Specifying models – fixed and random effects

Rob Davies

r.davies1@lancaster.ac.uk

Spring 2016

Aims for the class

1. Understand the use of *Information Criteria* and *Likelihood Ratio Tests* for evaluating models
2. Practise running linear mixed-effects models with varying fixed or random effects structures
3. Practise evaluating models

Linear mixed-effects models – Model comparison approach

We focus on building a series of models from the simplest to the most complex model supported by the data

- A minimal model – observed responses can be predicted only by the average response – the intercept
- Specific response values are then random deviations from this average
- Will our our capacity to predict responses be improved by adding other terms?
- What about the experimental factors we manipulate? Or variables carrying information about test conditions, particants or stimuli?

Linear mixed-effects models – Model comparison approach

We focus on building a series of models from the simplest to the most complex model supported by the data

- A minimal model – observed responses can be predicted only by the average response – the intercept
- Specific response values are then random deviations from this average
- Will our our capacity to predict responses be improved by adding other terms?
- What about the experimental factors we manipulate? Or variables carrying information about test conditions, particants or stimuli?

Linear mixed-effects models – Model comparison approach

We focus on building a series of models from the simplest to the most complex model supported by the data

- A minimal model – observed responses can be predicted only by the average response – the intercept
- Specific response values are then random deviations from this average
- Will our our capacity to predict responses be improved by adding other terms?
- What about the experimental factors we manipulate? Or variables carrying information about test conditions, particants or stimuli?

Linear mixed-effects models – Model comparison approach

We focus on building a series of models from the simplest to the most complex model supported by the data

- A minimal model – observed responses can be predicted only by the average response – the intercept
- Specific response values are then random deviations from this average
- Will our our capacity to predict responses be improved by adding other terms?
- What about the experimental factors we manipulate? Or variables carrying information about test conditions, particants or stimuli?

Simplicity and parsimony

We can expect to deal with a trade-off between too much and too little simplicity in model specification

- Models with too many parameters may tend to identify effects that are spurious
- Effects may be unintuitive and hard to explain *and* not reproduced in future samples
- Contrastingly, if we omit key variables we will likely mis-estimate the ‘true’ population value of the effects we seek to identify

Simplicity and parsimony

We can expect to deal with a trade-off between too much and too little simplicity in model specification

- Models with too many parameters may tend to identify effects that are spurious
- Effects may be unintuitive and hard to explain *and* not reproduced in future samples
- Contrastingly, if we omit key variables we will likely mis-estimate the 'true' population value of the effects we seek to identify

Simplicity and parsimony

We can expect to deal with a trade-off between too much and too little simplicity in model specification

- Models with too many parameters may tend to identify effects that are spurious
- Effects may be unintuitive and hard to explain *and* not reproduced in future samples
- Contrastingly, if we omit key variables we will likely mis-estimate the ‘true’ population value of the effects we seek to identify

We can use Information Criteria statistics like AIC or BIC to evaluate models

Understood within an approach: Information-theoretic methods– for linear models we can compare the F, R^2 and information criteria indices

- Information-theoretic methods are grounded in the insight that the distance between a model and observed outcomes corresponds to the 'information lost' when you use a model to approximate reality
- Information criteria – AIC or BIC – estimates of information loss
- The process of model selection aims to minimise information loss

We can use Information Criteria statistics like AIC or BIC to evaluate models

Understood within an approach: Information-theoretic methods– for linear models we can compare the F, R^2 and information criteria indices

- Information-theoretic methods are grounded in the insight that the distance between a model and observed outcomes corresponds to the 'information lost' when you use a model to approximate reality
- Information criteria – AIC or BIC – estimates of information loss
- The process of model selection aims to minimise information loss

We can use Information Criteria statistics like AIC or BIC to evaluate models

Understood within an approach: Information-theoretic methods– for linear models we can compare the F, R^2 and information criteria indices

- Information-theoretic methods are grounded in the insight that the distance between a model and observed outcomes corresponds to the 'information lost' when you use a model to approximate reality
- Information criteria – AIC or BIC – estimates of information loss
- The process of model selection aims to minimise information loss

Akaike Information Criteria: AIC

Akaike showed you could estimate information loss in terms of the likelihood of the model given the data

$$AIC = -2ln(l) + 2k \tag{1}$$

- $-2ln(l)$ -2 times the log of the likelihood of the model given the data
- (l) – likelihood
  - Is proportional to the probability of observed data conditional on some hypothesis being true

Akaike Information Criteria: AIC

Akaike showed you could estimate information loss in terms of the likelihood of the model given the data

$$AIC = -2ln(l) + 2k \tag{1}$$

- $-2ln(l)$ -2 times the log of the likelihood of the model given the data
- (l) – likelihood
    - Is proportional to the probability of observed data conditional on some hypothesis being true

Akaike Information Criteria: AIC

Akaike showed you could estimate information loss in terms of the likelihood of the model given the data

$$AIC = -2ln(l) + 2k \tag{1}$$

- $-2ln(l)$ -2 times the log of the likelihood of the model given the data
- (l) – likelihood
  - Is proportional to the probability of observed data conditional on some hypothesis being true

Akaike Information Criteria: AIC

Akaike showed you could estimate information loss in terms of the likelihood of the model given the data

$$AIC = -2ln(l) + 2k \tag{2}$$

- You want a more likely model – less information loss – closer to reality – you want more negative or lower AIC
- You can link models that are more likely – closer to reality – with models with smaller residuals
- Linear models with smaller residuals would have larger R^2 – more of observed outcome (response) variance is explained by better models

Akaike Information Criteria: AIC

Akaike showed you could estimate information loss in terms of the likelihood of the model given the data

$$AIC = -2ln(l) + 2k \tag{2}$$

- You want a more likely model – less information loss – closer to reality – you want more negative or lower AIC
- You can link models that are more likely – closer to reality – with models with smaller residuals
- Linear models with smaller residuals would have larger R^2 – more of observed outcome (response) variance is explained by better models

Akaike Information Criteria: AIC

Akaike showed you could estimate information loss in terms of the likelihood of the model given the data

$$AIC = -2ln(l) + 2k \tag{2}$$

- You want a more likely model – less information loss – closer to reality – you want more negative or lower AIC
- You can link models that are more likely – closer to reality – with models with smaller residuals
- Linear models with smaller residuals would have larger R^2 – more of observed outcome (response) variance is explained by better models

Bayesian Information Criteria: BIC

Schwartz proposed an alternative estimate

$$BIC = -2ln(l) + kln(N) \tag{3}$$

- $-2ln(l)$ – -2 times the log of the likelihood of the model given the data
- $+kln(N)$ – is the number of parameters in the model times the log of the sample size
- Crudely, the penalty for greater complexity is heavier in BIC
- Models with more parameters may fit the data better but some of those effects may be spurious

Bayesian Information Criteria: BIC

Schwartz proposed an alternative estimate

$$BIC = -2ln(l) + kln(N) \tag{3}$$

- $-2ln(l)$ – -2 times the log of the likelihood of the model given the data
- $+kln(N)$ – is the number of parameters in the model times the log of the sample size
- Crudely, the penalty for greater complexity is heavier in BIC
- Models with more parameters may fit the data better but some of those effects may be spurious

Bayesian Information Criteria: BIC

Schwartz proposed an alternative estimate

$$BIC = -2ln(l) + kln(N) \tag{3}$$

- $-2ln(l)$ – -2 times the log of the likelihood of the model given the data
- $+kln(N)$ – is the number of parameters in the model times the log of the sample size
- Crudely, the penalty for greater complexity is heavier in BIC
- Models with more parameters may fit the data better but some of those effects may be spurious

Bayesian Information Criteria: BIC

Schwartz proposed an alternative estimate

$$BIC = -2ln(l) + kln(N) \tag{3}$$

- $-2ln(l)$ – -2 times the log of the likelihood of the model given the data
- $+kln(N)$ – is the number of parameters in the model times the log of the sample size
- Crudely, the penalty for greater complexity is heavier in BIC
- Models with more parameters may fit the data better but some of those effects may be spurious

Model evaluation: using AIC and BIC

- Example – compare models varying in fixed effects: model 1, just main effects; model 2, main effects plus interactions
- Example – compare models varying in random effects: model 1, just random effect of subjects on intercepts; model 2, random effects of subjects and items on intercepts
- If the more complex model better approximates reality then it will be more likely given the data
  - BIC or AIC will be closer to negative infinity: $-2ln(l)$ will be larger
  - e.g. 10 is better than 1000, -1000 better than -10

Model evaluation: using AIC and BIC

- Example – compare models varying in fixed effects: model 1, just main effects; model 2, main effects plus interactions
- Example – compare models varying in random effects: model 1, just random effect of subjects on intercepts; model 2, random effects of subjects and items on intercepts
- If the more complex model better approximates reality then it will be more likely given the data
  - BIC or AIC will be closer to negative infinity: $-2ln(l)$ will be larger
  - e.g. 10 is better than 1000, -1000 better than -10

Model evaluation: using AIC and BIC

- Example – compare models varying in fixed effects: model 1, just main effects; model 2, main effects plus interactions
- Example – compare models varying in random effects: model 1, just random effect of subjects on intercepts; model 2, random effects of subjects and items on intercepts
- If the more complex model better approximates reality then it will be more likely given the data
  - BIC or AIC will be closer to negative infinity: $-2ln(l)$ will be larger
  - e.g. 10 is better than 1000, -1000 better than -10

Model evaluation: using AIC and BIC

- Example – compare models varying in fixed effects: model 1, just main effects; model 2, main effects plus interactions
- Example – compare models varying in random effects: model 1, just random effect of subjects on intercepts; model 2, random effects of subjects and items on intercepts
- If the more complex model better approximates reality then it will be more likely given the data
  - BIC or AIC will be closer to negative infinity: $-2ln(l)$ will be larger
  - e.g. 10 is better than 1000, -1000 better than -10

Model evaluation: using AIC and BIC

- Example – compare models varying in fixed effects: model 1, just main effects; model 2, main effects plus interactions
- Example – compare models varying in random effects: model 1, just random effect of subjects on intercepts; model 2, random effects of subjects and items on intercepts
- If the more complex model better approximates reality then it will be more likely given the data
  - BIC or AIC will be closer to negative infinity: $-2ln(l)$ will be larger
  - e.g. 10 is better than 1000, -1000 better than -10

Reporting standards; Using AIC and BIC

So far, I tend to see such analyses reported more in ecology, little in psychology

- Report briefly the model comparisons: "Compared a simpler model: model 1, just main effects; model 2, main effects plus interactions"
- Report the AIC or BIC for the different models
  - Report and explain the model selection *which model is more useful?*, based on aims and information criteria

Reporting standards; Using AIC and BIC

So far, I tend to see such analyses reported more in ecology, little in psychology

- Report briefly the model comparisons: "Compared a simpler model: model 1, just main effects; model 2, main effects plus interactions"
- Report the AIC or BIC for the different models
  - Report and explain the model selection *which model is more useful?*, based on aims and information criteria

Reporting standards; Using AIC and BIC

So far, I tend to see such analyses reported more in ecology, little in psychology

- Report briefly the model comparisons: "Compared a simpler model: model 1, just main effects; model 2, main effects plus interactions"
- Report the AIC or BIC for the different models
  - Report and explain the model selection *which model is more useful?*, based on aims and information criteria

Likelihood ratio test comparison

Likelihood ratio test comparisons appear increasingly wide-spread in psychology and are recommended for evaluating the relative utility of fixed or random effects

Baayen et al., 2008; Barr et al., 2013; Bates (LME developer list); glmm wiki dot; but see Bolker et al., 2009; Pinheiro & Bates, 2000

- The test statistic is the comparison of the likelihood of the simpler model with the more complex model
- Comparison by division $2log\frac{likelihood-complex}{likelihood-simple}$
- The likelihood ratio is compared to the χ^2 distribution for a significance test
- Assuming the null hypothesis that the simpler model is adequate
- With df equal to the difference in the number of parameters of the models being compared
- Alternatively, we can say: $deviance = deviance_{simple} - deviance_{complex}$ where $deviance = -2log(likelihood_{model})$

Likelihood ratio test comparison

Likelihood ratio test comparisons appear increasingly wide-spread in psychology and are recommended for evaluating the relative utility of fixed or random effects Baayen et al., 2008; Barr et al., 2013; Bates (LME developer list); glmm wiki dot; but see Bolker et al., 2009; Pinheiro & Bates, 2000

- The test statistic is the comparison of the likelihood of the simpler model with the more complex model
- Comparison by division $2log\frac{likelihood-complex}{likelihood-simple}$
- The likelihood ratio is compared to the χ^2 distribution for a significance test
- Assuming the null hypothesis that the simpler model is adequate
- With df equal to the difference in the number of parameters of the models being compared
- Alternatively, we can say: $deviance = deviance_{simple} - deviance_{complex}$ where $deviance = -2log(likelihood_{model})$

Likelihood ratio test comparison

Likelihood ratio test comparisons appear increasingly wide-spread in psychology and are recommended for evaluating the relative utility of fixed or random effects Baayen et al., 2008; Barr et al., 2013; Bates (LME developer list); glmm wiki dot; but see Bolker et al., 2009; Pinheiro & Bates, 2000

- The test statistic is the comparison of the likelihood of the simpler model with the more complex model
- Comparison by division $2log\frac{likelihood-complex}{likelihood-simple}$
- The likelihood ratio is compared to the χ^2 distribution for a significance test
- Assuming the null hypothesis that the simpler model is adequate
- With df equal to the difference in the number of parameters of the models being compared
- Alternatively, we can say: $deviance = deviance_{simple} - deviance_{complex}$ where $deviance = -2log(likelihood_{model})$

Likelihood ratio test comparison

Likelihood ratio test comparisons appear increasingly wide-spread in psychology and are recommended for evaluating the relative utility of fixed or random effects Baayen et al., 2008; Barr et al., 2013; Bates (LME developer list); glmm wiki dot; but see Bolker et al., 2009; Pinheiro & Bates, 2000

- The test statistic is the comparison of the likelihood of the simpler model with the more complex model
- Comparison by division $2log\frac{likelihood-complex}{likelihood-simple}$
- The likelihood ratio is compared to the χ^2 distribution for a significance test
- Assuming the null hypothesis that the simpler model is adequate
- With df equal to the difference in the number of parameters of the models being compared
- Alternatively, we can say: $deviance = deviance_{simple} - deviance_{complex}$ where $deviance = -2log(likelihood_{model})$

Likelihood ratio test comparison

Likelihood ratio test comparisons appear increasingly wide-spread in psychology and are recommended for evaluating the relative utility of fixed or random effects Baayen et al., 2008; Barr et al., 2013; Bates (LME developer list); glmm wiki dot; but see Bolker et al., 2009; Pinheiro & Bates, 2000

- The test statistic is the comparison of the likelihood of the simpler model with the more complex model
- Comparison by division $2log\frac{likelihood-complex}{likelihood-simple}$
- The likelihood ratio is compared to the χ^2 distribution for a significance test
- Assuming the null hypothesis that the simpler model is adequate
- With df equal to the difference in the number of parameters of the models being compared
- Alternatively, we can say: $deviance = deviance_{simple} - deviance_{complex}$ where $deviance = -2log(likelihood_{model})$

Likelihood ratio test comparison

Likelihood ratio test comparisons appear increasingly wide-spread in psychology and are recommended for evaluating the relative utility of fixed or random effects Baayen et al., 2008; Barr et al., 2013; Bates (LME developer list); glmm wiki dot; but see Bolker et al., 2009; Pinheiro & Bates, 2000

- The test statistic is the comparison of the likelihood of the simpler model with the more complex model
- Comparison by division $2log\frac{likelihood-complex}{likelihood-simple}$
- The likelihood ratio is compared to the χ^2 distribution for a significance test
- Assuming the null hypothesis that the simpler model is adequate
- With df equal to the difference in the number of parameters of the models being compared
- Alternatively, we can say: $deviance = deviance_{simple} - deviance_{complex}$ where $deviance = -2log(likelihood_{model})$

Model comparisons among mixed-effects models – use *anova()* function

```
anova(full.lmer0, full.lmer1)
```

- `anova(...,....)` compare pairs of models named in brackets
- `full.lmer0` including a simpler model – fewer parameters (predictors)
- `full.lmer1` compared to a more complex model – more predictors – includes simpler model as a special case

Model comparisons among mixed-effects models – use *anova()* function

```
anova(full.lmer0, full.lmer1)
```

- `anova(...,....)` compare pairs of models named in brackets
- `full.lmer0` including a simpler model – fewer parameters (predictors)
- `full.lmer1` compared to a more complex model – more predictors – includes simpler model as a special case

Model comparisons among mixed-effects models – use *anova()* function

```
anova(full.lmer0, full.lmer1)
```

- `anova(...,....)` compare pairs of models named in brackets
- `full.lmer0` including a simpler model – fewer parameters (predictors)
- `full.lmer1` compared to a more complex model – more predictors – includes simpler model as a special case

Running the *anova*(,) comparison will deliver AIC, BIC, and likelihood comparisons for varying models

```
> anova(full.lmer1, full.lmer2)
Data: subjects.behaviour.items.nomissing
Models:
full.lmer1: logrt ~ cAge + cTOWRE_wordacc + cTOWRE_nonwordacc + cART_HRminusFR +
full.lmer1:     (1 | subjectID) + (1 | item_name)
full.lmer2: logrt ~ cAge + cTOWRE_wordacc + cTOWRE_nonwordacc + cART_HRminusFR +
full.lmer2:     item_type + cLength + cOrtho_N + cBG_Mean + (1 | subjectID) +
full.lmer2:     (1 | item_name)
           Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
full.lmer1  8 -17983 -17925 8999.4   -17999
full.lmer2 12 -18319 -18232 9171.3   -18343 343.81      4  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Figure : Comparison of model with subject attribute predictors and model also with item effects

Model comparison

- AIC, BIC and LRT comparisons should be consistent in their indications – which model to prefer
- Can be tricky where dealing with complex sets of predictors – indicators may diverge
- Remember that BIC may penalise complexity more heavily – especially if conducting exploratory research
- Remember that may be obliged to include all effects built-in by design – if conducting confirmatory study

```
> anova(full.lmer0, full.lmer1)
Data: subjects.behaviour.items.nomissing
Models:
full.lmer0: logrt ~ (1 | subjectID) + (1 | item_name)
full.lmer1: logrt ~ cAge + cTOWRE_wordacc + cTOWRE_nonwordacc + cART_HRminusFR +
full.lmer1:     (1 | subjectID) + (1 | item_name)
           Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
full.lmer0  4 -17981 -17952 8994.4   -17989
full.lmer1  8 -17983 -17925 8999.4   -17999 10.116      4    0.03852 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model comparison

- AIC, BIC and LRT comparisons should be consistent in their indications – which model to prefer
- Can be tricky where dealing with complex sets of predictors – indicators may diverge
- Remember that BIC may penalise complexity more heavily – especially if conducting exploratory research
- Remember that may be obliged to include all effects built-in by design – if conducting confirmatory study

```
> anova(full.lmer0, full.lmer1)
Data: subjects.behaviour.items.nomissing
Models:
full.lmer0: logrt ~ (1 | subjectID) + (1 | item_name)
full.lmer1: logrt ~ cAge + cTOWRE_wordacc + cTOWRE_nonwordacc + cART_HRminusFR +
full.lmer1:     (1 | subjectID) + (1 | item_name)
           Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
full.lmer0  4 -17981 -17952 8994.4   -17989
full.lmer1  8 -17983 -17925 8999.4   -17999 10.116      4    0.03852 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model comparison

- AIC, BIC and LRT comparisons should be consistent in their indications – which model to prefer
- Can be tricky where dealing with complex sets of predictors – indicators may diverge
- Remember that BIC may penalise complexity more heavily – especially if conducting exploratory research
- Remember that may be obliged to include all effects built-in by design – if conducting confirmatory study

```
> anova(full.lmer0, full.lmer1)
Data: subjects.behaviour.items.nomissing
Models:
full.lmer0: logrt ~ (1 | subjectID) + (1 | item_name)
full.lmer1: logrt ~ cAge + cTOWRE_wordacc + cTOWRE_nonwordacc + cART_HRminusFR +
full.lmer1:     (1 | subjectID) + (1 | item_name)
           Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
full.lmer0  4 -17981 -17952 8994.4   -17989
full.lmer1  8 -17983 -17925 8999.4   -17999 10.116      4    0.03852 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model comparison

- AIC, BIC and LRT comparisons should be consistent in their indications – which model to prefer
- Can be tricky where dealing with complex sets of predictors – indicators may diverge
- Remember that BIC may penalise complexity more heavily – especially if conducting exploratory research
- Remember that may be obliged to include all effects built-in by design – if conducting confirmatory study

```
> anova(full.lmer0, full.lmer1)
Data: subjects.behaviour.items.nomissing
Models:
full.lmer0: logrt ~ (1 | subjectID) + (1 | item_name)
full.lmer1: logrt ~ cAge + cTOWRE_wordacc + cTOWRE_nonwordacc + cART_HRminusFR +
full.lmer1:     (1 | subjectID) + (1 | item_name)
           Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
full.lmer0  4 -17981 -17952 8994.4   -17989
full.lmer1  8 -17983 -17925 8999.4   -17999 10.116      4    0.03852 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A more realistic repeated measures model of word frequency and reading ability effects

Taking into account random effect of subjects and items

$$RT = \beta_0 + \beta_{ability} + \beta_{frequency} + \beta_{ability*frequency} + \beta_{subject} + \beta_{item} + \beta_{subject*frequency} + \beta_{item*ability} + \epsilon \quad (4)$$

- What about effect of random variation between participants?
  - Allow intercept to vary – random effect of subject – some have slower average some have faster average than average overall
  - Allow effect of frequency to vary – random effect of subject – subjects can be affected by word frequency in different ways

A more realistic repeated measures model of word frequency and reading ability effects

Taking into account random effect of subjects and items

$$RT = \beta_0 + \beta_{ability} + \beta_{frequency} + \beta_{ability*frequency} + \beta_{subject} + \beta_{item} + \beta_{subject*frequency} + \beta_{item*ability} + \epsilon \quad (4)$$

- What about effect of random variation between participants?
  - Allow intercept to vary – random effect of subject – some have slower average some have faster average than average overall
  - Allow effect of frequency to vary – random effect of subject – subjects can be affected by word frequency in different ways

A more realistic repeated measures model of word frequency and reading ability effects

Taking into account random effect of subjects and items

$$RT = \beta_0 + \beta_{ability} + \beta_{frequency} + \beta_{ability*frequency} + \beta_{subject} + \beta_{item} + \beta_{subject*frequency} + \beta_{item*ability} + \epsilon \quad (4)$$

- What about effect of random variation between participants?
  - Allow intercept to vary – random effect of subject – some have slower average some have faster average than average overall
  - Allow effect of frequency to vary – random effect of subject – subjects can be affected by word frequency in different ways

A more realistic repeated measures model of the item type and reading ability effects

Taking into account random effects of subjects and items

$$RT = \beta_0 + \beta_{ability} + \beta_{frequency} + \beta_{ability*frequency} + \beta_{subject} + \beta_{item} + \beta_{subject*frequency} + \beta_{item*ability} + \epsilon \quad (5)$$

- What about effect of random variation between stimuli?
  - Allow intercept to vary – random effect of items – some items harder and elicit slower average response than others
  - Allow effect of reading ability to vary – random effect of items?

A more realistic repeated measures model of the item type and reading ability effects

Taking into account random effects of subjects and items

$$RT = \beta_0 + \beta_{ability} + \beta_{frequency} + \beta_{ability*frequency} + \beta_{subject} + \beta_{item} + \beta_{subject*frequency} + \beta_{item*ability} + \epsilon \quad (5)$$

- What about effect of random variation between stimuli?
  - Allow intercept to vary – random effect of items – some items harder and elicit slower average response than others
  - Allow effect of reading ability to vary – random effect of items?

A more realistic repeated measures model of the item type and reading ability effects

Taking into account random effects of subjects and items

$$RT = \beta_0 + \beta_{ability} + \beta_{frequency} + \beta_{ability*frequency} + \beta_{subject} + \beta_{item} + \beta_{subject*frequency} + \beta_{item*ability} + \epsilon \quad (5)$$

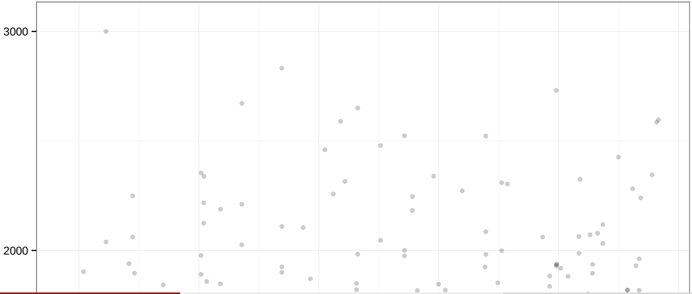
- What about effect of random variation between stimuli?
  - Allow intercept to vary – random effect of items – some items harder and elicit slower average response than others
  - Allow effect of reading ability to vary – random effect of items?

Random effect of subjects on intercepts

Differences in individuals' average response speed compared to the group average

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + \epsilon_{ij} \tag{6}$$

- β_0 common intercept, average outcome given the other effects
- $U_{0,j}$ adjustments to the intercept required to explain differences between common average and average for each j individual

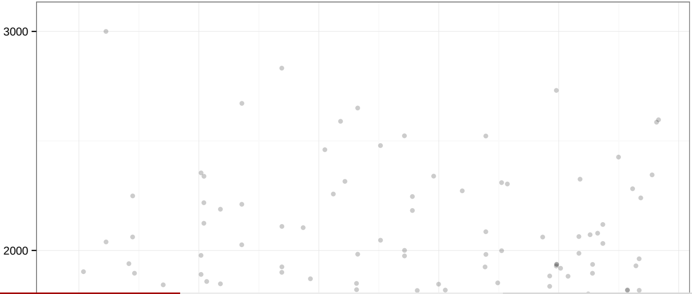


Random effect of subjects on intercepts

Differences in individuals' average response speed compared to the group average

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + \epsilon_{ij} \tag{6}$$

- β_0 common intercept, average outcome given the other effects
- $U_{0,j}$ adjustments to the intercept required to explain differences between common average and average for each j individual

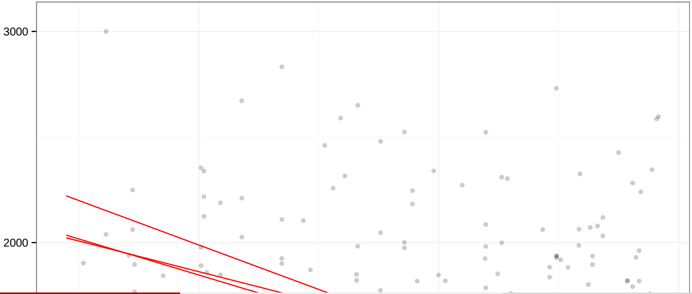


Random effect of subjects on slopes of fixed effects

Accounting for variation in *within-subjects* frequency effect on RTs

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + \epsilon_{ij} \quad (7)$$

- $\beta_1 X_i$ the group average frequency effect
- $U_{1,j}$ adjustments required to model differences between group average frequency effect and frequency effect for each j individual

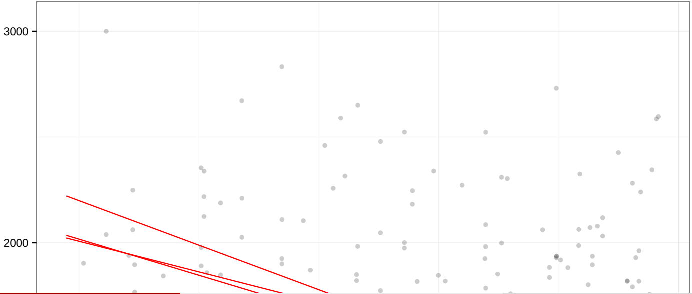


Random effect of subjects on slopes of fixed effects

Accounting for variation in *within-subjects* frequency effect on RTs

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + \epsilon_{ij} \tag{7}$$

- $\beta_1 X_i$ the group average frequency effect
- $U_{1,j}$ adjustments required to model differences between group average frequency effect and frequency effect for each j individual



Can allow random effects of both subjects and items

Solving the 'language-as-fixed-effect' problem in one model

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{8}$$

- $W_{0,i}$ random effect of items on intercepts – where W terms correspond to random effects of items
- $W_{1,i}$ may require a random effect of items on slopes of within-items effects e.g. effect of ability differences between participants

Can allow random effects of both subjects and items

Solving the 'language-as-fixed-effect' problem in one model

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{8}$$

- $W_{0,i}$ random effect of items on intercepts – where W terms correspond to random effects of items
- $W_{1,i}$ may require a random effect of items on slopes of within-items effects e.g. effect of ability differences between participants

Mixed effects models – both fixed effects and random effects

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \quad (9)$$

- $\beta_0 + \beta_1 X_i$ fixed effect of predictors – fixed because replicable by manipulation or selection
- $U_{0,j} + U_{1,j} + +W_{0,i} + W_{1,i} + \epsilon_{ij}$ random effects of group on intercepts and slopes – random because differences due to sampling

Mixed effects models – both fixed effects and random effects

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{9}$$

- $\beta_0 + \beta_1 X_i$ fixed effect of predictors – fixed because replicable by manipulation or selection
- $U_{0,j} + U_{1,j} + +W_{0,i} + W_{1,i} + \epsilon_{ij}$ random effects of group on intercepts and slopes – random because differences due to sampling

We could model random differences between participants items as fixed effects but we do not

We incorporate random effects terms in models to capture the spread, *the variance*, associated with random differences in intercepts or slopes

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{10}$$

- $\sigma^2_{U_{0,j}} + \sigma^2_{U_{1,j}} + \sigma^2_{U_{0,j}U_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{W_{0,j}} + \sigma^2_{W_{1,j}} + \sigma^2_{W_{0,j}W_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{\epsilon_{ij}}$ residuals – error variance left over
- Covariances are included because random intercepts and slopes may correlate e.g. if slow subjects are more vulnerable to difference in conditions

We could model random differences between participants items as fixed effects but we do not

We incorporate random effects terms in models to capture the spread, *the variance*, associated with random differences in intercepts or slopes

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{10}$$

- $\sigma^2_{U_{0,j}} + \sigma^2_{U_{1,j}} + \sigma^2_{U_{0,j} U_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{W_{0,j}} + \sigma^2_{W_{1,j}} + \sigma^2_{W_{0,j} W_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{\epsilon_{ij}}$ residuals – error variance left over
- Covariances are included because random intercepts and slopes may correlate e.g. if slow subjects are more vulnerable to difference in conditions

We could model random differences between participants items as fixed effects but we do not

We incorporate random effects terms in models to capture the spread, *the variance*, associated with random differences in intercepts or slopes

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{10}$$

- $\sigma^2_{U_{0,j}} + \sigma^2_{U_{1,j}} + \sigma^2_{U_{0,j}U_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{W_{0,j}} + \sigma^2_{W_{1,j}} + \sigma^2_{W_{0,j}W_{1,j}}$variances and covariance of random effects
- $\sigma^2_{\epsilon_{ij}}$ residuals – error variance left over
- Covariances are included because random intercepts and slopes may correlate e.g. if slow subjects are more vulnerable to difference in conditions

We could model random differences between participants items as fixed effects but we do not

We incorporate random effects terms in models to capture the spread, *the variance*, associated with random differences in intercepts or slopes

$$Y_{ij} = \beta_0 + \beta_1 X_i + U_{0,j} + U_{1,j} + W_{0,i} + W_{1,i} + \epsilon_{ij} \tag{10}$$

- $\sigma^2_{U_{0,j}} + \sigma^2_{U_{1,j}} + \sigma^2_{U_{0,j}U_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{W_{0,j}} + \sigma^2_{W_{1,j}} + \sigma^2_{W_{0,j}W_{1,j}}$ variances and covariance of random effects
- $\sigma^2_{\epsilon_{ij}}$ residuals – error variance left over
- Covariances are included because random intercepts and slopes may correlate e.g. if slow subjects are more vulnerable to difference in conditions

Estimation methods

An intuitive account of estimation in mixed-effects models

- If we knew the random effects, we could find the fixed effects estimates by minimising differences between predicted and observed outcomes – like linear modelling
- If we knew the fixed effects – the regression coefficients – we could work out the residuals and the random effects

Estimation methods

An intuitive account of estimation in mixed-effects models

- If we knew the random effects, we could find the fixed effects estimates by minimising differences between predicted and observed outcomes – like linear modelling
- If we knew the fixed effects – the regression coefficients – we could work out the residuals and the random effects

Estimation methods

An intuitive account of estimation in mixed-effects models

- At the start, we know neither, but we can move between partial estimation of fixed and random effect in an iterative approach
  - Using provisional values for the fixed effects to estimate the random effects
  - Using provisional values for the random effects to estimate the fixed effects again
  - To converge on the maximum likelihood estimates of effects – when the estimates stop changing

Estimation methods

An intuitive account of estimation in mixed-effects models

- At the start, we know neither, but we can move between partial estimation of fixed and random effect in an iterative approach
  - Using provisional values for the fixed effects to estimate the random effects
  - Using provisional values for the random effects to estimate the fixed effects again
  - To converge on the maximum likelihood estimates of effects – when the estimates stop changing

Estimation methods

An intuitive account of estimation in mixed-effects models

- At the start, we know neither, but we can move between partial estimation of fixed and random effect in an iterative approach
  - Using provisional values for the fixed effects to estimate the random effects
  - Using provisional values for the random effects to estimate the fixed effects again
  - To converge on the maximum likelihood estimates of effects – when the estimates stop changing

Estimation methods

An intuitive account of estimation in mixed-effects models

- At the start, we know neither, but we can move between partial estimation of fixed and random effect in an iterative approach
  - Using provisional values for the fixed effects to estimate the random effects
  - Using provisional values for the random effects to estimate the fixed effects again
  - To converge on the maximum likelihood estimates of effects – when the estimates stop changing

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- Restricted maximum likelihood ($REML = TRUE$)
  - REML estimates the variance components while taking into account the loss of degrees of freedom resulting from the estimation of the fixed effects
  - REML estimates vary if the fixed effects vary

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- Restricted maximum likelihood ($REML = TRUE$)
  - REML estimates the variance components while taking into account the loss of degrees of freedom resulting from the estimation of the fixed effects
  - REML estimates vary if the fixed effects vary

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- Restricted maximum likelihood ($REML = TRUE$)
  - REML estimates the variance components while taking into account the loss of degrees of freedom resulting from the estimation of the fixed effects
  - REML estimates vary if the fixed effects vary

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- REML estimates vary if the fixed effects vary
- Therefore not recommended to compare the likelihood of models varying in fixed effects fitted using REML (Pinheiro & Bates, 2000)
- REML method recommended for comparing the likelihood of models with *the same fixed effects* but *different random effects*
- ML method `REML= FALSE` recommended for comparing models with *different fixed effects* but *the same random effects*

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- REML estimates vary if the fixed effects vary
- Therefore not recommended to compare the likelihood of models varying in fixed effects fitted using REML (Pinheiro & Bates, 2000)
- REML method recommended for comparing the likelihood of models with *the same fixed effects* but *different random effects*
- ML method `REML= FALSE` recommended for comparing models with *different fixed effects* but *the same random effects*

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- REML estimates vary if the fixed effects vary
- Therefore not recommended to compare the likelihood of models varying in fixed effects fitted using REML (Pinheiro & Bates, 2000)
- REML method recommended for comparing the likelihood of models with *the same fixed effects* but *different random effects*
- ML method `REML= FALSE` recommended for comparing models with *different fixed effects* but *the same random effects*

Restricted maximum likelihood ($REML = TRUE$) and maximum likelihood ($REML = FALSE$) methods

- REML estimates vary if the fixed effects vary
- Therefore not recommended to compare the likelihood of models varying in fixed effects fitted using REML (Pinheiro & Bates, 2000)
- REML method recommended for comparing the likelihood of models with *the same fixed effects* but *different random effects*
- ML method `REML= FALSE` recommended for comparing models with *different fixed effects* but *the same random effects*

Model comparisons among mixed-effects models fitted using maximum likelihood ($REML = FALSE$) method

```
full.lmer.0 <- lmer(logrt ~

(1|subjectID) + (1|item_name),

data = ML.all.correct, REML = F)
```

- Maximum likelihood ($REML = FALSE$)
  - ML estimation methods can be used to fit models with varying fixed effects but the same random effects

Model comparisons among mixed-effects models fitted using maximum likelihood ($REML = FALSE$) method

```
full.lmer.0 <- lmer(logrt ~

(1|subjectID) + (1|item_name),

data = ML.all.correct, REML = F)
```

- Maximum likelihood ($REML = FALSE$)
  - ML estimation methods can be used to fit models with varying fixed effects but the same random effects

Examine the fixed effects then the random effects

Start by examining models varying in the *fixed effects* but constant in the *random effects*

- Compare maximum likelihood ($REML = FALSE$) models varying in fixed effects
  - Think about simpler models as simplifications or subsets of more complex models
- Add effects of interest in blocks or sets of predictors

Examine the fixed effects then the random effects

Start by examining models varying in the *fixed effects* but constant in the *random effects*

- Compare maximum likelihood ($REML = FALSE$) models varying in fixed effects
  - Think about simpler models as simplifications or subsets of more complex models
- Add effects of interest in blocks or sets of predictors

Examine the fixed effects then the random effects

Start by examining models varying in the *fixed effects* but constant in the *random effects*

- Compare maximum likelihood ($REML = FALSE$) models varying in fixed effects
    - Think about simpler models as simplifications or subsets of more complex models
- Add effects of interest in blocks or sets of predictors

Compare maximum likelihood models with varying fixed effects but the same random effects

```
full.lmer.1 <- lmer(logrt ~

  zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

  (1|subjectID) + (1|item_name),

  data = ML.all.correct, REML = FALSE)

summary(full.lmer.1)
```

- `zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +` add fixed effects
- `(1|subjectID) + (1|item_name)` to random effects of subjects and items on intercepts

Compare maximum likelihood models with varying fixed effects but the same random effects

```
full.lmer.1 <- lmer(logrt ~

  zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

  (1|subjectID) + (1|item_name),

  data = ML.all.correct, REML = FALSE)

summary(full.lmer.1)
```

- `zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +` add fixed effects
- `(1|subjectID) + (1|item_name)` to random effects of subjects and items on intercepts

Compare the model 1 with just subject effects to the model 2 also with item effects

```
full.lmer.2 <- lmer(logrt ~

zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

item_type + zLength + zOrtho_N +

(1|subjectID) + (1|item_name),

data = ML.all.correct, REML = F)

summary(full.lmer.2)
```

- `item_type + zLength + zOrtho_N` add item effects
- Everything else stays the same

Compare the model 1 with just subject effects to the model 2 also with item effects

```
full.lmer.2 <- lmer(logrt ~

zAge + zTOWRE_wordacc + zTOWRE_nonwordacc +

item_type + zLength + zOrtho_N +

(1|subjectID) + (1|item_name),

data = ML.all.correct, REML = F)

summary(full.lmer.2)
```

- `item_type + zLength + zOrtho_N` add item effects
- Everything else stays the same

Compare model 2 with subject and item effects to model 3 specifying interactions between subject and item effects

```
full.lmer.3  <- lmer(logrt ~

    (zAge + zTOWRE_wordacc + zTOWRE_nonwordacc)*

    (item_type + zLength + zOrtho_N) +

    (1|subjectID) + (1|item_name),

    data = ML.all.correct, REML = FALSE)
summary(full.lmer.3)
```

- Notice that the `(something)*(something)` get you interactions and main effects for all possible pairs of variables in the first set `(zAge + zTOWRE_wordacc + zTOWRE_nonwordacc)` and the second set `(item_type + zLength + zOrtho_N)` of predictors

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

We can use the *anova()* function to do the comparison

```
anova(full.lmer.0, full.lmer.1, full.lmer.2, full.lmer.3)
```

- `anova(full.lmer.0, full.lmer.1, full.lmer.2, full.lmer.3)` compare the named models in pairs

See that *anova()* results present information criteria statistics for each model plus *likelihood ratio test* comparisons

Each step increase in complexity appears warranted by improved model fit to data

```
> anova(full.lmer.0, full.lmer.1, full.lmer.2, full.lmer.3)
Data: ML.all.correct
Models:
full.lmer.0: logrt ~ (1 | subjectID) + (1 | item_name)
full.lmer.1: logrt ~ zAge + zTOWRE_wordacc + zTOWRE_nonwordacc + (1 | subjectID) +
full.lmer.1:     (1 | item_name)
full.lmer.2: logrt ~ zAge + zTOWRE_wordacc + zTOWRE_nonwordacc + item_type +
full.lmer.2:     zLength + zOrtho_N + (1 | subjectID) + (1 | item_name)
full.lmer.3: logrt ~ (zAge + zTOWRE_wordacc + zTOWRE_nonwordacc) * (item_type +
full.lmer.3:     zLength + zOrtho_N) + (1 | subjectID) + (1 | item_name)
            Df    AIC    BIC logLik deviance    Chisq Chi Df Pr(>Chisq)
full.lmer.0  4 -17981 -17952 8994.4   -17989
full.lmer.1  7 -17982 -17931 8998.0   -17996   7.1782      3    0.06643 .
full.lmer.2 10 -18320 -18247 9169.7   -18340 343.5400      3    < 2e-16 ***
full.lmer.3 19 -18416 -18278 9227.0   -18454 114.5129      9    < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

Figure : Model comparisons

The distinction between exploratory and confirmatory studies may be useful here

Cumming (2014): between *pre specified* and *exploratory* or *question answering* and *question formulating* studies

- Confirmatory – pre specified – models test the effects of the variables or conditions you manipulated
- Exploratory – question formulating – models may be developed by building up complexity, with initially no clear idea about predictions

The distinction between exploratory and confirmatory studies may be useful here

Cumming (2014): between *pre specified* and *exploratory* or *question answering* and *question formulating* studies

- Confirmatory – pre specified – models test the effects of the variables or conditions you manipulated
- Exploratory – question formulating – models may be developed by building up complexity, with initially no clear idea about predictions

Examine the random 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

Examine the random 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

Examine the random 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

Examine the random 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

Model comparison and selection

Remember: Simplicity and parsimony

- Trade-off between too much and too little simplicity in model selection
- Building random slopes into your model may reveal that there is quite a bit of variation in some effects – enough random variation that fixed effects are not significant – though that variation may be of interest
- Estimation procedures may run into convergence problems where there is too much model complexity and not enough data

Model comparison and selection

Remember: Simplicity and parsimony

- Trade-off between too much and too little simplicity in model selection
- Building random slopes into your model may reveal that there is quite a bit of variation in some effects – enough random variation that fixed effects are not significant – though that variation may be of interest
- Estimation procedures may run into convergence problems where there is too much model complexity and not enough data

Model comparison and selection

Remember: Simplicity and parsimony

- Trade-off between too much and too little simplicity in model selection
- Building random slopes into your model may reveal that there is quite a bit of variation in some effects – enough random variation that fixed effects are not significant – though that variation may be of interest
- Estimation procedures may run into convergence problems where there is too much model complexity and not enough data