

CHAPTER 7

An introduction to survival analysis using generalized linear mixed models

Erik Hedlin and Alastair Franke

Hedlin, E., and A. Franke. 2017. An introduction to survival analysis using generalized linear mixed models. Pages 127–146 in D.L. Anderson, C.J.W. McClure, and A. Franke, editors. Applied raptor ecology: essentials from Gyrfalcon research. The Peregrine Fund, Boise, Idaho, USA. <https://doi.org/10.4080/are.2017/007>

7.1 Introduction

Important insight into population dynamics can be gained from studying the ways in which survival and fecundity are influenced by the environmental conditions to which free-living animals are exposed (Caughley 1977). For birds, understanding the factors that affect survival can be difficult, because individuals are simultaneously subjected to many sources of mortality. Further, survival data gathered in the field are frequently exposed to environmental variables that cannot be controlled for despite their impact on the response variable. Mixed models have become an increasingly attractive option for ecologists because they can address issues associated with random and hierarchical variation. In this chapter, we introduce the structure of generalized linear mixed models, and provide a working example using Peregrine Falcon (*Falco peregrinus*) nestling survival data.

7.1.1 Generalized linear mixed models

Ecological studies are often subjected to sources of variation that affect the response variable, but aren't interesting within the context of the research questions. Further, this variation is often grouped at different levels within the system that can lead to a lack of independence among experimental units. For example, ecologists often collect data over several years, but factors such as food abundance vary from year to year and can

influence the response variable. As a result, there may be consistent variation in survival according to the year nestlings are reared. In this case, the “year effect” on the species may not be of interest, but accounting for it within the model will preserve assumptions of independence while increasing fit and predictive power. When the year effect is included as a random variable in the model, the effect on the response variable is determined and compensated for by letting other variables of interest vary according to that effect (for more detail and an example of a simple linear model that incorporates random effects, see Chapter 1 of Pinheiro and Bates 2000). In this way, mixed models can be used to address many of the challenges typically associated with ecological studies.

7.1.2 Fixed vs. random

Due to several competing philosophies and definitions, deciding whether to treat a certain variable as fixed or random can be difficult (Kreft and De Leeuw 1998, Gelman 2005). With definitions ranging from how they are calculated to their level of interest (Gelman 2005), there is not one universal principle that defines random variables. Assigning variables as fixed or random can therefore be a subjective task (Schabenberger and Francis 2001). However, one must ensure that any variable treated as random within the model has more than five categorical levels. This is because random effect estimation attempts to determine among-level variance, but sufficient levels need to be present to do so accurately and without error (Bolker et al. 2009). If there are fewer than five levels (as we see below with our variable “year”), the random variable in question should be included as a fixed effect (Crawley 2002, Bolker et al. 2009).

7.1.3 The model for survival

In this chapter we investigate survival of nestlings in which there can only be two outcomes, a nestling survives = 1, or it dies = 0. This results in a binomial distribution in our response variable, and we therefore need to use some form of logistic regression. As noted below, our experimental unit is the nestling, and each nestling can be organized into the random blocks: “year” and “nest.” This introduces hierarchical variation, so a logistic mixed effect model, also termed generalized linear mixed model, is the best structure for our data. The general structure in matrix notation is:

Equation 7.1

$$Y = \beta X + \gamma Z + \epsilon$$

where Y is a vector containing the response, β is a vector containing the fixed effect regression coefficients, X represents the vector containing the predictor variables and their values, γ is the random effect, Z represents the

random effect design matrix, and ε represents the error (the variation not explained by the model).

Equation 7.2

$$Y = \begin{bmatrix} \text{SURVIVAL} \\ 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$X = \begin{bmatrix} \text{intercept}^a & \text{relhatch}^b & \text{hatch2}^c & \text{hatch3}^d & \text{hatch4}^e & \text{treatment}^f & \text{year2}^g & \text{year3}^h & \text{year4}^i \\ 1 & -4 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 5 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 2.72^a \\ -0.28^b \\ -1.49^c \\ -2.00^d \\ -3.49^e \\ 2.96^f \\ -1.49^g \\ -0.90^h \\ -1.56^i \end{bmatrix}$$

Equation 7.2 extends the matrix form of the fixed model structure by incorporating our data. Note that we added notation to indicate the matching Beta estimates and predictor variables. We can see that the response variable is a vector of survival observations for each of our experimental units. Next we have a matrix of predictor variables: hatch order within the brood, hatch date relative to the rest of the population, and treatment (whether an individual was food supplemented or not). The variable “year” has the qualities of a random effect because it can be used to group our experimental units, however, it’s included in the fixed structure of the model because it does not have sufficient levels (Bolker 2009). Looking at Equation 2 we can also see how the categorical variables HATCH and YEAR are treated within the model. It is unlikely that the effect of hatching first to fourth is linear and we therefore want to compare each level of hatch order to a reference. Although R automatically assigns reference based on either alphabetical or numerical value (first hatch in this case), we can specify which level of the variable we want as the reference by using the `relevel` function (run `?relevel` in R for more information). Year is also treated as a categorical variable with the year 2013 as our reference. If not treated categorically, the year 2013 would be treated as the value 2,013 and an incorrect effect based on this numerical

value would be calculated. Both of these categorical variables are converted to a dummy variable matrix composed of zeros and ones as seen above. Lastly, we estimate the vector of coefficients (β) for our predictor variables.

Missing from Equation 7.2 is the residual (ϵ), and random portion of our model (γZ). As seen in Fig. 7.1, random effects are calculated as deviations from the fixed intercept effect (β_0). In other words, each level of our random variable (nests in this case) is assigned a random intercept based on its estimated effect on survival. These random intercepts are all centered around the fixed intercept β_0 with a mean of zero, and a variance that is estimated from the nest effect as a whole. One can add complexity to the random portion of this model by including random slopes, but this is not necessary given the structure of our data.

The residual portion of the model is a covariance structure that estimates the variation not explained by the model. The residual covariance can assume different structures that account for a lack of independence in the data, but in its general form (the form we use in this chapter), the residual covariance is composed of an identity matrix (a diagonal matrix made of 1s and 0s) and the conditional residual variance. For more information on residual covariance, see Zuur et al. 2009.

7.2 Survival data

The data used in this chapter come from a research program conducted in the eastern Canadian Arctic around the community of Rankin Inlet, Nunavut (Court et al. 1989, 1990, Johnstone et al. 1996, Bradley et al. 1997, Franke et al. 2010, Anctil and Franke 2013, L'Hérault et al. 2013, Anctil et al. 2014), over a four-year period. A total of 270 Peregrine Falcon nestlings from 84 broods were monitored from hatch to fledge with two objectives, 1) to determine factors that affect nestling survival, and 2) to determine whether food is limiting during the brood rearing period. To better understand the extent of food limitation, a random subset of broods within the population was food supplemented ($n = 40$), while the remaining broods were kept as a control ($n = 44$). The response variable in the analysis is the categorical variable survival (1 = survived and 0 = died), we are including a priori predictor variables based on previous research conducted on the population (Anctil et al. 2014). These predictor variables are within-brood hatch order, hatch date relative to the whole population (0 = the yearly median), within-brood asynchrony (number of days an individual hatched after their oldest sibling), and a treatment effect (s = individual was food supplemented, c = individual was not food supplemented). Additionally, two grouping variables are included in each model to account for consistent variation within broods (random), and years (fixed). As in Chapter 6 (this volume), these data are organized in long format where each variable is represented as a column, and each data point is represented as a row (Table 7.1).

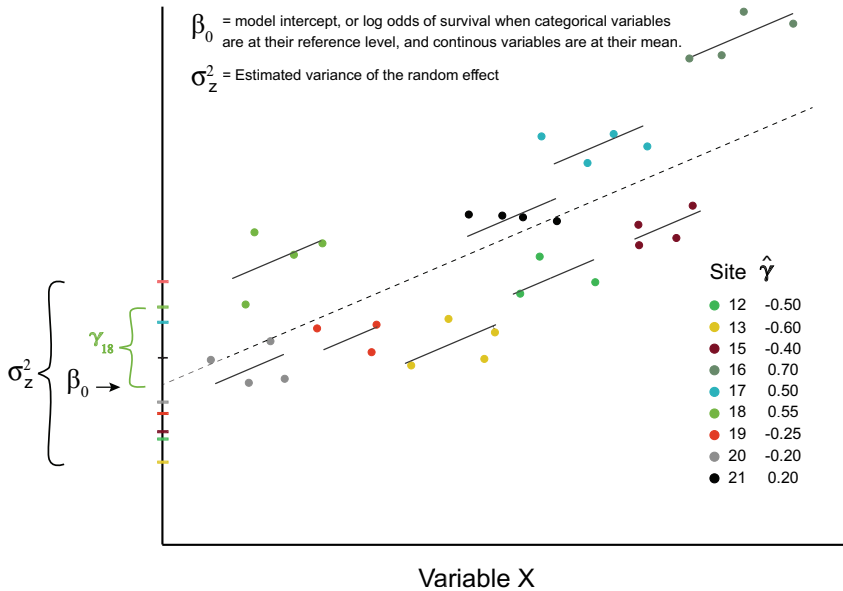


Figure 7.1 An illustration of random intercepts that vary across nest sites. Each point represents the log odds of survival across the range of variable X for a specific nestling, and individuals are grouped according to the brood in which they were reared (i.e., nestlings of the same color are siblings). Using this figure, we can see that our random intercepts are deviations around the fixed effect intercept (β_0), and therefore have a mean of zero. The variance depends on the data, and how diverse the brood effect is across our samples.

7.3 Using the model to estimate survival

The general goal of this analysis is to gather valid inference about nestling survival during the brood rearing period by using models that best approximate that process. Because our inference is based on the models we build and select, the strength of our inference is dependent on how accurately our top models approximate nestling survival. There are many methods of building and selecting models, but here we use an information theoretic (IT) approach as outlined by Burnham and Anderson (2002). The IT method of model selection differs from traditional null hypothesis testing in three important ways: 1) it examines multiple models simultaneously, 2) it uses measures of parsimony rather than P -values to evaluate model quality, and 3) inference can be drawn across multiple models to reduce bias (Symonds and Moussalli 2011). The process can be divided into three steps: 1) build a set of a priori candidate models that are ecologically justifiable, 2) fit and rank these models according to appropriate

Table 7.1. An example of how the data should be organized. Each column represents one variable, and each row represents one observation.

ID	TREATMENT	SURVIVAL	HATCH	RELHATCH	ASYNCHRONY	NESTYEAR
4b13	s	1	1	-2	0	4y2013
4g13	s	1	2	-1	1	4y2013
4bk13	s	1	3	0	2	4y2013
4r13	s	1	4	1	3	4y2013
7r13	c	1	1	-4	0	7y2013
7g13	c	1	2	-4	0	7y2013
7b13	c	1	3	-3	1	7y2013
7bk13	c	1	4	0	4	7y2013
8r13	c	1	1	-1	0	8y2013
8g13	c	1	2	0	1	8y2013
8b13	c	1	3	2	3	8y2013
8bk13	c	1	4	3	4	8y2013
12r13	c	0	1	-1	0	12y2013
12g13	c	0	2	0	1	12y2013
19r13	s	1	1	-5	0	19y2013
4b13	s	1	1	-2	0	4y2013

IT criteria, and 3) make inferences about nestling survival using parameters from the top model(s).

7.3.1 Build a candidate model set

Once a group of predictor variables is established, we begin building our candidate models. The candidate set (Table 7.2) usually includes a global model that contains all our predictor variables, followed by models that include different combinations of our predictor variables and their interactions. Lastly we include the ‘null’ model without predictor variables. Note that variable notation in this example candidate set has been changed to be more descriptive and that we only show fixed effects.

7.3.2 Fitting and ranking the models

Once we have our list of candidate models, we fit the models with the `glmer` function from the package `lme4` (Bates et al. 2015) using the syntax outlined in Table 7.3. Additional information on using this function can be accessed by running `?glmer`. Here we use the default Laplace Approximation to fit the candidate models. Although less accurate in estimating parameters than the Gauss Hermite Quadrature (both are common choices), Laplace Approximation is less computationally-demanding and

Table 7.2 Candidate model set for analysis.

m1 = Survival ~ treatment + asynchrony + hatch + relhatch
 m2 = Survival ~ relhatch + asynchrony
 m3 = Survival ~ relhatch + asynchrony + treatment
 m4 = Survival ~ relhatch * asynchrony + treatment
 m5 = Survival ~ hatch order + relhatch
 m6 = Survival ~ hatch order:relhatch + treatment
 m7 = Survival ~ treatment * hatch order
 m8 = Survival ~ treatment * relhatch
 m9 = Survival ~ treatment * asynchrony
 m10 = Survival ~ hatch order + treatment
 m11 = Survival ~ relhatch + hatch order + treatment
 m12 = Survival ~ asynchrony + treatment
 m13 = Survival ~ 1

Table 7.3 Various ways in which fixed and random variables can be coded in the package lme4

Fixed Effects	TREATMENT	Variable with no interaction
	TREATMENT : ASYNCHRONY	Specifies an interaction between these two variables
	TREATMENT * ASYNCHRONY = TREATMENT : ASYNCHRONY + TREATMENT + ASYNCHRONY	Indicates both an interaction and the two variables without an interaction.
Random Effects	(1 NEST)	Random intercept at the nest level
	(1 YEAR/NEST)= (1 YEAR)+(1 YEAR:NEST)	Nested random effects with an intercept varying among years, and among nests within years.
	... + YEAR + (1 YEAR:NEST)	Year specified as fixed effect due to insufficient levels.
	(1 REGION) + (1 NEST)	When random variables are crossed rather than nested (such as nests located in a specific region)

is sufficient for our purposes (see Bolker et al. 2009 for a summary of all methods used to fit glms). Additionally, we have included the bobyqa optimizer to suppress warnings of non-convergence.

To rank our models, we begin by fitting all the models in the candidate list (see Table 7.2).

```
# fit candidate models
mod1 <- glmer(SURVIVAL ~ TREATMENT + ASYNCHRONY + HATCH +
  RELHATCH + YEAR + (1|YEAR:NESTYEAR), dat,
  family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod2 <- glmer(SURVIVAL ~ RELHATCH + ASYNCHRONY +YEAR +
  (1|YEAR:NESTYEAR), dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod3 <- glmer(SURVIVAL ~ RELHATCH + ASYNCHRONY +
  TREATMENT + YEAR + (1|YEAR:NESTYEAR), dat,
  family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod4 <- glmer(SURVIVAL ~ RELHATCH:ASYNCHRONY + TREATMENT+
  YEAR + (1|YEAR:NESTYEAR), dat, family =
  binomial(logit), glmerControl(optimizer =
  "bobyqa"))
mod5 <- glmer(SURVIVAL ~ HATCH + RELHATCH +YEAR +
  (1|YEAR:NESTYEAR), dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod6 <- glmer(SURVIVAL ~ RELHATCH:HATCH + TREATMENT +YEAR +
  (1|YEAR:NESTYEAR), dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod7 <- glmer(SURVIVAL ~ TREATMENT*HATCH +YEAR +
  (1|YEAR:NESTYEAR), dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod8 <- glmer(SURVIVAL ~ TREATMENT*RELHATCH +YEAR +
  (1|YEAR:NESTYEAR), dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod9 <- glmer(SURVIVAL ~ TREATMENT*ASYNCHRONY + YEAR +
  (1|YEAR:NESTYEAR), dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod10 <- glmer(SURVIVAL ~ HATCH + TREATMENT + YEAR +
  (1|YEAR:NESTYEAR),dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))
mod11 <- glmer(SURVIVAL ~ RELHATCH + TREATMENT + HATCH +
  YEAR + (1|YEAR:NESTYEAR), dat, family =
  binomial(logit), glmerControl(optimizer =
  "bobyqa"))
```



```

mod12 <- glmer(SURVIVAL ~ ASYNCHRONY + TREATMENT + YEAR +
  (1|YEAR:NESTYEAR), dat, family =
  binomial(logit), glmerControl(optimizer =
  "bobyqa"))
mod13 <- glmer(SURVIVAL ~ 1 + YEAR + (1|YEAR:NESTYEAR),
  dat, family = binomial(logit),
  glmerControl(optimizer = "bobyqa"))

```

Once we have fitted our candidate models, we can use the `model.sel` function from the package MuMIn (Barton 2015) to rank our models by their second-order Akaike Information Criterion scores (AIC, Akaike 1974) corrected for small sample size (AICc, Hurvich and Tsai 1989). Founded on information theory, AIC rewards models for fit, and penalizes them for each additional parameter. This results in a measure of quality that balances goodness of fit and parsimony (Burnham and Anderson 2002). It may be true that a model with every possible factor will fit the data best, however this may lead to overparameterization and poor predictive ability (Burnham and Anderson 2002).

store models in a list

```

out.put<-model.sel(mod1,mod2,mod3,mod4,mod5,mod6,mod7,mod8,
  mod9,mod10,mod11,mod12,mod13)

```

build a model selection table

```

sel.table <-as.data.frame(out.put)[12:16]
(sel.table <- round(sel.table, digits = 2))

```

output

	df	logLik	AICc	delta	weight
mod11	10	-119.16	259.18	0.00	0.52
mod1	11	-118.33	259.69	0.51	0.40
mod3	8	-123.50	263.55	4.38	0.06
mod10	9	-123.70	266.10	6.92	0.02
mod12	7	-126.72	267.86	8.69	0.01
mod9	8	-126.35	269.25	10.08	0.00
mod7	12	-123.37	271.95	12.78	0.00
mod5	9	-130.49	279.67	20.49	0.00
mod6	10	-129.89	280.63	21.45	0.00
mod2	7	-134.43	283.29	24.12	0.00
mod8	8	-133.58	283.72	24.54	0.00
mod4	7	-136.47	287.37	28.20	0.00
mod13	5	-157.78	325.79	66.61	0.00

To properly interpret the summary table, it is important to understand the values in each column. In column one we simply have the model names, which correspond to the order in which we entered them into our list. In column two we can see the number of parameters (normally referred to as “ K ”, but here as “ df ”), followed in column three by the log likelihood, which is a measure of how well the model fit the data (higher numbers indicate better fit). Column four shows AICc, which is a calculation that is driven by log likelihood, but adds a component that penalizes models for every additional parameter used. Column five shows delta (short for Δ AICc). This is simply the AICc score of the respective model, minus that of the best (most parsimonious) model. This column enables us to easily compare all models to the highest ranked model. Finally, column six shows the AICc weights, which tell us the likelihood that the respective model is the best in the set.

We see model 11 [survival ~ relhatch + hatch order + treatment + year] ranked highest with an AICc score of 259.18. This model was followed closely by our global model with all of the predictor variables. If we look at the log likelihood of these two top models, we can see that our second best model actually fits our data better (-118.33 vs. -119.16), but received a penalized AICc score for having an extra parameter. The extra parameter included in model 1 therefore improved model fit, but not enough to justify its inclusion. The importance of our predictor variables is apparent due to the lowest rank assigned to null model. When looking at the delta (Δ)AICc and AICc weights of our top models, we can now understand the value of using an information theoretic approach. As is the case with most ecological processes, predicting nestling survival is complex and the best approximation of this process seems to lie somewhere between two of our candidate models. According to AICc, the top two models are nearly equally as likely (AICc weights of 0.52 and 0.40 respectively), and therefore we will want to average parameter estimates between these two.

Before we move on to parameter estimates, it is important to realize that all we can conclude so far is that model 11 is the best model *within our candidate set*; this says nothing about how the model actually fits the data. For linear models, we can calculate the coefficient of determination (R^2) to determine how much of the total variance is explained by the predictor variables, but this calculation is challenging to implement with mixed models due to the inclusion of random effects and link functions. Nakagawa and Schielzeth (2013) have provided a method of calculating a similar coefficient for GLMMs (termed R^2_{GLMM}) that is implemented by the `r.squaredGLMM` function again using the MuMIn package (Barton 2015).

```

# R2GLMM estimates
r.squaredGLMM(mod11)
# output
      R2m      R2c
0.4391978 0.7355235

r.squaredGLMM(mod1)
# output
      R2m      R2c
0.4400974 0.7328579

r.squaredGLMM(mod3)
# output
      R2m      R2c
0.4138954 0.6909614

```

The `r.squaredGLMM` function returns two numbers. The first is the marginal R^2_{GLMM} which reports the amount of variance explained by the fixed effects portion in our model, and the second is the conditional R^2_{GLMM} , or the total variance explained by both the fixed and the random effects of our model. These values range from 0 to 1, and give us an objective measure of how our models perform. Not surprisingly, there still seems to be little difference between our top models, which provides further evidence that parameter estimates should be averaged.

7.3.3 Drawing inference from parameter estimates

After we choose the model(s) that best predict nestling survival, we investigate the relative relationship of each predictor variable to survival. Because our top two models were equally parsimonious we will begin by model averaging our estimates. Model averaging produces parameter and error estimates that are weighted averages of the values derived from specified models using AICc weights. Depending on the AICc weights of the top models, we can model average our estimates in one of two ways. If support for one model is strong but not indisputable (i.e., AICc weight >0.90), subset model averaging, where only parameter estimates for variables in the top model are averaged, is preferred (Burnham and Anderson 2002, Lukacs et al. 2010, Symonds and Moussalli 2011). The second approach is full model averaging, and is used in cases where evidence of support is spread across a number of top models. In this case, inference is drawn from coefficient estimates of all models within the candidate set and coefficients are weighted by their AICc weight. Details about how these two calculations differ can be found in Lukacs et al. (2010).

Although support is not unequivocal for our top model, the combined weights for our top two is high (cum.wt = 0.92). For this reason, we will use subset model averaging using the `model.avg` function in the package MuMIn.

```
# averaged Parameter Estimates below a specified AICc
# threshold
avgmod.95 <- model.avg(out.put, subset = delta < 2,
  revised.var = TRUE)
modavg<-as.data.frame(avgmod.95["coefficients"])
modavg <- data.frame(modavg[2,])
modavg <- data.frame(t(modavg))
conf<-as.data.frame(confint(avgmod.95))
(modavg <- cbind(modavg,conf))

# output
```

	subset	2.5 %	97.5 %
coefficients..Intercept.	2.7994637	1.1049068	4.49402066
coefficients.RELHATCH	-0.2668755	-0.4812351	-0.05251593
coefficients.TREATMENTS	2.9618739	1.5377233	4.38602461
coefficients.HATCH2	-1.4056421	-2.5291693	-0.28211498
coefficients.HATCH3	-1.7933147	-3.1497185	-0.43691092
coefficients.HATCH4	-3.4205161	-5.4713535	-1.36967861
coefficients.YEAR2014	-1.5739238	-3.3332828	0.18543526
coefficients.YEAR2015	-1.0131320	-2.8596084	0.83334446
coefficients.YEAR2016	-1.6376443	-3.4225639	0.14727531
coefficients.ASYNCHRONY	-0.3457474	-0.8814261	0.18993130

Because our link function is logit (i.e., the calculation required to relate our predictor variables to response distribution, see Bolker et al. 2009), estimates reported here are Beta values, or slopes of each of our fixed effects in log odds. This means that for every unit of increase in the value of a parameter, the log odds of nestling survival changes by the respective estimate. If we look at our relative hatch date for example, a nestling's log odds of survival decreases by 0.27 for additional day. For the variable treatment we can see that as the level of this variable changes from control to supplemented, the log odds of survival increases to 2.96. If we look at the categorical variables HATCH and YEAR, we can see that each level of the respective variable has an estimate. This is because each level is being compared to a reference level. HATCH2, for instance, has an estimate of -1.41, which indicates that the log odds of survival will change by that amount relative to the reference (in this case first hatch). Categorical variables such as HATCH and YEAR lack an estimate value at their reference levels, because survival at these lev-

els is accounted for within the intercept. We can therefore see that nestling survival at the reference levels (e.g., first hatched, no food supplementation, hatched in the year 2013, and hatched at the mean date within the population) was relatively high: 2.80.

Model averaged estimates are more accurate than those generated from the top model alone but model averaging did not result in considerable change in parameter estimates—perhaps because the second-best model contained an uninformative parameter (see Arnold 2010 for more details). Therefore, we use only the top model to visualize our estimates. Here we use the `visreg` package (Breheny and Burchett 2013) to plot our fixed effects as they vary throughout their levels (Fig. 7.2).

```
# top model summary
data.frame(Estimate = round(fixef(mod1), 3))
# output
```

	Estimate
(Intercept)	2.717
RELHATCH	-0.285
TREATMENTS	2.966
HATCH2	-1.489
HATCH3	-2.004
HATCH4	-3.790
YEAR2014	-1.487
YEAR2015	-0.902
YEAR2016	-1.558

```
# visualize parameter estimates
graphics.off()
windows(6,10)
par(mar=c(0,2,2,0),mai=c(0.3,0.3,0.3,0.3),
    mfrow=c(3,2),mgp=c(1.5,0.5,0))
visreg(mod1, type = "conditional",
    cex.lab = 1, cex.axis=0.75,
    pch=19,
    font.lab=2,
    bty="l",
    tck=0.01)
```

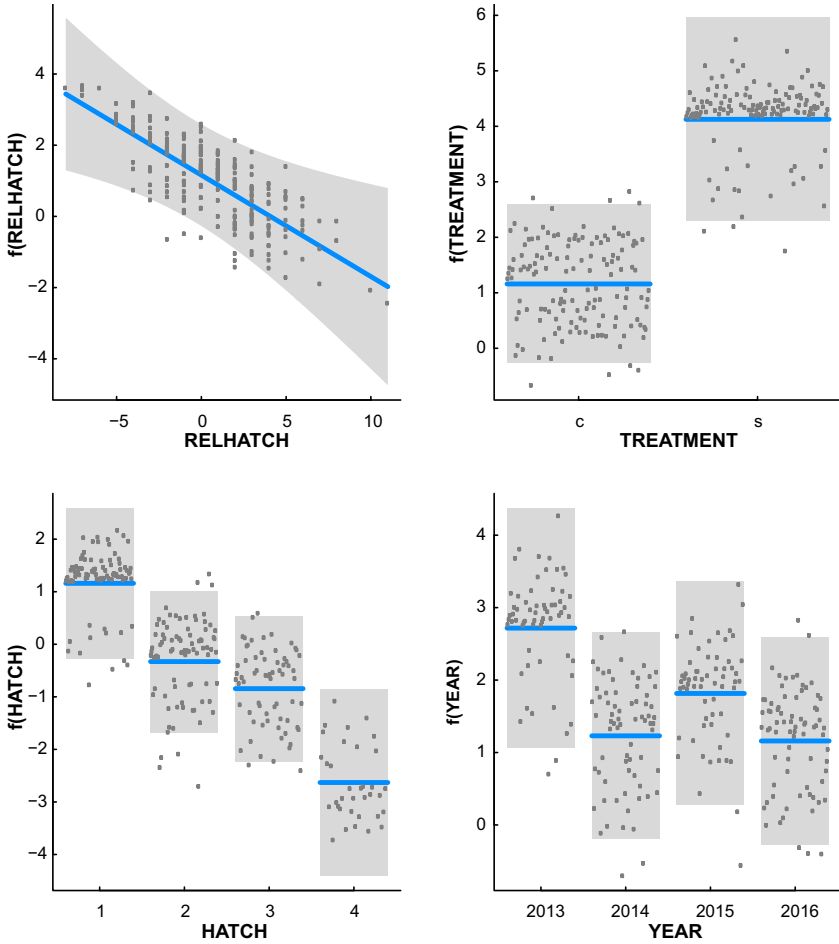


Figure 7.2. A display of the parameter estimates for the four fixed variables from our top model. Here we can see the log odds of nestling survival (y axis) across the range of each variable (x axis).

We can now draw inferences with regard to nestling survival within our population. We see that hatching early provides a substantial advantage in nestling survival, a pattern that is consistent across a variety of bird species (Verhulst and Nilsson 2008). We can see that food supplementation causes a considerable increase in the log odds of survival, which provides evidence for food limitation during the brood rearing period. We can also see that although there is little difference in survival between hatching second and third, there is a large negative difference between hatching first and last. Finally, we can see the variation in survival among the four years in which data were collected. Although variation in survival among years was not related to our research questions, it was correct to include year as a fixed effect because it explained variation in nestling survival in our study.

The objectives of this analysis were to firstly determine which factors influence nestling survival in a population of Peregrine Falcons near Rankin Inlet, Nunavut, Canada, and to investigate how these factors relate to survival. Secondly, we were interested in testing whether food was limiting during the brood rearing period. Given the positive parameter estimates of supplementation within our top model, we can conclude that evidence exists to show that food is indeed a limiting factor.

So far it has been convenient to present our estimates as log odds because: 1) no conversion is required because we are using a logit link function, and 2) relationships are easy to visualize in log odds because it is linearized (i.e., a one unit difference in the predictor equals one unit change in the log odds estimate), but it may be easier to view our parameter estimates in real world scenarios if we convert them to survival probability. Using the section of code below (UCLA Statistical Consulting Group: 2017), we use our top model to investigate how the predicted probability of survival differs for control and supplemented individuals across hatch order and relative hatch date. This involves letting survival vary across relative hatch date for both of our experimental groups, while holding hatch order constant at each level. This exercise is complicated by the fact that our random effect is also influencing survival. To include this group level variation, we average the change in survival probability across our groups (nests) as they progress from early to late hatch dates (UCLA Statistical Consulting Group: 2017). After we subset and plot these marginally averaged survival probabilities by hatch order, we produce a figure that depicts the combined effects of hatch order, food supplementation, and hatch date on survival (Fig. 7.3).

```

# get predicted probabilities across hatch and relhatch
# Predicted control probabilities
control<-subset(dat,TREATMENT=="c")
tmpdatc <- control[, c("RELHATCH", "HATCH", "NESTYEAR",
  "YEAR")]
jvaluesc<- with(control, seq(from = min(RELHATCH), to =
  max(RELHATCH)))
mc <-glmer(SURVIVAL ~ RELHATCH + HATCH + YEAR +
  (1|YEAR:NESTYEAR), dat, family=binomial(logit),
  glmerControl(optimizer = "bobyqa"))

# calculate predicted probabilities and store in a list
biprobsc <- lapply(levels(control$HATCH),
  function(stage) {
    tmpdatc$HATCH[] <- stage
    lapply(jvaluesc, function(j) {
      tmpdatc$RELHATCH <- j
      predict(mc, newdata = tmpdatc, type = "response")
    })
  })

# get means and quartiles for all jvalues for each level of
# hatchorder
plotdat2c <- lapply(biprobsc, function(X) {
  tempc <- t(sapply(X, function(x) {
    c(M=mean(x), quantile(x, c(.25, .75)))
  })))
  tempc <- as.data.frame(cbind(tempc, jvaluesc))
  colnames(tempc) <- c("PredictedProbability", "Lower",
    "Upper", "RELHATCH")
  return(tempc)
})
plotdat2c <- do.call(rbind, plotdat2c)
plotdat2c$HATCH <- factor(rep(levels(control$HATCH),
  each = length(jvaluesc)))

# See online for supplemented code

# build the plot
# set graphical parameters
graphics.off()
windows(6.5,6)
par(mar=c(0,1,1,0),oma=c(5,5,3,3),mfrow=c(2,2),mgp=c(1.5,
  0.5,0))

```



```

# Plot probabilities for FIRST hatch individuals
plotdat2c1<-subset(plotdat2c,HATCH=="1")
plotdat2s1<-subset(plotdat2s,HATCH=="1")

plot(plotdat2c1$PredictedProbability~plotdat2c1$RELHATCH,
     pch = 16,col = rgb(1,.5, .5,0.0),
     bty = "l",cex = 2,xaxt = "n",yaxt = "n",las = 1,
     xlim=c(-7,11),ylim=c(0.0,1.0))
lines(plotdat2c1$PredictedProbability ~
      plotdat2c1$RELHATCH,lwd = 1.5,lty = 2,col =
      rgb(0,0,0,1))
polygon(c(rev(plotdat2c1$RELHATCH),plotdat2c1$RELHATCH),
       c(rev(plotdat2c1$Lower),plotdat2c1$Upper),
       col = rgb(0,0,0,0.05),border = NA)
lines(plotdat2s1$PredictedProbability ~
      plotdat2s1$RELHATCH,lwd = 1.5,col = rgb(0.0,0,0,1))
polygon(c(rev(plotdat2s1$RELHATCH),plotdat2s1$RELHATCH),
       c(rev(plotdat2s1$Lower),plotdat2s1$Upper),
       col=rgb(0,0,0,0.05),border = NA)
axis(1, at = seq(-7,11,2),tcl = -0.3,lwd = 1, tck =
      0.01,labels = FALSE,cex.axis = 0.9)
axis(2,at = seq(0,1,0.2),tcl = -0.3,las = 1,lwd = 1,
      tck = 0.01, cex.axis = 0.95)
legend(-6.4,0.3,cex = 0.8,legend = " Supplemented",bty =
      "n",lty = 1,lwd = 1,col = c(rgb(0,0,0,1)))
legend(-6.4,0.25,cex = 0.8,legend = "Control",bty =
      "n",lty = 2,lwd = 1,col =c (rgb(0,0,0,1)))
text(8,0.95,cex=1,labels="a",bty="n")
# see online for second, third and fourth hatch # visuals

```

From Fig. 7.3 we can see that for both control and supplemented individuals at all hatch positions, survival starts relatively high at early hatch dates. Survival probabilities then decline as individuals hatch progressively later in the season. By plotting marginal survival probabilities, we can visualize how supplemental food increases survival across the range of dates and hatch orders. We can see that as individuals hatch later within the brood and later in the season, the difference between supplemented and control individuals gradually decreases. One might conclude that the accumulated consequences of hatching late begin to overwhelm the buffering effect of supplemental food. In other words, there is little difference between the supplemented and control nestlings that hatch both last in the brood and late in the season because the supplemental food is not enough to overcome the disadvantage of hatching last. This implies a

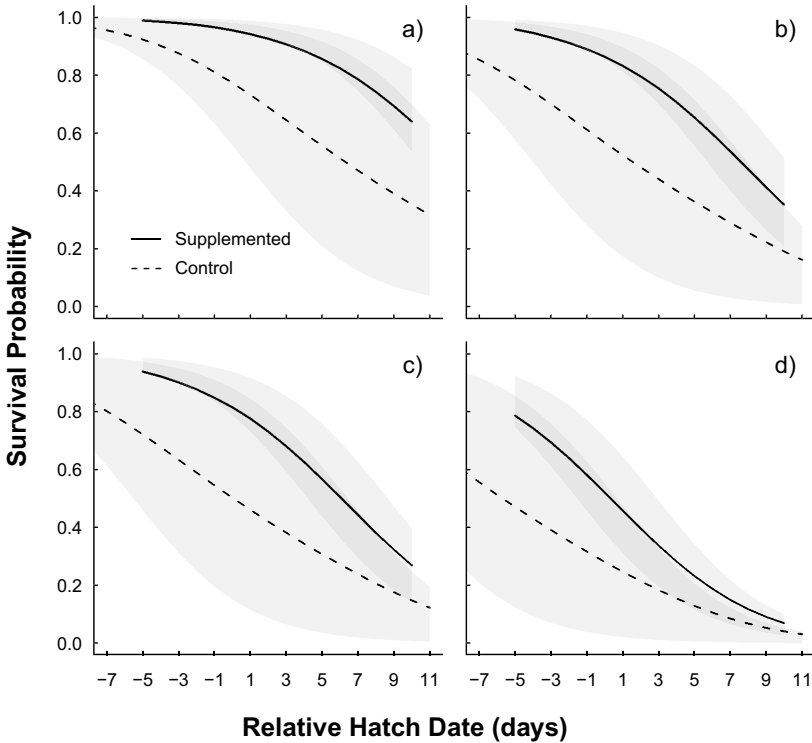


Figure 7.3. A comparison of marginal survival probabilities between control and supplemented nestlings across the range of hatch order, and relative hatch date for all years combined. Hatch order is indicated alphabetically starting at Hatch 1 = a, and ending with hatch 4 = d.

strong relationship between hatch date and food limitation, something that should be explored with further work.

We conclude this chapter by reiterating that although mixed models are becoming a widely used tool, they are complex and difficult to implement correctly. We introduced a relatively simple example of how they can be used to address hierarchical variation in nestling survival analysis, but we cannot possibly cover the full range of possibilities for these models. We recommend further reading to explore the concepts introduced here (Pinheiro and Bates 2000, Gelman and Hill 2007, Jiang 2007, Bolker 2008, Zuur et al. 2009).

Literature cited

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* AC-19:716–723.
- Anctil, A., and A. Franke. 2013. Intraspecific adoption and double nest switching in Peregrine falcons (*Falco peregrinus*). *Arctic* 66: 222–225.
- Anctil, A., A. Franke, and J. Bêty. 2014. Heavy rainfall increases nestling mortality of an arctic top predator: Experimental evidence and long-term trend in peregrine falcons. *Oecologia* 174:1033–1043.
- Arnold, T. W. 2010. Uninformative parameters and model selection using Akaike's Information Criterion. *Journal of Wildlife Management* 74:1175–1178.
- Barton, K. 2015. MuMIn: Multi-model inference. R package version 1.9.13. Version 1:18.
- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Package lme4. *Journal Of Statistical Software* 67:1–91.
- Bolker, B. M. 2008. *Ecological Models and Data in R*. Princeton University Press, Princeton, New Jersey, USA.
- Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24:127–135.
- Bradley, M., R. Johnstone, G. Court, and T. Duncan. 1997. Influence of Weather on Breeding Success of Peregrine Falcons in the Arctic. *Source. Auk* 114:786–791.
- Breheny, P., and W. Burchett. 2013. Visualization of regression models using visreg. R package:1–15.
- Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach*, Second Edition. Springer-Verlag, New York, New York, USA.
- Caughley, G. 1977. *Analysis of vertebrate populations*. John Wiley and Sons, London, UK.
- Court, G. S., D. M. Bradley, C. C. Gates, and D. A. Boag. 1989. Turnover and recruitment in a tundra population of Peregrine Falcons *Falco peregrinus*. *Ibis* 131:487–496.
- Court, G. S., C. C. Gates, D. A. Boag, J. D. Macneil, D. M. Bradley, A. C. Fesser, J. R. Patterson, G. B. Stenhouse, and L. W. Oliphant. 1990. A toxicological assessment of Peregrine Falcons, *Falco peregrinus tundrius*, Breeding in the Keewatin District of the Northwest-Territories, Canada. *Canadian Field Naturalist* 104:255–272.
- Crawley, M. 2002. *Statistical Computing: An Introduction to Data Analysis Using S-PLUS*. Wiley, Chichester, West Sussex, UK.
- Franke, A., M. Settingington, G. Court, and D. Birkholz. 2010. Long-term trends of persistent organochlorine pollutants, occupancy and repro-

- ductive success in Peregrine Falcons (*Falco peregrinus tundrius*) breeding near Rankin Inlet, Nunavut, Canada. *Arctic* 63:442–450.
- Gelman, A. 2005. Analysis of variance — why it is more important than ever. *The Annals of Statistics* 33:1–53.
- Gelman, A., and J. Hill. 2006. *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press, Cambridge, UK.
- Hurvich, C. M., and C. L. Tsai. 1989. Regression and time-series model selection in small sample sizes. *Biometrika* 76:297–307.
- Jian, J. 2007. *Linear and generalized linear mixed models and their applications*. Springer, New York, New York, USA.
- Johnstone, R. M., G. S. Court, A. C. Fesser, D. M. Bradley, L. W. Oliphant, and J. D. MacNeil. 1996. Long-term trends and sources of organochlorine contamination in Canadian Tundra Peregrine Falcons, *Falco peregrinus tundrius*. *Environmental Pollution* 93:109–120.
- Kreft, I., and J. De Leeuw. 1998. *Introducing Multilevel Modeling*. Sage Publications, London, UK.
- L'Hérault, V., A. Franke, N. Lecomte, A. Alogut, and J. Bêty. 2013. Landscape heterogeneity drives intra-population niche variation and reproduction in an arctic top predator. *Ecology and Evolution* 3:2867–2879.
- Lukacs, P. M., K. P. Burnham, and D. R. Anderson. 2010. Model selection bias and Freedman's paradox. *Annals of the Institute of Statistical Mathematics* 62:117–125.
- Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4:133–142.
- Pinheiro, J. C., and D. M. Bates. 2000. *Mixed effects models in S and S-Plus*. Springer, New York, New York, USA.
- Schabenberger, O., and P. J. Francis. 2001. *Contemporary statistical models for the plant and soil sciences*. CRC Press, Boca Raton, Florida, USA.
- Symonds, M. R. E., and A. Moussalli. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* 65:13–21.
- UCLA, U. S. C. (n.d.). *R Data Analysis Examples: Mixed Effects Logistic Regression*. <http://www.ats.ucla.edu/stat/r/dae/melogit.htm>.
- Verhulst, S., and J.-A. Nilsson. 2008. The timing of birds' breeding seasons: a review of experiments that manipulated timing of breeding. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* 363:399–410.
- Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York, New York, USA.