Generalized Linear Mixed-effects models

Rob Davies

r.davies1@lancaster.ac.uk

Spring 2016

Aims for the class

1. Understand the reasons for using Generalized Linear Mixed-effects models (GLMMs) to analyze discrete outcome variables
2. Recognize the limitations of alternative methods for analyzing discrete outcome variables
3. Practise running GLMMs with varying fixed or random effects structures
4. Practise reporting the results of GLMMs

Discrete outcome variables are very common

Understand the reasons for using Generalized Linear Mixed-effects models (GLMMs) to analyze discrete outcome variables

- We often need to analyze outcomes that are discrete or categorical
  - The accuracy of responses (correct vs. incorrect)
  - The membership of one group out of two groups (e.g. impaired vs. unimpaired participant, fixation to left vs. right visual field)
  - Also, outcomes like: membership of one group out of multiple groups (categories); frequency of occurrence of an event; membership of ordered categories (e.g. Likert ratings scales)

Discrete outcome variables are very common

Understand the reasons for using Generalized Linear Mixed-effects models (GLMMs) to analyze discrete outcome variables

- We often need to analyze outcomes that are discrete or categorical
  - The accuracy of responses (correct vs. incorrect)
  - The membership of one group out of two groups (e.g. impaired vs. unimpaired participant, fixation to left vs. right visual field)
  - Also, outcomes like: membership of one group out of multiple groups (categories); frequency of occurrence of an event; membership of ordered categories (e.g. Likert ratings scales)

Discrete outcome variables are very common

Understand the reasons for using Generalized Linear Mixed-effects models (GLMMs) to analyze discrete outcome variables

- We often need to analyze outcomes that are discrete or categorical
  - The accuracy of responses (correct vs. incorrect)
  - The membership of one group out of two groups (e.g. impaired vs. unimpaired participant, fixation to left vs. right visual field)
  - Also, outcomes like: membership of one group out of multiple groups (categories); frequency of occurrence of an event; membership of ordered categories (e.g. Likert ratings scales)

Discrete outcome variables are very common

Understand the reasons for using Generalized Linear Mixed-effects models (GLMMs) to analyze discrete outcome variables

- We often need to analyze outcomes that are discrete or categorical
  - The accuracy of responses (correct vs. incorrect)
  - The membership of one group out of two groups (e.g. impaired vs. unimpaired participant, fixation to left vs. right visual field)
  - Also, outcomes like: membership of one group out of multiple groups (categories); frequency of occurrence of an event; membership of ordered categories (e.g. Likert ratings scales)

Recognize the limitations of alternative methods for analyzing response accuracy

You will see alternative methods used very frequently

- We often see response accuracy analyzed using one of the following methods:
  - The accuracy of responses (correct vs. incorrect) is counted e.g. as the number of correct responses (or errors) per subject, per condition
  - 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
  - See the same approach if the outcome is group membership (e.g. impaired vs. healthy) or visual field (e.g. left vs. right) etc.

Recognize the limitations of alternative methods for analyzing response accuracy

You will see alternative methods used very frequently

- We often see response accuracy analyzed using one of the following methods:
  - The accuracy of responses (correct vs. incorrect) is counted e.g. as the number of correct responses (or errors) per subject, per condition
  - 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
  - See the same approach if the outcome is group membership (e.g. impaired vs. healthy) or visual field (e.g. left vs. right) etc.

Recognize the limitations of alternative methods for analyzing response accuracy

You will see alternative methods used very frequently

- We often see response accuracy analyzed using one of the following methods:
  - The accuracy of responses (correct vs. incorrect) is counted e.g. as the number of correct responses (or errors) per subject, per condition
  - 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
  - See the same approach if the outcome is group membership (e.g. impaired vs. healthy) or visual field (e.g. left vs. right) etc.

Recognize the limitations of alternative methods for analyzing response accuracy

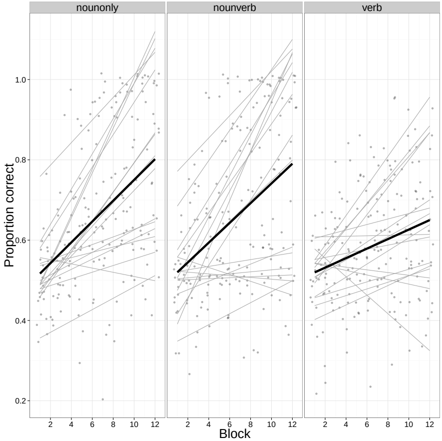
You will see alternative methods used very frequently

- We often see response accuracy analyzed using one of the following methods:
    - The accuracy of responses (correct vs. incorrect) is counted e.g. as the number of correct responses (or errors) per subject, per condition
    - 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
    - See the same approach if the outcome is group membership (e.g. impaired vs. healthy) or visual field (e.g. left vs. right) etc.

Recognize the limitations

ANOVA or regression over proportions can lead to spurious results – accuracy is bounded between 1 and 0, parametric model predictions or confidence intervals are not

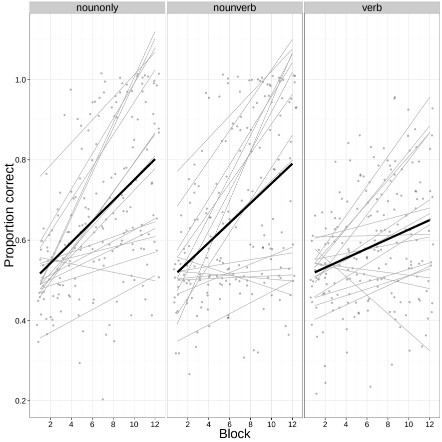
- Each plot point shows percentage accuracy of responses per subject per learning trials block and experimental learning condition
- Light grey lines show per-subject linear model best fit line for block effect
- Black lines show best fit for block effect for all participants



Recognize the limitations

ANOVA or regression over proportions can lead to spurious results – accuracy is bounded between 1 and 0, parametric model predictions or confidence intervals are not

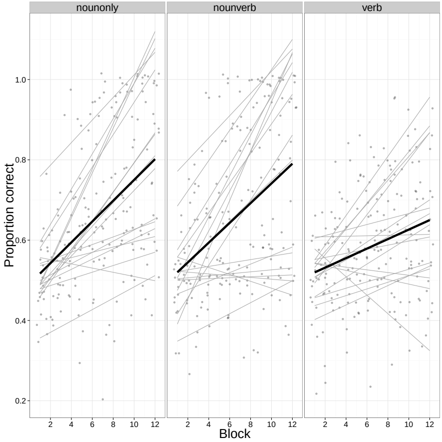
- Each plot point shows percentage accuracy of responses per subject per learning trials block and experimental learning condition
- Light grey lines show per-subject linear model best fit line for block effect
- Black lines show best fit for block effect for all participants



Recognize the limitations

ANOVA or regression over proportions can lead to spurious results – accuracy is bounded between 1 and 0, parametric model predictions or confidence intervals are not

- Each plot point shows percentage accuracy of responses per subject per learning trials block and experimental learning condition
- Light grey lines show per-subject linear model best fit line for block effect
- Black lines show best fit for block effect for all participants



Recognize the limitations

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, 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 $p \sim .5$

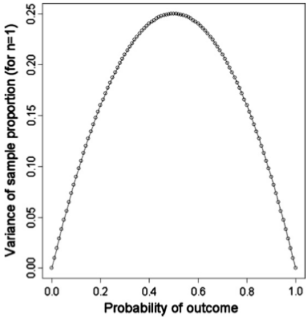


Figure : Jaeger (2008) variance of sample proportion dependent on probability that a response is correct (or probability that is 1 of 0,1 outcomes)

Recognize the limitations

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, 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 $p \sim .5$

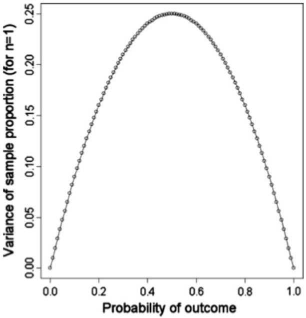


Figure : Jaeger (2008) variance of sample proportion dependent on probability that a response is correct (or probability that is 1 of 0,1 outcomes)

Recognize the limitations

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, 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 $p \sim .5$

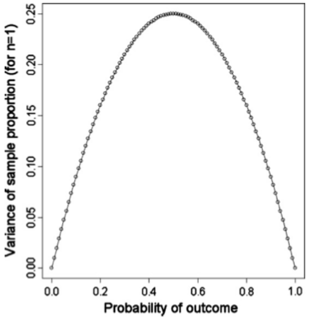


Figure : Jaeger (2008) variance of sample proportion dependent on probability that a response is correct (or probability that is 1 of 0,1 outcomes)

Recognize the limitations of traditional methods for analyzing response accuracy

The methods are common in the psychological literature but can give the wrong results

- 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, using ANOVA or linear models on accuracy (proportions of responses correct) can tell you, wrongly, that the interaction is significant
- Common remedies e.g. the arcsine root transformation do not work (Jaeger, 2008)

Recognize the limitations of traditional methods for analyzing response accuracy

The methods are common in the psychological literature but can give the wrong results

- 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, using ANOVA or linear models on accuracy (proportions of responses correct) can tell you, wrongly, that the interaction is significant
- Common remedies e.g. the arcsine root transformation do not work (Jaeger, 2008)

Recognize the limitations of traditional methods for analyzing response accuracy

The methods are common in the psychological literature but can give the wrong results

- 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, using ANOVA or linear models on accuracy (proportions of responses correct) can tell you, wrongly, that the interaction is significant
- Common remedies e.g. the arcsine root transformation do not work (Jaeger, 2008)

Recognize the limitations of traditional methods for analyzing response accuracy

The methods are common in the psychological literature but can give the wrong results

- 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, using ANOVA or linear models on accuracy (proportions of responses correct) can tell you, wrongly, that the interaction is significant
- Common remedies e.g. the arcsine root transformation do not work (Jaeger, 2008)

What we need, then, is a method that allows us to analyze categorical outcomes

We find the appropriate method in Generalized Linear Models – Generalized Linear Mixed-effects Models for repeated measures data

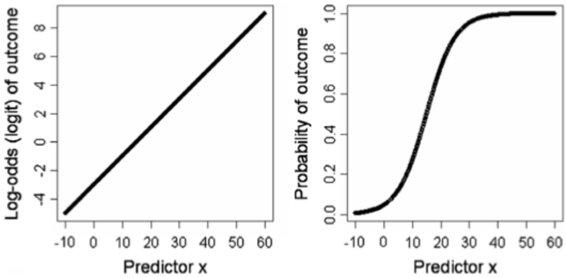


Figure : Jaeger (2008) the effect of some predictor x on a categorical outcome y: on the left the effect in logit space; on the right the effect in probability space

What we need, then, is a method that allows us to analyze categorical outcomes

The logistic transformation takes p the probability of an event with two possible outcomes, and turns it into a logit: the natural logarithm of the odds of the event

- The problem is how to estimate effects on a bounded outcome with a linear model
- Transforming a probability to odds $o = \frac{p}{1-p}$ is a partial solution
- Odds – the ratio of the probability of occurrence to non-occurrence or of correct vs. incorrect – are continuous and scaled from zero to infinity
- Using the logarithm of the odds $logit = ln\frac{p}{1-p}$ removes the boundary at zero because log odds ranges from negative to positive infinity

What we need, then, is a method that allows us to analyze categorical outcomes

The logistic transformation takes *p* the probability of an event with two possible outcomes, and turns it into a logit: the natural logarithm of the odds of the event

- The problem is how to estimate effects on a bounded outcome with a linear model
- Transforming a probability to odds $o = \frac{p}{1-p}$ is a partial solution
- Odds – the ratio of the probability of occurrence to non-occurrence or of correct vs. incorrect – are continuous and scaled from zero to infinity
- Using the logarithm of the odds $logit = ln\frac{p}{1-p}$ removes the boundary at zero because log odds ranges from negative to positive infinity

What we need, then, is a method that allows us to analyze categorical outcomes

The logistic transformation takes p the probability of an event with two possible outcomes, and turns it into a logit: the natural logarithm of the odds of the event

- The problem is how to estimate effects on a bounded outcome with a linear model
- Transforming a probability to odds $o = \frac{p}{1-p}$ is a partial solution
- Odds – the ratio of the probability of occurrence to non-occurrence or of correct vs. incorrect – are continuous and scaled from zero to infinity
- Using the logarithm of the odds $logit = ln\frac{p}{1-p}$ removes the boundary at zero because log odds ranges from negative to positive infinity

What we need, then, is a method that allows us to analyze categorical outcomes

The logistic transformation takes p the probability of an event with two possible outcomes, and turns it into a logit: the natural logarithm of the odds of the event

- The problem is how to estimate effects on a bounded outcome with a linear model
- Transforming a probability to odds $o = \frac{p}{1-p}$ is a partial solution
- Odds – the ratio of the probability of occurrence to non-occurrence or of correct vs. incorrect – are continuous and scaled from zero to infinity
- Using the logarithm of the odds $logit = ln\frac{p}{1-p}$ removes the boundary at zero because log odds ranges from negative to positive infinity

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)

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)

To illustrate GLMMs we use the word learning data set

Padraic Monaghan, our colleagues, and I examined the accuracy of responses in a word learning study: noun-verb-learning-study.csv

The word learning data set

- 48 adults participated in learning trials (12 blocks of 24)
- In each trial, observed 2 objects undergoing a different motion (one on the left, one on the right), and heard a sentence of fake words
- Words were either assigned to "refer to" the objects (nouns) or to the motions (verbs)
- Task is to indicate whether the heard sentence referred to the action on the left or the right of the screen to test if they could learn the object-or-motion to word associations

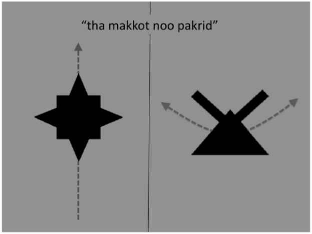


Figure : Example of a learning trial. Two moving objects are observed. Arrows indicate the movement path of the object. The four word phrase is simultaneously heard, with "tha" and "noo" function words and "makkot" and or "pakrid" referring to the motion and or object

The word learning data set

- 48 adults participated in learning trials (12 blocks of 24)
- In each trial, observed 2 objects undergoing a different motion (one on the left, one on the right), and heard a sentence of fake words
- Words were either assigned to "refer to" the objects (nouns) or to the motions (verbs)
- Task is to indicate whether the heard sentence referred to the action on the left or the right of the screen to test if they could learn the object-or-motion to word associations

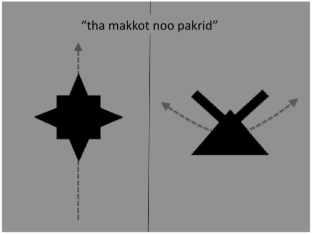


Figure : Example of a learning trial. Two moving objects are observed. Arrows indicate the movement path of the object. The four word phrase is simultaneously heard, with "tha" and "noo" function words and "makkot" and or "pakrid" referring to the motion and or object

The word learning data set

- 48 adults participated in learning trials (12 blocks of 24)
- In each trial, observed 2 objects undergoing a different motion (one on the left, one on the right), and heard a sentence of fake words
- Words were either assigned to "refer to" the objects (nouns) or to the motions (verbs)
- Task is to indicate whether the heard sentence referred to the action on the left or the right of the screen to test if they could learn the object-or-motion to word associations

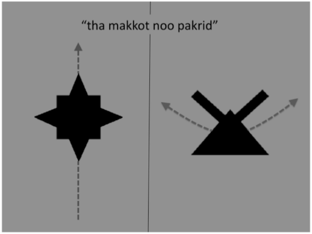


Figure : Example of a learning trial. Two moving objects are observed. Arrows indicate the movement path of the object. The four word phrase is simultaneously heard, with "tha" and "noo" function words and "makkot" and or "pakrid" referring to the motion and or object

The word learning data set

- 48 adults participated in learning trials (12 blocks of 24)
- In each trial, observed 2 objects undergoing a different motion (one on the left, one on the right), and heard a sentence of fake words
- Words were either assigned to “refer to” the objects (nouns) or to the motions (verbs)
- Task is to indicate whether the heard sentence referred to the action on the left or the right of the screen to test if they could learn the object-or-motion to word associations

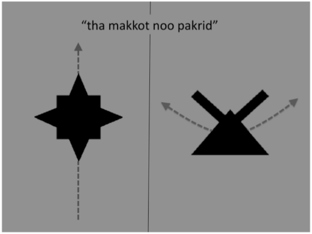


Figure : Example of a learning trial. Two moving objects are observed. Arrows indicate the movement path of the object. The four word phrase is simultaneously heard, with “tha” and “noo” function words and “makkot” and or “pakrid” referring to the motion and or object

Phenomena and data sets in the social sciences often have a multilevel structure

This is true for the word learning data set, which has a repeated measures design, requiring the use of mixed-effects models

Variation in learning between participants

We can reasonably attempt to model random effects of subjects on intercepts and the slope of the learning block and experimental condition effects

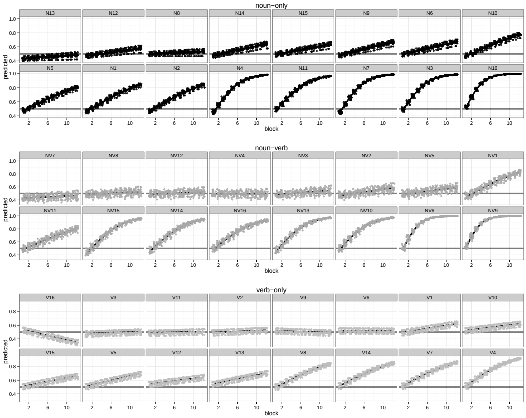


Figure : Graph showing the Generalized Linear Mixed-effects Model predicted values of the probability that a response is correct, for each participant, in each trial

A small change in R *lmer* code allows us to extend what we know about linear mixed-effects models to conduct *generalized linear mixed-effects models*

Models varying in fixed effects with constant random effects (of subjects or items on intercepts)

We start with an empty model

```
all.merged.glmm0 <- glmer(accuracy ~
    (1|Subjecta) + (1|targetobject) + (1|targetaction),
    data = all.merged, family = binomial)
summary(all.merged.glmm0)
```

- `glmer()` the function name changes because now we want a *generalized* linear mixed-effects model of accuracy
- `family = binomial` accuracy is a binary outcome variable so assume a binomial probability distribution

Models varying in fixed effects with constant random effects (of subjects or items on intercepts)

We start with an empty model

```
all.merged.glmm0 <- glmer(accuracy ~

    (1|Subjecta) + (1|targetobject) + (1|targetaction),

    data = all.merged, family = binomial)

summary(all.merged.glmm0)
```

- `glmer()` the function name changes because now we want a *generalized* linear mixed-effects model of accuracy
- `family = binomial` accuracy is a binary outcome variable so assume a binomial probability distribution

Add a predictor variable coding for the effect of experimental condition

```
all.merged.glmm1 <- glmer(accuracy ~

  Experiment +

  (1|Subjecta) + (1|targetobject) + (1|targetaction),

   data = all.merged, family = binomial)
summary(all.merged.glmm1)
```

- `Experiment` learning conditions coded with the “Experiment” variable, a factor with levels “nounonly”, “verbonly”, “nounverb”
- Notice have included random effects of stimulus object and motion sample units (object items, motion-action items) on intercepts

Add a predictor variable coding for the effect of experimental condition

```
all.merged.glmm1 <- glmer(accuracy ~

  Experiment +

  (1|Subjecta) + (1|targetobject) + (1|targetaction),

   data = all.merged, family = binomial)
summary(all.merged.glmm1)
```

- `Experiment` learning conditions coded with the “Experiment” variable, a factor with levels “nounonly”, “verbonly”, “nounverb”
- Notice have included random effects of stimulus object and motion sample units (object items, motion-action items) on intercepts

Add a predictor variable coding for learning trial block

```
all.merged.glmm2 <- glmer(accuracy ~

    Experiment + block +

    (1|Subjecta) + (1|targetobject) + (1|targetaction),

    data = all.merged, family = binomial)
summary(all.merged.glmm2)
```

- `block` there were 12 blocks of 24 learning trials, and here block is treated as a numeric variable

Add effect of experimental condition and block interaction

```
all.merged.glmm3 <- glmer(accuracy ~

    Experiment*block +

    (1|Subjecta) + (1|targetobject) + (1|targetaction),

    data = all.merged, family = binomial)
summary(all.merged.glmm3)
```

- Notice that the `(something)*(something)` get you interactions and main effects for all possible pairs of variables

How do we know if increasing *model complexity* by adding predictors actually helps us to account for variation in outcome values?

The Likelihood Ratio Test (LRT) comparison tells us main and interaction effects are justified by improved model fit to data

```
anova(all.merged.glmm0, all.merged.glmm1, all.merged.glmm2, all.merged.glmm3)
```

```
> anova(all.merged.glmm0, all.merged.glmm1, all.merged.glmm2, all.merged.glmm3)
Data: all.merged
Models:
all.merged.glmm0: accuracy ~ (1 | Subjecta) + (1 | targetobject) + (1 | targetaction)
all.merged.glmm1: accuracy ~ Experiment + (1 | Subjecta) + (1 | targetobject) +
all.merged.glmm1:     (1 | targetaction)
all.merged.glmm2: accuracy ~ Experiment + block + (1 | Subjecta) + (1 | targetobject) +
all.merged.glmm2:     (1 | targetaction)
all.merged.glmm3: accuracy ~ Experiment * block + (1 | Subjecta) + (1 | targetobject) +
all.merged.glmm3:     (1 | targetaction)
                 Df   AIC   BIC  logLik deviance    Chisq Chi Df Pr(>Chisq)
all.merged.glmm0  4 17259 17289 -8625.5    17251
all.merged.glmm1  6 17260 17306 -8624.2    17248   2.5788      2     0.2754
all.merged.glmm2  7 16928 16981 -8457.2    16914 333.9987      1  < 2.2e-16 ***
all.merged.glmm3  9 16888 16956 -8434.9    16870  44.6814      2  1.984e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure : LRT comparisons of models varying in fixed effects

When the goal of a confirmatory analysis is to test hypotheses about one or more critical fixed effects, what random-effects structure should one use?

- Current recommendations (Barr et al., 2013; JML): Maximal random effects structure
- If you are testing effects manipulated according to a prespecified – confirmatory study – design
  - Test random intercepts – subjects and items
  - Test random slopes for all within-subjects or within-items fixed effects
- But some authors (Bates et al., 2015; arXiv) argue we should test for the utility of adding model complexity

When the goal of a confirmatory analysis is to test hypotheses about one or more critical fixed effects, what random-effects structure should one use?

- Current recommendations (Barr et al., 2013; JML): Maximal random effects structure
- If you are testing effects manipulated according to a prespecified – confirmatory study – design
  - Test random intercepts – subjects and items
  - Test random slopes for all within-subjects or within-items fixed effects
- But some authors (Bates et al., 2015; arXiv) argue we should test for the utility of adding model complexity

When the goal of a confirmatory analysis is to test hypotheses about one or more critical fixed effects, what random-effects structure should one use?

- Current recommendations (Barr et al., 2013; JML): Maximal random effects structure
- If you are testing effects manipulated according to a prespecified – confirmatory study – design
  - Test random intercepts – subjects and items
  - Test random slopes for all within-subjects or within-items fixed effects
- But some authors (Bates et al., 2015; arXiv) argue we should test for the utility of adding model complexity

When the goal of a confirmatory analysis is to test hypotheses about one or more critical fixed effects, what random-effects structure should one use?

- Current recommendations (Barr et al., 2013; JML): Maximal random effects structure
- If you are testing effects manipulated according to a prespecified – confirmatory study – design
    - Test random intercepts – subjects and items
    - Test random slopes for all within-subjects or within-items fixed effects
- But some authors (Bates et al., 2015; arXiv) argue we should test for the utility of adding model complexity

When the goal of a confirmatory analysis is to test hypotheses about one or more critical fixed effects, what random-effects structure should one use?

- Current recommendations (Barr et al., 2013; JML): Maximal random effects structure
- If you are testing effects manipulated according to a prespecified – confirmatory study – design
  - Test random intercepts – subjects and items
  - Test random slopes for all within-subjects or within-items fixed effects
- But some authors (Bates et al., 2015; arXiv) argue we should test for the utility of adding model complexity

Examine the utility of random effects by comparing models with the same fixed effects but varying random effects

I am just going to assume we need both random effects of subjects and of items on intercepts so I focus on *random slopes*

```
all.merged.glmm4 <- glmer(accuracy ~

    Experiment*block +

    (block + 1|Subjecta) + (block + 1|targetobject)
    + (block + 1|targetaction),

    data = all.merged, family = binomial)
summary(all.merged.glmm4)
```

- `(block + 1|Subjecta) ...` learning block is manipulated within-subjects and within-items – so we must account for random variation between subjects, between stimulus objects, or between stimulus actions, in the block effect on response accuracy

The model summary indicates correlations between random intercepts and slopes for the items of 1

Bates et al. (2015) argue this shows that model complexity cannot really be sustained – we do not really need random effects of items on the slope of the block effect

```
> summary(all.merged.glmm4)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximatio
 Family: binomial  ( logit )
Formula: accuracy ~ Experiment * block + (block + 1 | Subjecta) + (block +
   Data: all.merged

     AIC      BIC   logLik deviance df.resid
 16500.3  16613.3  -8235.1  16470.3    13809

Scaled residuals:
     Min       1Q   Median       3Q      Max
-10.0695  -0.9979   0.4049   0.8602   1.5032

Random effects:
 Groups       Name        Variance  Std.Dev. Corr
 Subjecta     (Intercept) 4.414e-02 0.210087
              block       2.576e-02 0.160486 -0.68
 targetobject (Intercept) 4.830e-03 0.069501
              block       4.952e-05 0.007037 1.00
 targetaction (Intercept) 1.255e-02 0.112039
              block       4.329e-06 0.002081 1.00
Number of obs: 13824, groups:  Subjecta, 48; targetobject, 9; targetaction, 9

Fixed effects:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)              -0.26701    0.14383  -1.856   0.0634 .
Experimentnounverb       -0.01944    0.17076  -0.114   0.9094
Experimentverbonly        0.26595    0.18433   1.443   0.1491
block                     0.18715    0.04205   4.450 8.58e-06 ***
Experimentnounverb:block  0.01172    0.05940   0.197   0.8435
Experimentverbonly:block -0.12866    0.05925  -2.172   0.0299 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Extreme correlations (near 0 or 1) between random effects on intercepts and on slopes of fixed effects suggest the level of complexity in the model cannot really be justified

Note also that the variances for the random effects of items on the slopes of the block effect are very small

- We should see if a simpler model, excluding the correlations between item random effectscan be estimated

```
> summary(all.merged.glmm4)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximatio
 Family: binomial  ( logit )
Formula: accuracy ~ Experiment * block + (block + 1 | Subjecta) + (block +
   Data: all.merged

     AIC      BIC   logLik deviance df.resid
 16500.3  16613.3  -8235.1  16470.3    13809

Scaled residuals:
     Min       1Q   Median       3Q      Max
-10.0695  -0.9979   0.4049   0.8602   1.5032

Random effects:
 Groups       Name        Variance  Std.Dev. Corr
 Subjecta     (Intercept) 4.414e-02 0.210087
              block       2.576e-02 0.160486 -0.68
 targetobject (Intercept) 4.830e-03 0.069501
              block       4.952e-05 0.007037 1.00
 targetaction (Intercept) 1.255e-02 0.112039
              block       4.329e-06 0.002081 1.00
```

Try running the model removing just the problematic correlations – item-wise – between random effects

```
all.merged.glmm5 <- glmer(accuracy ~

    Experiment*block +

    (block + 1|Subjecta) + (1|targetobject) + (1|targetaction) +
    (block + 0|targetobject) + (block + 0|targetaction),

     data = all.merged, family = binomial)
summary(all.merged.glmm5)
```

- `(1|targetobject) + (1|targetaction) + (block + 0|targetobject)` the terms like `(1|targetobject)` specify a random effect (of items) on intercepts, while the terms `(block + 0|targetobject)` specify a random effect of items on the slope of the fixed effect (here, of block)

We want to simplify the model so we want to see no difference between the simpler and the more complex models

- Compare models with random intercepts and slopes, and either (1.) correlations between all random intercepts and slopes (glmm4) or (2.) random effects and just correlations between intercepts and slopes for subjects random effects not for items random effects (glmm5)
- If no difference between these models in relative fit then the item random effects correlations do not add anything to model utility

```
> anova(all.merged.glmm4, all.merged.glmm5)
Data: all.merged
Models:
all.merged.glmm5: accuracy ~ Experiment * block + (block + 1 | Subjecta) + (1 |
all.merged.glmm5:     targetobject) + (1 | targetaction) + (block + 0 | targetobject) +
all.merged.glmm5:     (block + 0 | targetaction)
all.merged.glmm4: accuracy ~ Experiment * block + (block + 1 | Subjecta) + (block +
all.merged.glmm4:     1 | targetobject) + (block + 1 | targetaction)
                 Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
all.merged.glmm5 13 16497 16595 -8235.4    16471
all.merged.glmm4 15 16500 16613 -8235.1    16470 0.4755      2     0.7884
```

We want to simplify the model so we want to see no difference between the simpler and the more complex models

- Compare models with random intercepts and slopes, and either (1.) correlations between all random intercepts and slopes (glmm4) or (2.) random effects and just correlations between intercepts and slopes for subjects random effects not for items random effects (glmm5)
- If no difference between these models in relative fit then the item random effects correlations do not add anything to model utility

```
> anova(all.merged.glmm4, all.merged.glmm5)
Data: all.merged
Models:
all.merged.glmm5: accuracy ~ Experiment * block + (block + 1 | Subjecta) + (1 |
all.merged.glmm5:     targetobject) + (1 | targetaction) + (block + 0 | targetobject) +
all.merged.glmm5:     (block + 0 | targetaction)
all.merged.glmm4: accuracy ~ Experiment * block + (block + 1 | Subjecta) + (block +
all.merged.glmm4:     1 | targetobject) + (block + 1 | targetaction)
                 Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
all.merged.glmm5 13 16497 16595 -8235.4    16471
all.merged.glmm4 15 16500 16613 -8235.1    16470 0.4755      2     0.7884
```

Reporting standards

Model comparisons – Using AIC, BIC and LRT

- Report briefly the model comparisons e.g. “Compared a simpler model: model 1, just main effects; model 2, main effects plus interactions”
- Report the AIC or BIC for the different models, or LRT for pair-wise comparisons
    - Report and explain the model selection choice, based on the aims of the study and the information criteria comparisons results

Reporting standards

Model comparisons – Using AIC, BIC and LRT

- Report briefly the model comparisons e.g. "Compared a simpler model: model 1, just main effects; model 2, main effects plus interactions"
- Report the AIC or BIC for the different models, or LRT for pair-wise comparisons
  - Report and explain the model selection choice, based on the aims of the study and the information criteria comparisons results

Reporting standards

Model comparisons – Using AIC, BIC and LRT

- Report briefly the model comparisons e.g. “Compared a simpler model: model 1, just main effects; model 2, main effects plus interactions”
- Report the AIC or BIC for the different models, or LRT for pair-wise comparisons
  - Report and explain the model selection choice, based on the aims of the study and the information criteria comparisons results

Reporting the model

- Summary of fixed effects – just like in linear models – with confidence intervals and p-values
- Report random effects variance and covariance (if applicable)

Fixed effects	Estimated coefficient	SE	Wald confidence intervals 2.50%	97.50%	z	Pr(>|z|)
(Intercept)	-0.2669	0.1544	-0.5482	0.0143	-1.7300	0.0840
Experimental condition (noun vs. noun-verb)	-0.0193	0.1797	-0.3539	0.3149	-0.1100	0.9150
Experimental condition (noun vs. verb)	0.2661	0.2028	-0.0948	0.6266	1.3100	0.1900
Block effect	0.1871	0.0421	0.1049	0.2694	4.4500	< 0.0001
Experimental condition (noun-verb):block interaction	0.0117	0.0594	-0.1047	0.1282	0.2000	0.8440
Experimental condition (verb):block interaction	-0.1287	0.0596	-0.2447	-0.0126	-2.1600	0.0310

Random effects	Name	Variance	Std.Dev.	Corr
Subject effect on intercepts	(Intercept)	0.0447	0.2115	
Random effect of subjects on slopes of block effects		0.0258	0.1606	-0.6800
Item effect (action) on intercepts	block	< 0.0001	0.0023	
Random effect of items (actions) on slopes of block effects	block	0.0001	0.0096	
Item effect (objects) on intercepts	(Intercept)	0.0153	0.1238	
Random effect of items (objects) on slopes of block effects	(Intercept)	0.0084	0.0919	

AIC	BIC	logLik	deviance
16496.736	16594.681	-8235.368	16470.736

13824 observations, 48 participants, 8 target action stimuli plus null action, 8 target object stimuli plus null object

Figure : Model summary table – word learning study – GLMM

Reporting the model

- Summary of fixed effects – just like in linear models – with confidence intervals and p-values
- Report random effects variance and covariance (if applicable)

Fixed effects	Estimated coefficient	SE	Wald confidence intervals 2.50%	97.50%	z	$\Pr(>\lvert z\rvert)$
(Intercept)	-0.2669	0.1544	-0.5482	0.0143	-1.7300	0.0840
Experimental condition (noun vs. noun-verb)	-0.0193	0.1797	-0.3539	0.3149	-0.1100	0.9150
Experimental condition (noun vs. verb)	0.2661	0.2028	-0.0948	0.6266	1.3100	0.1900
Block effect	0.1871	0.0421	0.1049	0.2694	4.4500	< 0.0001
Experimental condition (noun-verb):block interaction	0.0117	0.0594	-0.1047	0.1282	0.2000	0.8440
Experimental condition (verb):block interaction	-0.1287	0.0596	-0.2447	-0.0126	-2.1600	0.0310
Random effects						
	Name	Variance	Std.Dev.	Corr		
Subject effect on intercepts	(Intercept)	0.0447	0.2115			
Random effect of subjects on slopes of block effects		0.0258	0.1606	-0.6800		
Item effect (action) on intercepts	block	< 0.0001	0.0023			
Random effect of items (actions) on slopes of block effects	block	0.0001	0.0096			
Item effect (objects) on intercepts	(Intercept)	0.0153	0.1238			
Random effect of items (objects) on slopes of block effects	(Intercept)	0.0084	0.0919			
	AIC	BIC	logLik	deviance		
	16496.736	16594.681	-8235.368	16470.736		

13824 observations, 48 participants, 8 target action stimuli plus null action, 8 target object stimuli plus null object

Figure : Model summary table – word learning study – GLMM