Fit statistics, based on the value of the log-likelihood; estimates of variance components; estimated fixed effects and their standard errors. Correlations between the fixed effects have been suppressed using `corr=FALSE`. No tests of effects for the t-values are provided. However, you can get approximate tests of overall effects using the `Anova` function with a capitol "A" in the `car` package:

```
> require(car)
> Anova(n2)

Analysis of Deviance Table (Type II tests)

Response: inc
     Chisq Df Pr(>Chisq)
age  414.2  7   < 2e-16
gear  18.8  1    1.4e-05
```

This does a *Type II* anova, meaning that all main effects are adjusted for each other, all two factor interactions are adjusted for other two-factor interactions and main effects, and so on. This is the most appropriate general approach to analysis of variance.

This appears to be the appropriate baseline model for these data, as it includes fixed intrinsic age effects, predicted random effects for years, and also models within-fish correlation by allowing each fish to have its own overall level of growth that applies to all increments on that fish. The estimated within fish correlation is $.0502/(.0502 + .0268 + 0.0825) = 0.314$. The estimate of σ_f^2

is larger than the estimate of σ_h^2, but smaller than the estimate of σ^2. Apparently most of the variation in these data is between fish and the random error not otherwise modeled in (2).

The "estimates" of the year random effects, which are of primary interest to the fish ecologists, can be recovered using the rather obscure commands

```
> out<- rbind(t(ranef(n2)$year),
+             sqrt(attr(ranef(n2, postVar=TRUE)$year, "postVar")[1,1, ]))
> rownames(out) <- c("Estimate", "SE")
> out

             1983    1984     1985     1986    1987    1988    1989
Estimate 0.16840 0.11339 -0.26101 -0.18988 0.05771 0.07438 0.03701
SE       0.02512 0.02349  0.02345  0.02266 0.01837 0.01936 0.02499
```

Here is a reasonable graph that summarizes the fish effects, namely a density estimate.

```
> plot(density(ranef(n2)$fish[, 1]), xlab="Increment",
+      main="Fish effects, n2")
```


Fish effects, n2
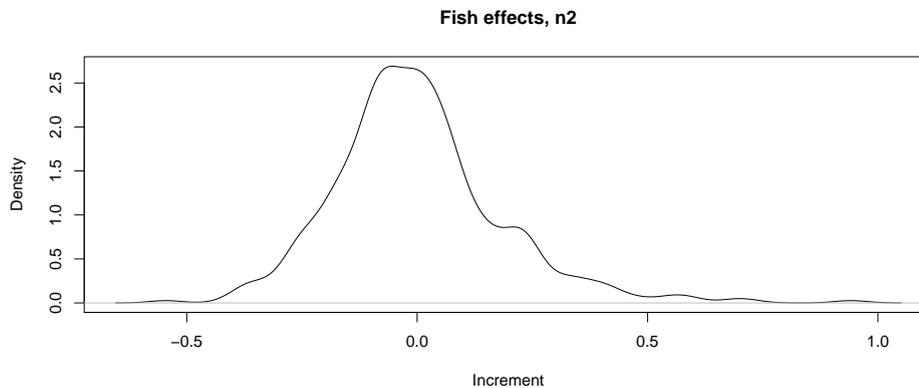
Mixed-effects with random effect interactions

A valid criticism of the fixed effects model is that if a year×age interaction is present, the model becomes too complex to be of much practical use. A mixed model with interactions is much simpler:

$$y_{cka} = \iota_a + h_{c+a-1} + (ih)_{a,c+a-1} + \gamma_{ck} + f_{ck} + e_{cka} \tag{3}$$

This adds only one new variance term σ_{ih}^2 for interaction rather than a large number of fixed effects terms. The predictions for the year main effects are still valid and useful.

```
> print(n3 <-lmer(inc~0+age+gear+(1|year)+(1|year:age)+
+           (1|fish),data),corr=FALSE)

Linear mixed model fit by REML
Formula: inc ~ 0 + age + gear + (1 | year) + (1 | year:age) + (1 | fish)
   Data: data
  AIC  BIC logLik deviance REMLdev
 1135 1201   -556     1090    1111
Random effects:
```

```
 Groups    Name          Variance Std.Dev.
 fish     (Intercept) 0.0498    0.2232
 year:age (Intercept) 0.0370    0.1923
 year     (Intercept) 0.0050    0.0707
 Residual              0.0795    0.2820
Number of obs: 1710, groups: fish, 420; year:age, 26; year, 7


Fixed effects:
      Estimate Std. Error t value
age1    1.3975     0.0884   15.81
age2    1.3193     0.0929   14.20
age3    1.1859     0.1027   11.55
age4    1.1582     0.1259    9.20
age5    1.1485     0.1485    7.73
age6    1.2198     0.1508    8.09
age7    1.1954     0.2074    5.76
gearT   0.1169     0.0287    4.07

> anova(n2,n3)

Data: data
Models:
n2: inc ~ 0 + age + gear + (1 | year) + (1 | fish)
n3: inc ~ 0 + age + gear + (1 | year) + (1 | year:age) + (1 | fish)
   Df  AIC  BIC logLik Chisq Chi Df Pr(>Chisq)
n2 11 1126 1186   -552
n3 12 1114 1179   -545  14.8      1    0.00012
```

The interaction model provides a decrease in AIC. Also, the two models can be compared using the asymptotic χ^2 distribution of the difference in deviances, with p-value of zero to three decimals. The `year:age` interaction is larger than the `year` variance, suggesting that for this particular data set the interactions are more important than previously claimed. The plot of the year effects shown above indicates that the apparently poor years of 1985 and 1986 may not have been as poor as suggested in the no-interaction model, and the years 1983 and 1984 might be better than previously reported.

```
> out1 <- rbind(t(ranef(n3)$year),
+       sqrt(attr(ranef(n3,postVar=TRUE)$year,"postVar")[1,1,]))
> rownames(out1) <- c("Estimate", "SE")
> out1

             1983     1984      1985     1986     1987     1988    1989
Estimate 0.005976 0.01495 -0.01558 -0.0716 0.01927 0.03657 0.01041
SE       0.066462 0.06314  0.06228  0.0597 0.05708 0.05558 0.05512
```

We can summarize the three models using a couple of graphs. First, define a useful function that uses the `plotrix` package.

```
> coefplot <- function(xaxis,coefs,ses,col="black",...) {
+  require(plotrix)
+  plotCI(xaxis,coefs,ses,...)
+  lines(xaxis,coefs,lty=2,col=col)}
```

```
> m1.coefs <- summary(m1)$coef
> par(mfrow=c(1,2))
> coefplot(1:7, fixef(n2)[1:7], sqrt(diag(vcov(n2))[1:7]), xlab="Age",
+         pch="N", ylab="Increment", ylim=c(1.0, 1.55))
> coefplot(.1+(1:7), fixef(n3)[1:7], sqrt(diag(vcov(n3))[1:7]),
+         pch="I", add=TRUE, col="red", scol="red")
> coefplot(1:7, m1.coefs[1:7,1], m1.coefs[1:7,2], xlab="Age",
+         pch="F", ylab="Increment", col="blue", scol="blue", add=TRUE)
> legend("bottomleft", legend=c("NoInt", "Int", "Fixed"),
+         col=c("black", "red", "blue"), pch=c("N","I","F"))
> # year plot
> coefplot(1983:1989,out[1, ], out[2, ],
+         pch="N", xlab="Year", ylab="Increment", ylim=c(-.5, .3))
> coefplot(.1+(1983:1989),out1[1, ],out1[2, ], add=TRUE,
+         pch="I", col="red", scol="red")
> coefplot(1983:1989, c(0,m1.coefs[8:13, 1]), c(0,m1.coefs[8:13, 2]),
+         pch="F", xlab="Age", ylab="Increment",
+         col="blue", scol="blue", ylim=c(1.0, 1.55), add=TRUE)
> legend("bottomleft", legend=c("NoInt", "Int", "Fixed"),
+         col=c("black", "red", "blue"), pch=c("N", "I", "F"))
```
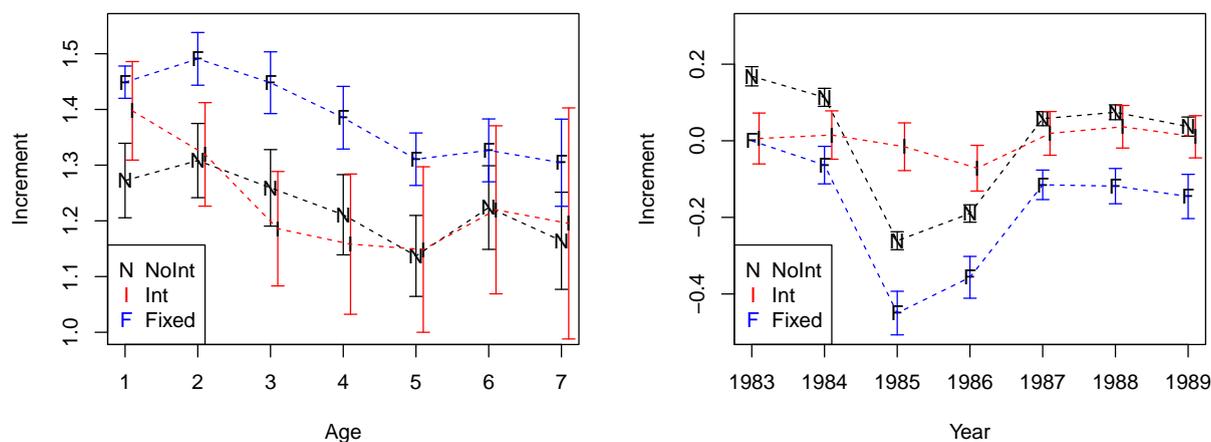

Next steps

1. Delete year 1 increments and refit.

2. *Nonlinear* mixed effects: rather than fitting a factor for age t, fit $g(t, \theta)$. For example, a usual The von Bertalanffy function posits that

$$y(t) = L_\infty(1 - \exp(K(t - t_0))) + \text{ error}$$

where L_∞ is the theoretical limit to size, K is a rate parameter, and t_0 is a negative value that guarantees that y is positive at birth, $t = 0$. Now the data we use is repeated

increments on a scale, and so the increment for age $t \geq 1$ is, apart from error terms,

$$
\begin{aligned}
\text{increment} \quad &= \quad y(t) - y(t-1) \\
&= \quad L_\infty(1 - \exp(K(t - t_0))) - [L_\infty(1 - \exp(K(t - 1 - t_0)))] \\
&= \quad L_\infty \exp(K(t - t_0))(1 - \exp(-K))
\end{aligned}
$$

We could fit with a mean function like

```
y~Linf*exp(K*(t-t0))*(1-exp(-K)) + (1|fYear) + (1|fish)
```

to estimate L_∞, K and t_0, using `nlmer`.

3. Explore the "estimated" ih random effects.