

Using *R* for Analyzing Delay Discounting Choice Data

Logistic regression is available in a wide range of statistical software packages, but the analysis of discounting choice data requires the use of tools that allow for repeated measures logistic regression to capture the differential dependencies within and between subjects.

Performing such an analysis requires generalized multilevel modeling which allows both the specification of a binomial outcome as well as defining the random effect structure (a technique for modeling data dependencies). I will present a series of examples using the lme4 package in *R*, because of its broad accessibility at no cost.

The following analyses all assume that the data is in long format where each choice is in a separate row. In my examples, the columns in the data typically include Subject, Trial, SSdelay, SSmagnitude, LLdelay, LLmagnitude, and Waited (the latter is coded 1 if the subject chose the LL and 0 if the SS was chosen). Additional variables could specify day, session, condition, IQ, etc. The basic syntax for fitting repeated measures logistic regression in *R* is illustrated in the following commands:

```
myr<-glmer(Waited ~ magnitudeRatio + delayRatio - 1 +
(magnitudeRatio + delayRatio - 1 | Subject),
data=myd, family=binomial)

myr<-glmer(Waited ~ log(LLmagn/SSmagn) +
log((LLdelay+1)/(SSdelay+1)) - 1 +
(log(LLmagn/SSmagn) + log((LLdelay+1)/(SSdelay+1)) - 1 |
Subject), data=myd, family=binomial)
```

The first command provides the basic framework, but its use would necessitate the creation of two new columns for each row in which the magnitude and delay ratios were computed (and transformed, if desired). The second command computes the ratios and their log transform as

part of the fit. I will unpack the first command due to its greater simplicity, but the same principles apply in the other variations encountered.

The `glmer` command (typically pronounced “glimmer”) fits a generalized linear mixed effects model in *R* (multilevel models can also be called mixed effects models). The dependent variable, `Waited`, is followed by the fixed effects specification, `magnitudeRatio + delayRatio - 1`. The command will compute a regression weight for each of the two predictors in the regression equation but will omit the intercept (the minus one part of the specification). Typical applications of regression would not exclude the intercept, but its exclusion forces the logistic equation to produce a 0.5 probability of waiting when the prediction equation sums to zero. The equation sums to zero only when the predictors are zero which would never occur for the first *R* command (because the ratios could only be zero if the LL magnitude and LL delay were both zero); but, the equation would sum to zero for the second *R* command when the choice alternatives have the same magnitudes and delays because each ratio would be 1.0 and $\log(1.0) = 0.0$. Thus, I recommend using log-transformed ratios and the omission of the intercept. The two regression weights would then specify the overall sensitivity to the magnitude and delay ratios, respectively, for each subject and the group as a whole.

The next part of the command “`(magnitudeRatio + delayRatio - 1 | Subject)`” specifies the random effect structure of the model. In multilevel modeling, the structure specified here indicates that the individual subjects’ regression weights for the magnitude and delay ratios are assumed to randomly vary around the group’s estimate (the fixed effect estimates). It is again necessary to exclude the intercept because otherwise the model would fit intercepts for each individual subject even though the group’s intercept is assumed to be zero (corresponding to indifference when the ratios equal one in the log version). The specification of the random

effects tends to be the poorest understood aspect of a multilevel model because it is possible to use structures in which each regression weight is *not* allowed to vary across subjects. In general, if a model is greatly improved (as judged by a lower AIC) by including a variable as a random effect, then it should be included in order to capture the correct dependencies among the data. However, overly complex random effect structures can create problems with model convergence and may not replicate due to overfitting. Only within-subject variables can be considered as random effects, and interactions between these variables can also be included as random effects assuming that there is adequate sampling of the joint parameter space at the individual subject level (McClelland & Judd, 1993).

The final two parts of the model specification are straightforward. The “data = myd” simply allows the user to specify the variable that contains the data set (in this case “myd”) which was previously populated using a read statement; see the Supplemental Materials for examples. The “family=binomial” part of the command indicates that the outcome variable, Waited, is binomial thus forcing the model to fit the data using a logistic regression.

To obtain the results of the fit, the user simply types “summary(myr)” where *myr* is the variable that stores the results of the models (in the examples above, the “myr<-” part of the command is storing the results in the *myr* variable). An example of the critical part of the summary is shown below:

```
AIC      BIC      logLik      deviance      df.resid
9985.2  10024.0 -4987.6    9975.2       17231

Fixed effects:
                                         Estimate      SE   z value Pr(>|z|)
log(LLmagn/SSmagn)                  7.3663  0.6049 12.178  <2e-16 ***
log((LLdelay + 1)/(SSdelay + 1)) -6.7511  0.6933 -9.737  <2e-16 ***
```

The AIC, etc., row provides a series of fit indices for the model. The AIC and BIC metrics are on an arbitrary scale and only interpretable when compared to the AIC/BIC of other models of the

same data (i.e., same rows of data), the same outcome variable (i.e., not transformed variations), and the same family (here, binomial). AIC and BIC penalize models based on their complexity (like adjusted R^2), and lower values are better. A difference of 2 or less is inconsequential, differences between 2 and 6 provide modest evidence in favor of the model with the lower AIC/BIC, differences greater than 6 represent strong evidence, and greater than 10 very strong evidence (Raftery, 1995).

Scientists interested in the individual differences in these parameters can obtain the individual regressions weights using the “coef(myr)” command. The first five rows from this data set are shown below:

	<code>log(LLmagn/SSmagn)</code>	<code>log((LLdelay + 1)/(SSdelay + 1))</code>
1	5.878101	-1.355578
2	6.087047	-1.769283
3	6.525303	-1.864917
4	2.945848	-4.011453
5	6.677519	-7.725642

When individual differences are observed or when behavior changes across trials, researchers are often interested in what is causing these effects. Thus, additional predictors can be added to the regression model to attempt to account for between-subject differences or within-subject changes. For example, when another data set was fit by including age (centered) as a moderator of the sensitivity to the magnitude ratio and to delay (the SS delay was always zero in this study), the fixed effects estimates were:

Fixed effects:						
	Estimate	Std. Error	z value	Pr(> z)		
<code>log(LLmagn/SSmagn)</code>	14.600012	0.947377	15.411	<2e-16	***	
<code>log(LLdelay)</code>	-1.463772	0.093580	-15.642	<2e-16	***	
<code>c.age:log(LLmagn/SSmagn)</code>	0.150579	0.069216	2.175	0.0296	*	
<code>c.age:log(LLdelay)</code>	-0.004282	0.006516	-0.657	0.5111		

The Wald Z tests indicate that there is modest evidence suggesting that the sensitivity to the magnitude ratio is increasing across the lifespan. Another approach is to use a likelihood ratio

(LR) test of this model versus a simpler model excluding the age interactions; this test was not significant. However, Pinheiro and Bates (2004) provide evidence that the LR test is unreliable when comparing multilevel models with different fixed effects except when the sample size is large, so the Wald Z is generally preferred for this type of model testing (note, however, that the LR is preferred when testing models that differ only in their random effects, Bolker et al., 2008). The Supplemental Material includes multiple examples to help navigate the syntax for running various models.