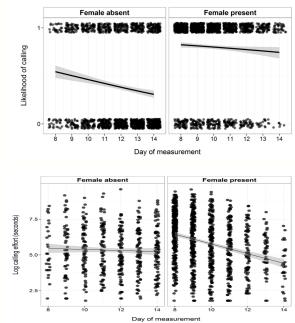


Advancing in R combining approaches



Outline

- How to advance even further
- More complex analyses
 - glmm
 - nlme
 - ZAP
- Suggested workflow

So far...

- Data manipulation
- Data visualization
- Univariate regression
- Multiple regression
- ANCOVA
- Generalized models
- Mixed models
- Nonlinear models

What is left?

- For example, in linear models universe:
 - Generalized linear mixed models
 - Nonlinear mixed effects models
 - Survival models
 - Zero-altered models
- Other techniques unrelated to lm()
 - Canonical analyses (data reduction)

glmms

- <http://glmm.wikidot.com/faq>
- You will need to do a lot of reading, and be able to cope with uncertainty

DISCLAIMERS:

- (GLIMMs are hard - harder than you may think based on what you may have learned in your second statistics class, which probably focused on picking the appropriate sums of squares terms and degrees of freedom for the numerator and denominator of an F-test). GLIMMs are hard because they are more powerful (they can handle complex designs, lack of balance, crossed random factors, some kinds of non-normally distributed responses, etc.), are more general, and can be used in more situations. You should at least a general familiarity with classical mixed-models/experimental designs but you should also probably read something about modern mixed model approaches (Littell et al. [18] and Pinheiro and Bates [21] are two places to start, although Pinheiro and Bates is probably the better place to start if you are interested in the theory anyway). If you are going to use generalized linear mixed models, you should understand generalized linear models.
- All of the issues that arise in regular linear or generalized linear models (e.g.: inseparability of p-values alone for those statistical analyses, need to understand how models are parameterized) still apply to understand the principle of marginality and how interactions can be treated; dangers of overfitting, which are not mitigated by stepwise procedures, are still there.
- When SAS (or Statst, or Genstat/AS-REML, or ...) and R differ in their answers, R may not be wrong. Both SAS and R may be "right" but proceeding in a different way/answering different questions/using a different philosophical approach (or both).
- The advice in this FAQ comes with **absolutely no warranty of any sort**.

Inference on (G)LMMs

What are the p-values listed by summary(glmerfit) etc.? Are they reliable?

By default, in keeping with the tradition in analysis of generalized linear models, lme4 and similar packages display the Wald Z-statistics for each parameter in the model summary. These have one big advantage: they're convenient to compute.

Example: how to generate p-values for glmms?

Tests of single parameters

From worst to best:

- Wald Z-tests
- For balanced, nested LMMs where df can be computed: Wald t-tests
- Likelihood ratio test, either by setting up the model so that the parameter can be isolated/dropped (via anova or drop1), or via computing likelihood profiles
- MCMI or parametric bootstrap confidence intervals

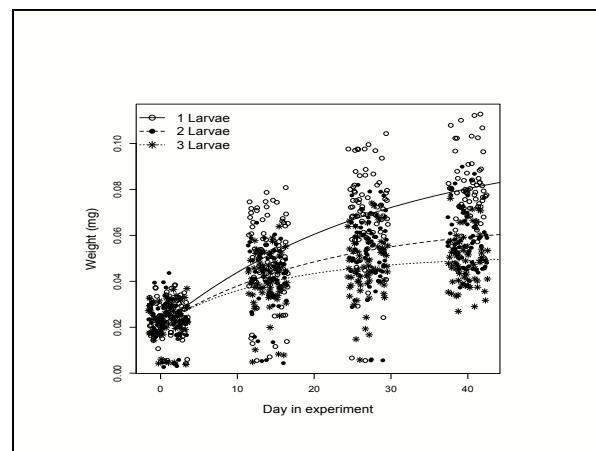
Tests of effects (i.e. testing that several parameters are simultaneously zero)

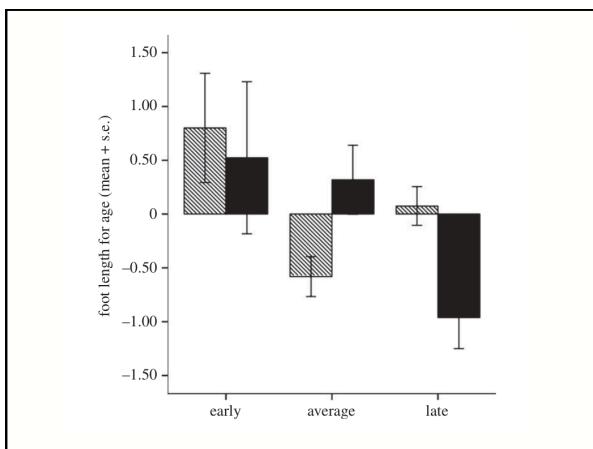
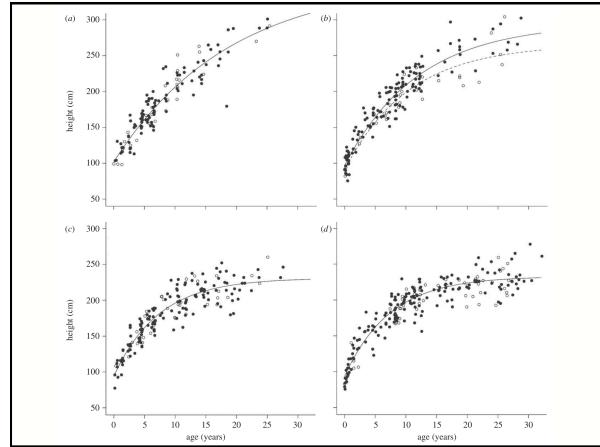
From worst to best:

- Wald chi-square tests (e.g. car::Anova)
- Likelihood ratio test (via anova or drop1)
- For balanced, nested LMMs where df can be computed: conditional F-tests
- For LMMs: conditional F-tests with df correction (e.g. Kenward-Roger in lmerTest package)
- MCMI or parametric bootstrap confidence intervals: bootstrap comparisons (nonparametric bootstrapping must be implemented carefully to account for grouping factors)

Is the likelihood ratio test reliable for mixed models?

- It depends
- Not for fixed effects in finite-size cases (see [21]): may depend on 'denominator degrees of freedom' (number of groups) and/or total number of samples - total number of parameters
- Conditional F-tests are preferred for LMMs, if denominator degrees of freedom are known





Fixed effect	Estimate	95% CI (lower, upper)	P
Zero-altered			
(Intercept)	-4.621	(-5.026, -4.195)	<0.001 ***
Female presence	2.981	(2.523, 3.477)	<0.001 ***
Diet	0.019	(-0.159, 0.197)	0.815
Day	-0.120	(-0.227, -0.003)	0.036 *
Body condition (beginning week 2)	0.364	(0.010, 0.751)	0.062
Mate availability history	-0.640	(-0.865, -0.439)	<0.001 ***
Female presence × day	0.776	(0.577, 0.993)	<0.001 ***
Mate availability history × diet	0.053	(-0.025, 0.134)	0.188
Mate availability history × day	0.012	(-0.038, 0.067)	0.600
Mate availability history × diet × day	0.018	(-0.011, 0.045)	0.210
Poisson			
(Intercept)	4.066	(3.739, 4.400)	<0.001 ***
Female presence	0.736	(0.471, 0.991)	<0.001 ***
Diet	0.348	(0.251, 0.461)	<0.001 ***
Day	-0.195	(-0.265, -0.133)	<0.001 ***
Body condition (beginning week 2)	0.825	(0.563, 1.153)	<0.001 ***
Mate availability history	-0.075	(-0.203, 0.070)	0.296
Female presence × day	-0.215	(-0.339, -0.094)	<0.001 ***
Mate availability history × diet	0.023	(-0.048, 0.100)	0.204
Mate availability history × day	0.083	(0.028, 0.076)	<0.001 ***
Mate availability history × diet × day	-0.020	(-0.033, -0.007)	<0.001 ***
Variance component			
Zero-altered			
ID	4.969	(3.844, 6.105)	
Poisson			
ID	1.819	(1.557, 2.331)	
Residual	1.273	(1.181, 1.391)	

