PSYC402-week-20-GLMM

Rob Davies (Lancaster University)

Targets for Week 20

1. Understand the reasons for using Generalized Linear Mixed-effects models (GLMMs) when we analyze discrete outcome variables
2. Recognize the limitations of alternative methods for analyzing such outcomes
3. Practice running GLMMs with varying random effects structures
4. Practice reporting the results of GLMMs, including through the use of model plots

We want to continue to evelop the capacity to understand mixed-effects models

1. Recognize where data have a multilevel structure
   - In Week 20, the structure comes from repeated measures and longitudinal elements of the study design
2. Recognize where multilevel or mixed-effects models are required
3. Distinguish the elements of a mixed-effects model, including fixed effects and random effects

Develop the capacity to work practically in R with mixed-effects models, to:

1. Be able to specify a mixed-effects model in glmer() code
2. Be able to compare and evaluate alternate model specifications

Develop the capacity to talk about and present the results, to:

1. Be able to describe in words and summary tables the results of a GLMM
2. Be able to visualize the effects estimates from a GLMM

The key idea to get us started

Categorical outcomes cannot be analyzed using linear models (ANOVA or t-test or linear models or linear mixed-effects models) without having to make some important compromises.

Discrete or categorical outcome variables

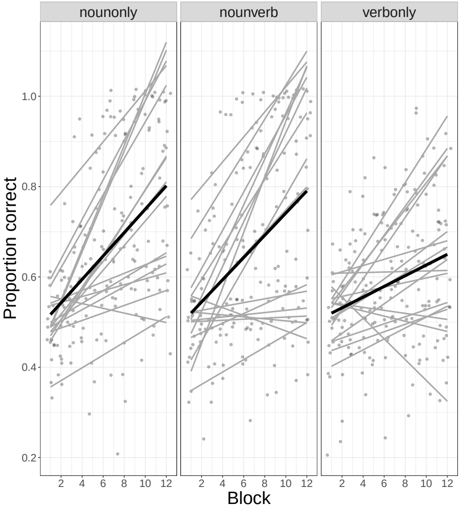
- The accuracy of responses: is a response correct or incorrect?
- The membership of one group out of two groups: e.g., is a participant impaired or unimpaired; e.g., was a recorded eye movement, a fixation, to the left or to the right visual field?
- Responses that can be coded in terms of ordered categories: e.g., a response on a (Likert) ratings scale
- Also, outcomes like membership of one group out of multiple groups (categories): e.g., is a participant in one of several groups like religious or ethnic or degree class group?
- Also, outcomes like frequency of occurrence of an event, e.g., how many arrests are made at a particular city location?

Recognize the limitations of alternative methods for analyzing response accuracy

- The accuracy of responses (correct vs. incorrect) is counted, e.g., as the number of correct responses (or errors) per subject, for each level of each condition or factor;
- The raw number of correct or incorrect responses, or the percentage, or the proportion of responses that are correct or incorrect out of the total number of responses is analyzed as the outcome variable in ANOVA or regression.

Accuracy is bounded between 1 and 0, parametric model predictions or confidence intervals are not

- Monaghan et al. (2015) artificial word learning study
- Plot showing the proportion of responses correct for each participant
- In each of 12 blocks of 24 learning trials, in each learning condition
- Each grey line shows the linear model prediction of the proportion correct, for each person, by learning block, in each condition



ANOVA or regression require the assumption of homogeneity of variance but for binary outcomes like accuracy the variance is proportional to the mean

- Given a binary outcome e.g. response is correct or incorrect
- For every trial, there is a probability p that the response is correct
- The variance of the proportion of trials (per condition) with correct responses is dependent on p and greater when $p \sim .5$, the probability that a response will be correct

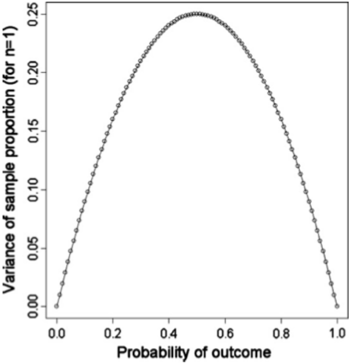


Figure 1: Jaeger (2008) variance of sample varies

Summary: Limitations of traditional methods

- Linear models assume outcomes are unbounded so allow predictions that are impossible when outcomes are, in fact, bounded as is the case for accuracy or other categorical variables
- Linear models assume homogeneity of variance but that is unlikely and anyway cannot be predicted in advance when outcomes are categorical variables
- If we are interested in the effect of an interaction between two effects, using ANOVA or linear models on accuracy (proportions of responses correct) can tell you, wrongly, that the interaction is significant

Understanding the Generalized part of the Generalized Linear Mixed-effects Models in practical terms

Understand GLMMs as analyses where the outcome variable is transformed

- Transforming a probability to odds $o = \frac{p}{1-p}$ is a partial solution
  - Odds are, for example, the ratio of the probability of the response being correct compared to the probability of the response being incorrect
  - And odds are continuous numeric quantities that are scaled from zero to infinity.
- We can then use the (natural) logarithm of the odds $logit = ln\frac{p}{1-p}$
  - Because using the logarithm removes the boundary at zero because log odds ranges from negative to positive infinity.

Understanding the Generalized part of the Generalized Linear Mixed-effects Models in practical terms

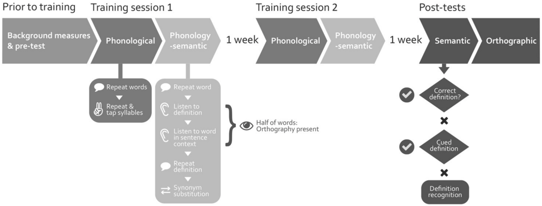
We can think of logistic models as working like linear models with log-odds outcomes

$$ln\frac{p}{1-p} = logitp = \beta_0 + \beta_1 X_1 \ldots \tag{1}$$

- We can describe the predicted log odds of a response of one type as the sum of the effects
- log odds range from negative to positive infinity (logit of 0 corresponds to proportion of .5)

The data we will work with: the word learning study

- Ricketts, Dawson, and Davies (in press) investigated how school-aged children learn words
  - Half of the words were taught with access to the orthographic form (orthography present condition) and the other half were taught without orthographic forms (orthography absent condition)
  - About half of the children were told that some words would appear with their written form (explicit group); the remaining children did not receive these instructions (incidental group)
- Looked at post-(intervention)-test measures of knowledge at 2 time points: including new word spelling test



The data we will work with: the word learning study

```r
head(long.orth)
```

```
## # A tibble: 6 x 30
##   Participant Time  Study  Instructions Version Word   Consistency_H Orthography
##   <fct>       <fct> <fct>  <fct>        <fct>   <fct>          <dbl> <fct>
## 1 EOF001      1     Study~ explicit     a       Accol~          1.91 absent
## 2 EOF001      1     Study~ explicit     a       Catac~          3.51 present
## 3 EOF001      1     Study~ explicit     a       Contr~          1.75 absent
## 4 EOF001      1     Study~ explicit     a       Debac~          2.90 present
## 5 EOF001      1     Study~ explicit     a       Dorma~          1.63 absent
## 6 EOF001      1     Study~ explicit     a       Epigr~          1.38 present
## # ... with 22 more variables: Measure <fct>, Score <dbl>, WASImRS <dbl>,
## #   TOWREsweRS <dbl>, TOWREpdeRS <dbl>, CC2regRS <dbl>, CC2irregRS <dbl>,
## #   CC2nwRS <dbl>, WASIvRS <dbl>, BPVSRS <dbl>, Spelling.transcription <fct>,
## #   Levenshtein.Score <dbl>, zTOWREsweRS <dbl>, zTOWREpdeRS <dbl>,
## #   zCC2regRS <dbl>, zCC2irregRS <dbl>, zCC2nwRS <dbl>, zWASIvRS <dbl>,
## #   zBPVSRS <dbl>, mean_z_vocab <dbl>, mean_z_read <dbl>, zConsistency_H <dbl>
```

Data collected by Ricketts et al. (in press) in a longitudinal study of the impacts of learning conditions on child word learning: spelling test response Score (accuracy) outcome

Research questions

1. Does the presence of orthography promote greater word learning?
   - We predicted that children would demonstrate greater orthographic learning for words that they had seen (orthography present condition) versus not seen (orthography absent condition)
2. Will orthographic facilitation be greater when the presence of orthography is emphasized explicitly during teaching?
   - We expected to observe an interaction between instructions and orthography, with the highest levels of learning when the orthography present condition was combined with explicit instructions
3. Does word consistency moderate the orthographic facilitation effect?
   - For orthographic learning, we expected that the presence of orthography might be particularly beneficial for words with higher spelling-sound consistency, with learning highest when children saw and heard the word, and these codes provided overlapping information.

Category coding

We follow recommendations to use sum contrast coding for the experimental factors

- Orthography, absent (-1) vs. present (+1)
- Instructions, incidental (-1) vs. explicit (+1)
- Time, test time 1 (-1) vs. time 2 (+1)

Category coding practicalities

```
library(memisc)
# -- check
contrasts(long.orth$Orthography)
```

```
##         present
## absent        0
## present       1
```

```
# -- change
contrasts(long.orth$Orthography) <- contr.sum(2, base = 1)
# -- check
contrasts(long.orth$Orthography)
```

```
##          2
## absent  -1
## present  1
```

Specify the analysis: fixed effects

We are testing the effects of factors in a 2 x 2 factorial design embedded in a longitudinal study

- Time: time 1 versus time 2
- Orthography: present versus absent conditions
- Instructions: explicit versus incidental conditions
- Standardized spelling-sound consistency
- Interaction between the effects of Orthography and Instructions
- Interaction between the effects of Orthography and consistency

Specify a random intercepts model

```
long.orth.min.glmer <- glmer(Score ~

                             Time + Orthography + Instructions + zConsistency_H
                             Orthography:Instructions +

                             Orthography:zConsistency_H +

                                    (1 | Participant) +

                                    (1 |Word),

                                  family = "binomial",
                                  glmerControl(optimizer="bobyqa",
                                               optCtrl=list(maxfun=2e5)),

                                  data = long.orth)

summary(long.orth.min.glmer)
```

Specify a random intercepts model

- `glmer()` for a *generalized* linear mixed-effects model of accuracy
- `(1 | Participant)` random effects of participants on intercepts
- `(1 |Word)` random effects of stimulus on on intercepts
- `family = binomial` accuracy is a binary outcome variable (correct, incorrect) so we assume a binomial probability distribution
- We change the underlying mathematical engine (the optimizer) to cope with greater model complexity `glmerControl(optimizer="bobyqa", ...)`

```
long.orth.min.glmer <- glmer(Score ~

                    Time + Orthography + Instructions +
                    zConsistency_H +

                    Orthography:Instructions +

                    Orthography:zConsistency_H +

                    (1 | Participant) +

                    (1 |Word),

                    family = "binomial",
                    glmerControl(optimizer="bobyqa",
                                 optCtrl=list(maxfun=2e5)),

                    data = long.orth)

summary(long.orth.min.glmer)
```

Specify a random intercepts model

We specify the main effects with:

```
Time + Orthography + Instructions + zConsistency_H +
```

We specify the interaction effects with:

```
Orthography:Instructions +

Orthography:zConsistency_H +
```

Read the results

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Score ~ Time + Orthography + Instructions + zConsistency_H +
##     Orthography:Instructions + Orthography:zConsistency_H + (1 |
##     Participant) + (1 | Word)
##    Data: long.orth
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
##
##      AIC      BIC   logLik deviance df.resid
##   1040.4   1086.7   -511.2   1022.4     1254
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.0994 -0.4083 -0.2018  0.2019  7.4940
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Participant (Intercept) 1.840    1.357
##  Word        (Intercept) 2.224    1.491
## Number of obs: 1263, groups:  Participant, 41; Word, 16
##
## Fixed effects:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -1.88631    0.50839  -3.710 0.000207 ***
## Time2                              0.10027    0.16665   0.602 0.547380
## Orthography2                       0.46080    0.12313   3.742 0.000182 ***
## Instructionsincidental            -0.08458    0.46067  -0.184 0.854327
## zConsistency_H                    -0.61809    0.38400  -1.610 0.107485
## Orthography2:Instructionsincidental -0.01157    0.16637  -0.070 0.944551
## Orthography2:zConsistency_H        0.01461    0.08311   0.176 0.860444
```

Read the results

- We see variances for random effects of participants and stimuli on intercepts
- Fixed effects estimates
- Effect of presence of Orthography appears to be significant
- Positive coefficient suggests log odds of correct response *greater* when words learned in the presence of orthography (word form)

```
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  Participant (Intercept) 1.840    1.357
##  Word        (Intercept) 2.224    1.491
## Number of obs: 1263, groups:  Participant, 41; Word, 16
##
## Fixed effects:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -1.878465   0.443946  -4.231 2.32e-05 ***
## Time2                        0.050136   0.083325   0.602    0.547
## Orthography2                 0.455009   0.086813   5.241 1.59e-07 ***
## Instructions1                0.042290   0.230336   0.184    0.854
## zConsistency_H              -0.618092   0.384004  -1.610    0.107
## Orthography2:Instructions1   0.005786   0.083187   0.070    0.945
## Orthography2:zConsistency_H  0.014611   0.083105   0.176    0.860
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Visualize the effects estimates

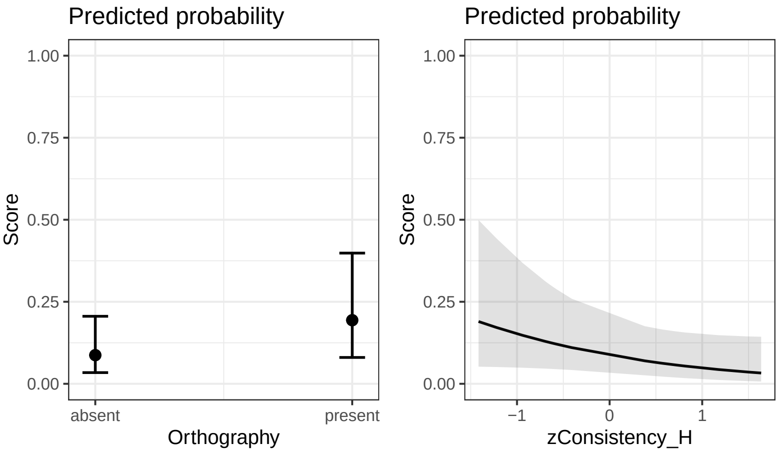


Figure 3: Effect of orthography condition (present versus absent) on probability of a response being correct; see also trend such that probability of a response being correct in the spelling test if the target word spelling-sound consistency is greater

What random effects should we include

- If you are testing effects manipulated according to a pre-specified design then you should:
  - Test random intercepts – due to random differences between subjects or between items (or other sample grouping variables)
  - Test random slopes for all within-subjects or within-items (or other) fixed effects

What random effects should we include

This means that specification of the random effects structure requires two sets of information:

1. What are the fixed effects?
2. What are the grouping variables: did you test multiple participants using multiple stimuli (e.g., words . . .) or did you test participants under multiple different conditions (e.g., levels of experimental condition factors)?

For our example, a maximal model would include

```
long.orth.max.glmer <- glmer(Score ~

                     Time + Orthography + Instructions + zConsistency_H +

                     Orthography:Instructions +

                     Orthography:zConsistency_H +

                     (Time + Orthography + zConsistency_H + 1 | Participant) +

                     (Time + Orthography + Instructions + 1 |Word),

                     family = "binomial",
                     glmerControl(optimizer="bobyqa",
                                  optCtrl=list(maxfun=2e5)),

                            data = long.orth)

summary(long.orth.max.glmer)
```

GLMMs with complex random effects can be trouble

```
## boundary (singular) fit: see ?isSingular

## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: Score ~ Time + Orthography + Instructions + zConsistency_H +
##     Orthography:Instructions + Orthography:zConsistency_H + (Time +
##     Orthography + zConsistency_H + 1 | Participant) + (Time +
##     Orthography + Instructions + 1 | Word)
##    Data: long.orth
## Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
##
##      AIC      BIC   logLik deviance df.resid
##   1053.6   1192.4   -499.8    999.6     1236
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1011 -0.4027 -0.1722  0.2036  7.0320
##
## Random effects:
##  Groups      Name                  Variance Std.Dev. Corr
##  Participant (Intercept)           1.91559  1.3840
##              Time2                 0.02270  0.1507    0.58
##              Orthography2          0.07997  0.2828    0.81 -0.01
##              zConsistency_H        0.06558  0.2561    0.46  0.99 -0.16
##  Word        (Intercept)           1.81892  1.3487
##              Time2                 0.18701  0.4324   -0.01
##              Orthography2          0.09376  0.3062   -0.58 -0.81
##              Instructionsincidental 0.85106  0.9225    0.57  0.05 -0.38
## Number of obs: 1263, groups:  Participant, 41; Word, 16
##
```

Bad signs

We can see that a model has difficulty if we see things like:

1. Convergence warnings, obviously
2. Very very small random effects variances
3. Extreme random effects correlations of ± 1.00

Examine the utility of random effects by comparing models with the same fixed effects but varying random effects: add one random effect at a time

```
long.orth.2.glmer <- glmer(Score ~

                        Time + Orthography + Instructions + zConsistency_H +

                        Orthography:Instructions +

                        Orthography:zConsistency_H +

                        (dummy(Orthography) + 1 || Participant) +

                        (dummy(Orthography) + 1 || Word),

                        family = "binomial",
                           glmerControl(optimizer="bobyqa",
                                        optCtrl=list(maxfun=2e5)),

                           data = long.orth)

summary(long.orth.2.glmer)
```

Add terms for the random effects of participant and word on the slopes of the within-subjects and within-items Orthography effect

Examine the utility of random effects by comparing models with the same fixed effects but varying random effects: add one random effect at a time

```
long.orth.3.glmer <- glmer(Score ~

                        Time + Orthography + Instructions + zConsistency_H +

                        Orthography:Instructions +

                        Orthography:zConsistency_H +

                        (dummy(Orthography) + 1 || Participant) +

                        (dummy(Orthography) + dummy(Instructions) + 1 || Word),

                        family = "binomial",
                           glmerControl(optimizer="bobyqa",
                                        optCtrl=list(maxfun=2e5)),

                           data = long.orth)

summary(long.orth.3.glmer)
```

Add terms for the random effects of participant and word on the slopes of the within-subjects and within-items Orthography and Instructions effects

Comparison of models varying in random effects

```r
anova(long.orth.min.glmer, long.orth.2.glmer)
```

```
## Data: long.orth
## Models:
## long.orth.min.glmer: Score ~ Time + Orthography + Instructions + zConsistency_H + Orthography:Instructions +
## long.orth.2.glmer: Score ~ Time + Orthography + Instructions + zConsistency_H + Orthography:Instructions + O
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## long.orth.min.glmer    9 1040.4 1086.7 -511.20   1022.4
## long.orth.2.glmer     11 1041.0 1097.6 -509.51   1019.0 3.3909  2     0.1835
```

Comparison of models varying in random effects

```
anova(long.orth.min.glmer, long.orth.3.glmer)
```

```
## Data: long.orth
## Models:
## long.orth.min.glmer: Score ~ Time + Orthography + Instructions + zConsistency_H + Orthography:Instructions +
## long.orth.3.glmer: Score ~ Time + Orthography + Instructions + zConsistency_H + Orthography:Instructions + O
##                     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## long.orth.min.glmer    9 1040.4 1086.7 -511.20   1022.4
## long.orth.3.glmer     12 1036.5 1098.2 -506.24   1012.5 9.9115  3    0.01933 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Evaluation

- The model comparison summary indicates that the addition of a random effect of words on the slope of the Instructions effect is justified by significantly improved model fit to data ($\chi^2 = 9.9115, 3df, p = 0.01933$).

Evaluation

- The model comparison summary indicates that the addition of a random effect of words on the slope of the Instructions effect is justified by significantly improved model fit to data ($\chi^2 = 9.9115, 3df, p = 0.01933$).
- While adding the random effect of Orthography does not improve model fit significantly, you see researchers allowing a generous p-value threshold for inclusion of terms (i.e. it is ok to add variables even where $.p < .2$)

Summary advice

My advice, then, is to consider whether random effects should be included in a model based on

1. Theoretical reasons, in terms of what your understanding of a study design allows and requires, with respect to random differences between groups (classes, participants, stimuli etc.) or stimuli
2. Model convergence, as when models do or do not converge
3. Over a series of model comparisons, an evaluation of whether model fit is improved by the inclusion of the random effect

Reporting model results

Explain approach We used mixed-effects models to analyse data because ...

Reporting model results

Explain how you get from the study design to the model you use We took a hypothesis driven approach, estimating the fixed effects of time (Time 1 versus Time 2), Orthography (absent versus present), Instructions (incidental versus explicit) and consistency (standardized H), as well as the interaction between orthography and instructions and the interaction between orthography and consistency

Reporting model results

Outline the model comparison or model selection work Likelihood ratio test comparison of models showed that a model with ... fit the data better than a model with just random intercepts ($\chi^2(df) = ..., p = ...$)

Help the reader with a concise summary of estimates A tabled summary of coefficient estimates, presenting fixed and random effects

Show and tell Use figures – model prediction plots, as seen – to help the reader to see what the fixed effects estimates imply

Use appendices or supplementary materials To give the reader full information on models fit, model comparisons

Reporting model results

Which model do we report?

- Note that given the model comparison results we have seen, I would probably report the estimates from `long.orth.3.glmer`

Reporting model results

Which model do we report?

- Note that given the model comparison results we have seen, I would probably report the estimates from `long.orth.3.glmer`
- The model appears to include the most comprehensive account of random effects while still being capable of converging

Summary

- We focused on the need to use Generalized Linear Mixed-effects Models (GLMMs)

Summary

- We focused on the need to use Generalized Linear Mixed-effects Models (GLMMs)
- We identified the kind of outcome data (like response accuracy) that requires analysis using GLMMs

Summary

- We focused on the need to use Generalized Linear Mixed-effects Models (GLMMs)
- We identified the kind of outcome data (like response accuracy) that requires analysis using GLMMs
- Alternative methods, and their limitations, were discussed

Summary

- We examined a study that incorporates repeated measures, a 2 x 2 factorial design, and a longitudinal aspect

Summary

- We examined a study that incorporates repeated measures, a 2 x 2 factorial design, and a longitudinal aspect
- We discussed the need to use effect coding for factors

Summary

- We examined a study that incorporates repeated measures, a 2 x 2 factorial design, and a longitudinal aspect
- We discussed the need to use effect coding for factors
- We worked through a random intercepts GLMM, and identified the critical elements - We then moved on to considering the question of what random effects we should include in the model

Summary

- We examined a study that incorporates repeated measures, a 2 x 2 factorial design, and a longitudinal aspect
- We discussed the need to use effect coding for factors
- We worked through a random intercepts GLMM, and identified the critical elements - We then moved on to considering the question of what random effects we should include in the model
- We considered how to report the analysis and results