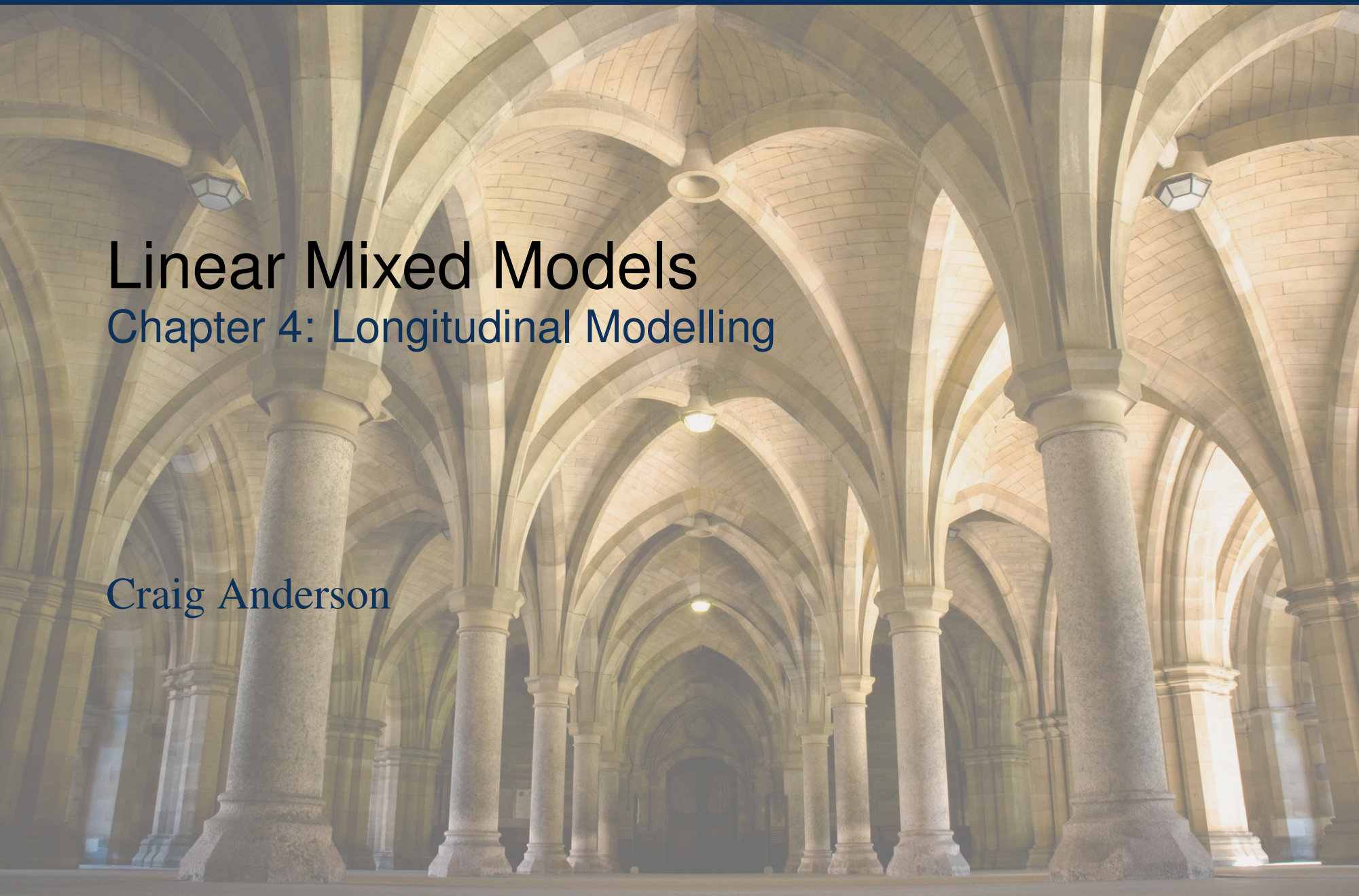
University of Glasgow



Linear Mixed Models

Chapter 4: Longitudinal Modelling

Craig Anderson

Longitudinal Data

Longitudinal studies

Designs in which the outcome variable is measured repeatedly over time (for at least some study participants).

Longitudinal designs allow investigation of events that occur in time, *e.g.* growth, aging or the temporal pattern of response to treatment.

Repeated measures

Older term applied to a special set of longitudinal designs characterised by measurement at a common set of occasions, usually in an experimental setting.

Analysis of repeated measures

- **Repeated measures** refers to multiple measurements on the same experimental unit (or subject).
- Measures taken on the same subject tend to be more similar than measures taken on different subjects.
- Measures made close in time on the same subject tend to be more highly correlated than measures made far apart in time.
- The analysis of repeated measures data accounts for the presence of correlation between the observations obtained on the same subject and for possible non-constant variances.

Analysis of repeated measures

Correlation due to measurements made on the same subject

- Two measures on the same subject are more likely to be closer to each other than two measures on the different subjects.
- They are positively correlated simply because they possess common effects from that same subject.
- This is comparable to measures on the same whole-plot unit in a split-plot type of experiment.

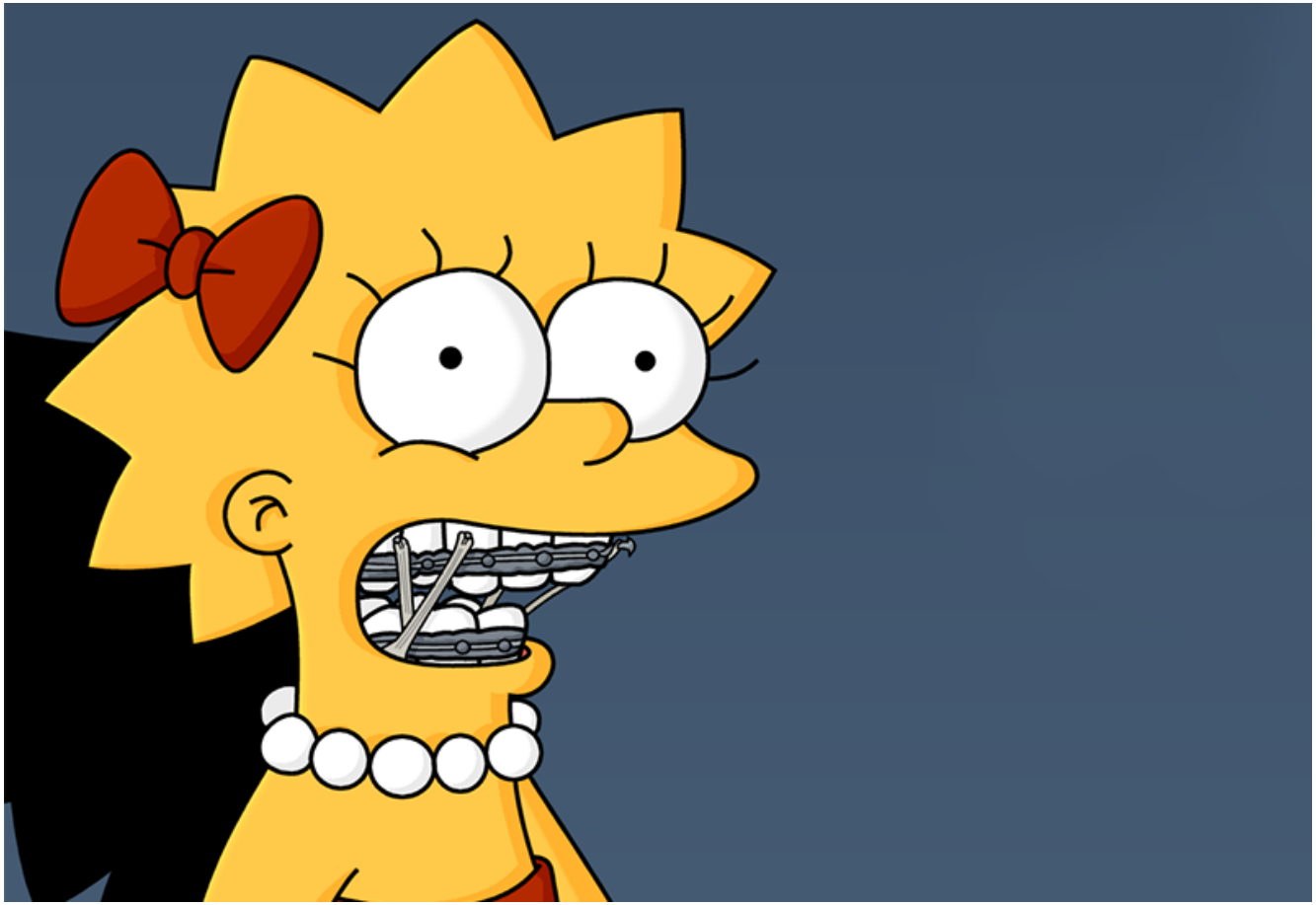
Analysis of repeated measures

Correlation due to measurements made close in time

- Two measures made close in time on the same subject are likely to be more highly correlated than two measures made far apart in time.
- This distinguishes repeated measures covariance from split-plot covariance structure.
- In a split-plot experiment, levels of sub-plot treatment are randomly assigned to sub-plot units within whole-plot units. This results in equal correlation between all pairs of measurements in the same whole-plot unit.

Orthodontics Example

- A study was carried out to measure the change in an orthodontic measure over time for a group of young patients.



Orthodontics Example

Study Design

- Investigators followed the growth of 27 children (16 male, 11 female) in North Carolina.
- The children received skull x-rays every two years from age 8 to age 14.
- From these x-rays, the investigators estimated the distance between the pituitary and pterygomaxillary fissure.

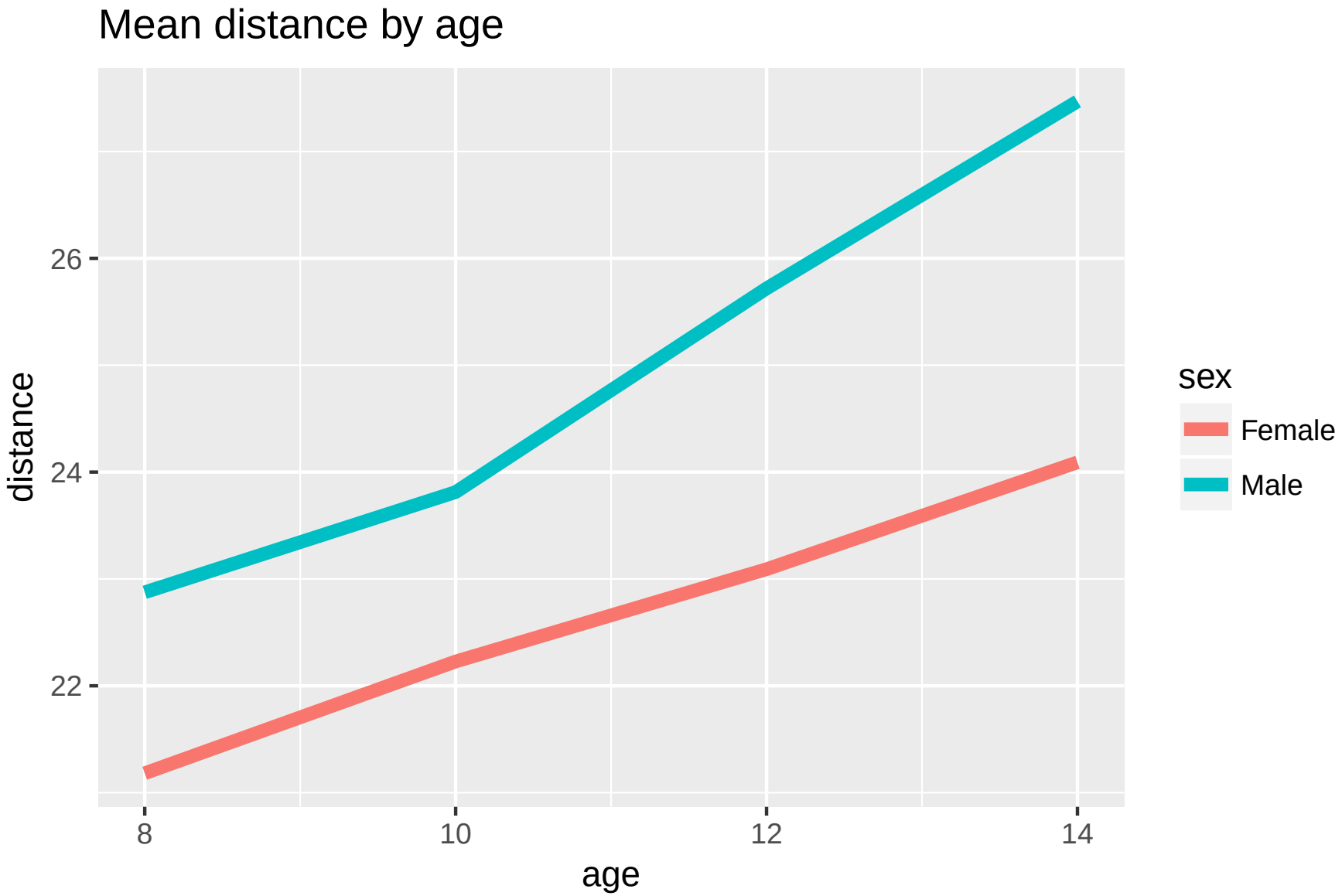
Orthodontics Example

Data

The dataset `Orthodontics` contains the following variables:

- `distance`: the distance between the pituitary and pterygomaxillary fissure (mm).
- `age`: the age of the subject (years).
- `Subject`: a factor indicating the subject on which the measurement was made.
- `Sex`: a factor indicating the sex of the patient.

Group profile plot



Mixed models for repeated measures data

Four-step procedure for mixed model analyses

1. Model the mean structure, usually by specification of the fixed effects.
2. Specify the covariance structures for between-subject and within-subject effects.
3. Fit the mean model, accounting for the covariance structure using GLS. This step might include making the mean model more parsimonious.
4. Make statistical inferences based on the results of Step 3. This step might also include making the mean model more parsimonious.

Step 1: the mean model

The model can be written in the form:

$$y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + e_{ijk}$$

where

- y_{ijk} is the kth hour **distance** measurement for the ith sex, jth subject.
- μ is the overall mean;
- α_i is the fixed effect for sex i;
- β_k is the fixed effect of the kth hour;
- $(\alpha\gamma)_{ik}$ is the fixed effect of the interaction between the ith sex and the kth hour;

Step 1: the mean model

- Obviously we only have one measurement of each patient at each age (ie no replicates).
- Any differences between individual subjects are measured via the random error terms.

Mean model

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik}.$$

Step 2: the covariance structure

- Recall that the variance structure of general linear mixed model has two components, **G** and **R**.
- **G** is the covariance structure for the random effects (analogous to between-subject variation).
- **R** is the the covariance structure for the within subject measurements (within-subject variation).
- Repeated measures are assumed to be correlated within subjects and independent across subjects.
- The covariance structure for our model is therefore typically defined by specifying a block diagonal form for **R**.

Types of covariance structure

Diagonal covariance matrix in each block

Example with four time measurements:

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \sigma_1^2 & 0 \\ 0 & 0 & 0 & \sigma_1^2 \end{bmatrix}$$

- Simplest covariance structure: equal variances, zero within-subject error correlation.
- This is usually not a reasonable structure for repeated measures data because the repeated measures within a subject are often correlated.

Types of covariance structure

Unstructured

Example with four time measurements:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix}$$

- Most complex covariance structure: unstructured, where observations for each pair of times have their own unique correlation.
- Many times this is more complicated than is necessary.

Types of covariance structure

Compound symmetry

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_B^2 + \sigma_E^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 \\ & \sigma_B^2 + \sigma_E^2 & \sigma_B^2 & \sigma_B^2 \\ & & \sigma_B^2 + \sigma_E^2 & \sigma_B^2 \\ & & & \sigma_B^2 + \sigma_E^2 \end{bmatrix}$$

Types of covariance structure

Compound symmetry

- The simplest correlation model is compound symmetry (CS), also referred to as exchangeable correlation structure.
- It assumes that correlation is constant regardless of the distance between time points.
- Might be reasonable in situations where the repeated measurements are not obtained over time.
- Unlikely to be a valid assumption for observations collected over time.

Types of covariance structure

First order autoregressive or AR(1) covariance structure

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

- The AR(1) model assumes that the correlation between all pairs of *adjacent* observations is ρ.
- In general, observations d units apart have correlation ρ^d.

Types of covariance structure

Toeplitz covariance structure

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ & 1 & \rho_1 & \rho_2 \\ & & 1 & \rho_1 \\ & & & 1 \end{bmatrix}$$

- Toeplitz covariance structure is similar to AR(1), but is more general.
- Still assumes that observations separated by a common distance (d) share the same correlation, but this correlation is no longer a power function of ρ.
- Instead, it can be any value between 0 and 1.

Types of covariance structure

AR(1) and Toeplitz covariance structures

+ The correlation between observations is a function of their distance in time, which appears to be a reasonable assumption in repeated measures.

- These structures do not model unequally spaced time points. In cases where this applies, we instead consider a spatial covariance structure.

Types of covariance structure

Spatial power covariance structure

$$\sigma^2 \begin{bmatrix} 1 & \rho^{|t_1-t_2|} & \rho^{|t_1-t_3|} & \rho^{|t_1-t_4|} \\ & 1 & \rho^{|t_2-t_3|} & \rho^{|t_2-t_4|} \\ & & 1 & \rho^{|t_3-t_4|} \\ & & & 1 \end{bmatrix}$$

- The spatial power structure provides a direct generalization of the AR(1) structure, allowing for unequal spacing of the repeated measures.
- Useful when the time points are unequally spaced and the correlations decline as a function of time difference.

Types of covariance structure

Spatial exponential covariance structure

$$\sigma^2 \begin{bmatrix} 1 & e^{-\frac{d_{12}}{\rho}} & e^{-\frac{d_{13}}{\rho}} & e^{-\frac{d_{14}}{\rho}} \\ & 1 & e^{-\frac{d_{23}}{\rho}} & e^{-\frac{d_{24}}{\rho}} \\ & & 1 & e^{-\frac{d_{34}}{\rho}} \\ & & & 1 \end{bmatrix}$$

- An alternative specification based on exponential functions rather than powers.
- Again, useful for modelling correlations for repeated measures obtained at unequally spaced time points.

Selecting an appropriate covariance model

- Using a covariance structure that is too simple means that important correlations are ignored. This could lead to increased Type I error rates for the fixed effects.
- Using a covariance structure that is more complex than necessary could lead to reduced power and efficiency in the test of fixed effects.

How to select a covariance model

- Use subject matter knowledge and information from the data collection process and previous relevant studies.
- Use graphical tools to examine the patterns of correlations over time.
- Compare the relative fit of various covariance structures (likelihood ratio tests for nested variance structures, information criteria otherwise).

Fitting the mixed model in R

- The `lme4` package does not allow you to specify correlation structure.
- Therefore we instead need to use the older `nlme` package.

```
library(nlme)
mod1 <- gls(distance ~ Sex * I(age-11), data=Orthodont,
            correlation = corSymm(form = ~ 1|Subject),
            weights=varIdent(form = ~ 1|age))
```

Fitting the mixed model in R

- The `gls()` function fits a linear model structure using generalised least squares.
- We rescale the age variable to centre it around the mean (11) - the `I()` function allows us to pass this into the model rather than creating a new variable.
- We can specify a `correlation` argument - here we make it unstructured using `corSymm()`.
- We can specify a `weights` argument - here we use `varIdent()` to allow each age group to have a different variance.

Output: Fixed Effects

```
summary(mod1)

Coefficients:
                            Value Std.Error  t-value p-value
(Intercept)             24.937123 0.4728666 52.73606  0.0000
SexFemale               -2.271743 0.7408396 -3.06644  0.0028
I(age - 11)              0.826804 0.0822177 10.05627  0.0000
SexFemale:I(age - 11) -0.350439 0.1288104 -2.72058  0.0076
```

- We obtain estimates for our fixed effects.
- All of the terms in our model appear to be significant.

Output: Correlation Matrix

- The `nlme` package has various functions to explore the correlation structures.
- Here we use `corMatrix()` to extract correlation matrices.

```
corMatrix(mod1$modelStruct$corStruct)[[1]]
```

```
          [,1]      [,2]      [,3]      [,4]
[1,] 1.0000000 0.5681970 0.6589493 0.5220393
[2,] 0.5681970 1.0000000 0.5806064 0.7249210
[3,] 0.6589493 0.5806064 1.0000000 0.7396207
[4,] 0.5220393 0.7249210 0.7396207 1.0000000
```

Output: Correlation Matrix

- This is the correlation matrix for the first block of R (patient 1).
- The matrix will be the same for every other patient, since every block has identical covariance structure.

```
          age 8     age 10    age 12    age 14
age 8  1.0000000 0.5681970 0.6589493 0.5220393
age 10 0.5681970 1.0000000 0.5806064 0.7249210
age 12 0.6589493 0.5806064 1.0000000 0.7396207
age 14 0.5220393 0.7249210 0.7396207 1.0000000
```

Information criteria

- Choosing a suitable covariance structure is crucial to mixed models, but often there is more than one sensible option.
- There are two main ways of choosing the best structure - a likelihood ratio test or information criteria.
- Information criteria are suitable in all cases regardless of which structures we are comparing.
- Likelihood ratio tests can only be used when comparing two nested covariance structure.

Information criteria

- **Information criteria** can be used to objectively select a model with the most appropriate covariance structure.
- The *smaller* the information criterion value is, the better the model is.
- Information criteria attach penalties to the -2 Res Log Likelihood value; the more parameters included, the bigger the penalty.

Information criteria

- Two commonly used information criteria are **Akaike's AIC** and **Schwarz's BIC**.
- Generally speaking, BIC tends to choose less complex models than AIC.
- Choosing a model that is too simple inflates Type I error rate, but choosing a model which is too complex can lead to a loss of power.
- If Type I error control is the highest priority, we might want to choose AIC.
- On the other hand, if loss of power is more of a concern, BIC might be preferable.

Information criteria

Akaike Information Criterion

$$\text{AIC} = -2\log\text{likelihood} + 2k$$

AIC (with finite sample size correction)

$$\text{AICc} = -2\log\text{likelihood} + 2k + \frac{2k(k+1)}{(n-k-1)}$$

Bayesian Information Criterion

$$\text{BIC} = -2\log\text{likelihood} + k\log(n)$$

Choosing a covariance structure

- The basic idea for repeated measures analysis is that, *among plausible within-subject covariance models* given a particular study, the model that minimizes AIC (or BIC) is preferable.
- When the two criteria are close, the simpler model is generally preferred.

Orthodontics example

- We can now think about testing alternative covariance structures for our orthodontics example.
- Based on our results, we might think about a compound symmetry model.
- We can also look at an AR1 model for the purposes of comparison.
- We can fit models with each of these and compare them using information criteria.

Compound symmetry

```
mod2 <- gls(distance ~ Sex * I(age-11), data=Orthodont,
            correlation = corCompSymm(form = ~ 1|Subject),
            weights=varIdent(form = ~ 1|age))
```

```
           [,1]      [,2]      [,3]      [,4]
[1,] 1.0000000 0.6352894 0.6352894 0.6352894
[2,] 0.6352894 1.0000000 0.6352894 0.6352894
[3,] 0.6352894 0.6352894 1.0000000 0.6352894
[4,] 0.6352894 0.6352894 0.6352894 1.0000000
```

- Here we have $\hat{\rho} = 0.635$.

AR(1) model

```
mod3 <- gls(distance ~ Sex * I(age-11), data=Orthodont,
            correlation = corAR1(form = ~ 1|Subject),
            weights=varIdent(form = ~ 1|age))
```

```
           [,1]      [,2]      [,3]      [,4]
[1,] 1.0000000 0.6332579 0.4010156 0.2539463
[2,] 0.6332579 1.0000000 0.6332579 0.4010156
[3,] 0.4010156 0.6332579 1.0000000 0.6332579
[4,] 0.2539463 0.4010156 0.6332579 1.0000000
```

- Here we have $\hat{\rho} = 0.633$.

Comparison of covariance structures

	Unstructured	Compound Symmetry	AR(1)
AIC	452.5	450.0	460.8
AICc	457.0	451.8	462.6
BIC	489.6	473.8	484.6

- Based on all three criteria, compound symmetry is the preferred model.
- However, we note that AR(1) is second best for BIC, while unstructured is second best for AIC.

Likelihood ratio test

Comparison of models with nested covariance structures

1. Fit the two models with nested covariance structures.
2. Compute the difference of the -2 Res log likelihood values between the two models.
3. Compute the difference in the number of parameters estimated from the two models.
4. Compare the result from Step 2 with a χ^2 distribution with the degrees of freedom obtained from Step 3.

Note: A likelihood ratio test to compare two different covariance structures can only be applied to two nested covariance structures.

Likelihood ratio test

- The null hypothesis is that the special case fits the data just as well as the more general covariance structure.
- It can be shown that under the null hypothesis, the -2 log of the ratio of the two likelihood functions is approximately distributed as $\chi^2(r)$.
- Here r is the difference in the number of parameters estimated from the two models.
- A small p-value indicates the more general covariance structure is needed.

Likelihood ratio test

- The compound symmetry structure is a special case of unstructured variance.
- In our example we obtain it by fixing
  $\sigma_{12} = \sigma_{13} = \ldots = \sigma_{34}$
- Our compound symmetry structure has 5 fewer parameters than the unstructured one.
- We therefore use the $\chi^2(5)$ distribution.

Likelihood ratio test

- We have to be careful that we remember the limitations of a likelihood ratio test.
- Recall that this approach is conservative when we set a parameter equal to the boundary constraint (for example, variance component equals zero).
- In that case, the asymptotic distribution for the likelihood ratio statistics is often a mixture of chi-squared distributions.

Likelihood ratio tests for orthodontics data

```
anova(mod1, mod2)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod1     1 14 452.5468 489.5683 -212.2734
mod2     2  9 449.9724 473.7719 -215.9862 1 vs 2 7.425576  0.1909
```

- The p-value suggests no evidence of a difference between the models.
- We therefore prefer the model with fewer parameters (compound symmetry).

Compound symmetry

- When we fitted the compound symmetry model we included the command `weights=varIdent(form = ~1|age)`.
- This allows for the possibility of non-constant variance between timepoints (ie four different variance parameters).
- We can test whether this assumption is necessary.

Compound symmetry

```
mod2 <- gls(distance ~ Sex * I(age-11), data=Orthodont,
            correlation = corCompSymm(form = ~ 1|Subject),
            weights=varIdent(form = ~ 1|age))

mod4 <- gls(distance ~ Sex * I(age-11), data=Orthodont,
            correlation = corCompSymm(form = ~ 1|Subject))

anova(mod2,mod4)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mod2     1  9 449.9724 473.7719 -215.9862
mod4     2  6 445.7572 461.6236 -216.8786 1 vs 2 1.784873  0.6182
```

- The p-value suggests no evidence of a difference between the models.
- We therefore prefer the model with constant variance (mod4) since it has fewer parameters.

Inference

- Now that we have identified the best covariance structure, we can carry out inference.
- We can explore the results for both the mean model and the covariance structure separately.
- We make final adjustments to the mean model at this stage if required.

Inference: Mean model

```
summary(mod4)

Coefficients:
                            Value Std.Error  t-value p-value
(Intercept)             24.968750 0.4860003 51.37600  0.0000
SexFemale               -2.321023 0.7614161 -3.04830  0.0029
I(age - 11)              0.784375 0.0775011 10.12082  0.0000
SexFemale:I(age - 11) -0.304830 0.1214209 -2.51052  0.0136
```

- All of our fixed effects are significant ($p < 0.05$).
- $\hat{\mu} = 24.97$, $\hat{\alpha} = -2.32$, $\hat{\beta} = 0.78$ and $\hat{\alpha\beta} = -0.30$.
- Note that these estimates are different to those obtained previously when we had a different covariance structure (unstructured).

Inference: Mean model

Fixed effect estimates (unstructured covariance)

```
Coefficients:
                         Value Std.Error  t-value p-value
(Intercept)          24.937123 0.4728666 52.73606  0.0000
SexFemale            -2.271743 0.7408396 -3.06644  0.0028
I(age - 11)           0.826804 0.0822177 10.05627  0.0000
SexFemale:I(age - 11) -0.350439 0.1288104 -2.72058  0.0076
```

Fixed effect estimates (compound symmetry)

```
Coefficients:
                         Value Std.Error  t-value p-value
(Intercept)          24.968750 0.4860003 51.37600  0.0000
SexFemale            -2.321023 0.7614161 -3.04830  0.0029
I(age - 11)           0.784375 0.0775011 10.12082  0.0000
SexFemale:I(age - 11) -0.304830 0.1214209 -2.51052  0.0136
```

Getting the right covariance structure is important!

Inference: Covariance structure

- We can also explore our final covariance structure.
- It is important that we understand the implications of our variance and covariance parameters.
- This helps to explain the relationship between different timepoints.

Inference: Covariance structure

Compound symmetry

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ & 1 & \rho & \rho \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_B^2 + \sigma_E^2 & \sigma_B^2 & \sigma_B^2 & \sigma_B^2 \\ & \sigma_B^2 + \sigma_E^2 & \sigma_B^2 & \sigma_B^2 \\ & & \sigma_B^2 + \sigma_E^2 & \sigma_B^2 \\ & & & \sigma_B^2 + \sigma_E^2 \end{bmatrix}$$

Inference: Covariance Structure

```
getVarCov(mod4)

Marginal variance covariance matrix
       [,1]   [,2]   [,3]   [,4]
[1,] 5.2207 3.2986 3.2986 3.2986
[2,] 3.2986 5.2207 3.2986 3.2986
[3,] 3.2986 3.2986 5.2207 3.2986
[4,] 3.2986 3.2986 3.2986 5.2207
```

- The estimated variance term $\hat{\sigma}^2$ is 5.22.
- We can also compute our correlation estimate $\hat{\rho} = 3.2986/5.2207 = 0.63$.
- There is a moderately strong correlation between a child's orthodontic distance at different ages.

Inference: Correlation Matrix

```
corMatrix(mod4$modelStruct$corStruct)[[1]]

          [,1]      [,2]      [,3]      [,4]
[1,] 1.0000000 0.6318381 0.6318381 0.6318381
[2,] 0.6318381 1.0000000 0.6318381 0.6318381
[3,] 0.6318381 0.6318381 1.0000000 0.6318381
[4,] 0.6318381 0.6318381 0.6318381 1.0000000
```

- We can also obtain the correlation matrix directly.
- This is the same as we would obtain by dividing the covariance matrix by the variance $\hat{\sigma}^2$.

Conclusions: Orthodontics Example

- The orthodontic distance tends to be smaller for girls than boys.
- The distance increases as a child gets older.
- There is an age-sex interaction - the increase in distance is smaller for girls.
- There is a moderately strong correlation between repeated measures on the same child.
- This correlation remains the same regardless of how many years apart the observations are taken.

Longitudinal Modelling

- The phrase *repeated measures* is typically used for designed experiments where data are obtained at fixed timepoints.
- In many cases, our data will not have this 'tidy' structure.
- Instead each subject may have different numbers of observations taken at different timepoints.
- This is typically known as *longitudinal data*.

Longitudinal Modelling

- Longitudinal modelling is similar to repeated measures analysis in many ways.
- We still believe that observations taken on the same subject will have more in common.
- We still believe that observations closer together in time may have more in common.
- However, it is no longer possible to model correlation using the 'simple' structures used for repeated measures analysis.

Longitudinal Modelling

- The correlation structures used in repeated measures analysis require us to estimate the correlation between our observed timepoints.
- When we have fixed timepoints, we only have to estimate a small number of correlation parameters, and we have lots of data for each one.
- If we have different timepoints for each individual, we need to estimate correlations between each possible pair, but lack the data to do so.
- Therefore we have to think about alternative approaches for this type of data.

Random coefficient models

- In the **random coefficient** model (Chapter 3), the regression coefficients for continuous explanatory variables are assumed to be *random effects*.
- Data arise from independent subjects or clusters from a larger population of interest.
- The regression model for each subject or cluster can be assumed to be a random deviation from some population regression model.
- We can use these models for longitudinal data.
- Here we can think about time as our continuous explanatory variable - we are trying to estimate a 'slope' for each subject.

Random coefficient models

- For the random intercept and random slope we assume

$$\begin{bmatrix} a_i \\ b_i \end{bmatrix} \overset{\text{iid}}{\sim} N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{bmatrix} \right)$$

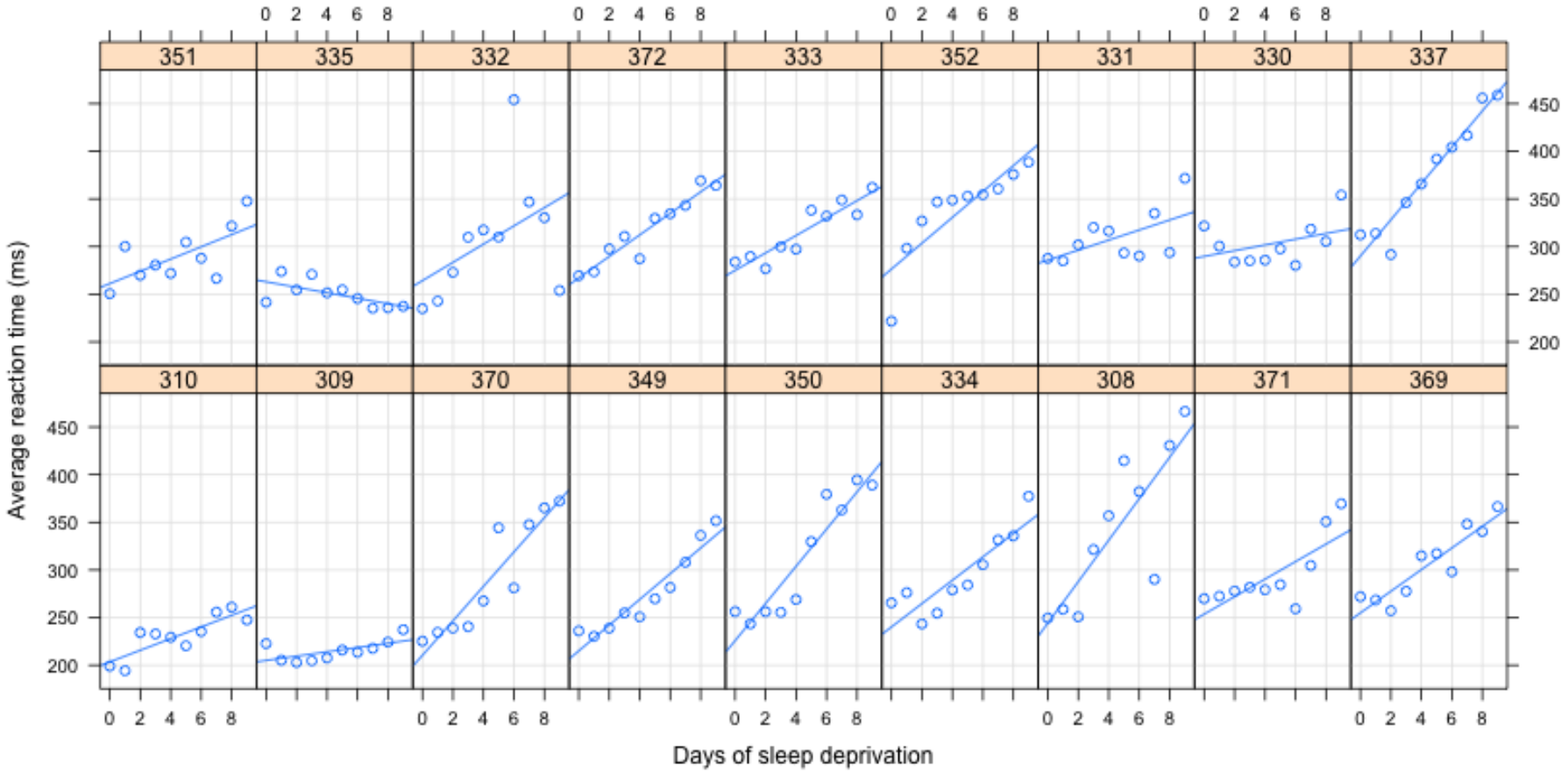
- The fixed effects of the model are the intercept α and the slope β.
- These are the expected values of the intercepts and slopes for the population.
- We can also obtain predicted values (EBLUPs) for each subject - this shows the predicted response over time for each subject.

Example

Sleep Deprivation Data

- This laboratory experiment measured the effect of sleep deprivation on cognitive performance.
- There were 18 subjects, chosen from the population of interest (long-distance truck drivers).
- The trial was carried out over 10 days, and the subjects were restricted to 3 hours sleep per night.
- Each subject's reaction time was measured daily. The reaction time in this dataset is the average of several measurements.
- These data are balanced in that each subject is measured the same number of times and on the same occasions.

Reaction time versus days, by subject



Comments on the sleep data plot

- The plot is a “trellis” or “lattice” plot where the data for each subject are presented in a separate panel.
- A reference line fit by simple linear regression to the panel’s data has been added to each panel.
- The panels have been ordered not by subject number (which is essentially a random order) but according to increasing intercept for the simple linear regression.
- If the slopes and the intercepts are highly correlated we should see a pattern across the panels in the slopes.

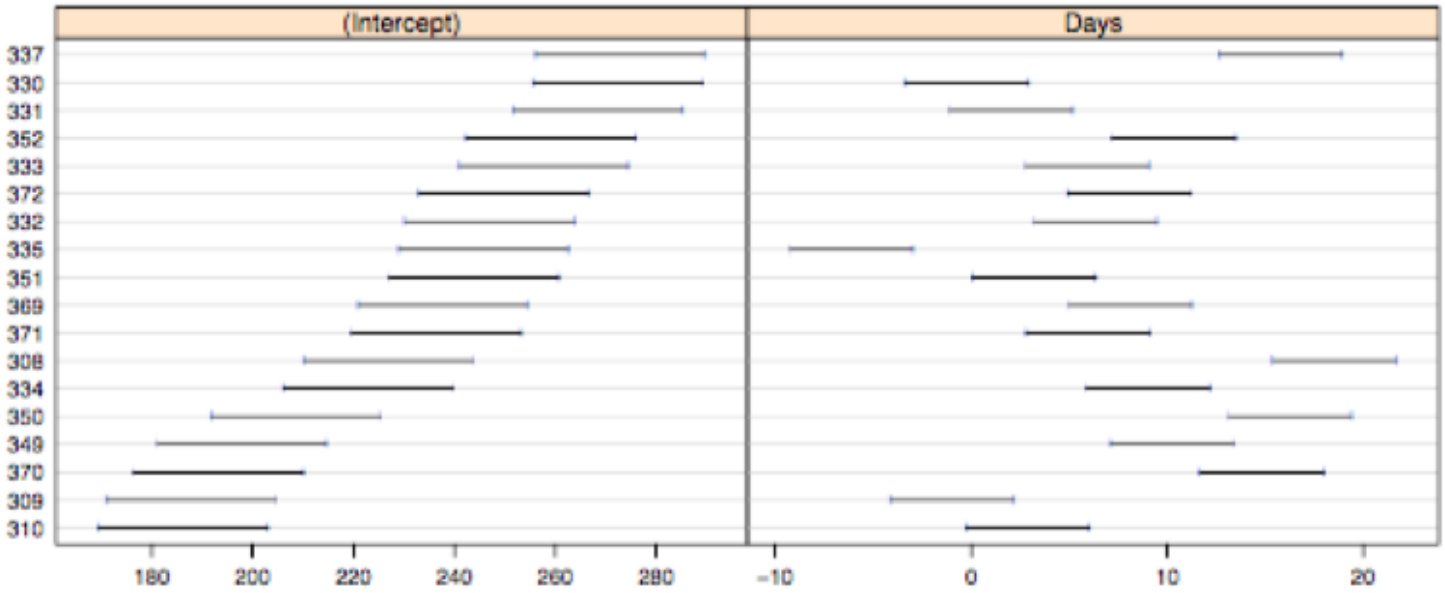
Assessing the linear fits

- In most cases a simple linear regression provides an adequate fit to the within-subject data.
- Patterns for some subjects (e.g. 350, 352 and 371) deviate from linearity but the deviations are neither widespread nor consistent in form.
- There is considerable variation in the intercept (estimated reaction time without sleep deprivation) across subjects (ranges from 200-300 ms).
- There is also a great deal of variation in the slope (increase in reaction time per day) across subjects (ranges from 0-20 ms/day).

Assessing the linear fits

- We can examine this variation further by plotting confidence intervals for these intercepts and slopes.
- Because we use a pooled variance estimate and have balanced data, the intervals have identical widths.
- We again order the subjects by increasing intercept so we can check for relationships between slopes and intercepts

95% CI on within-subject intercept and slope



- These intervals reinforce our earlier impressions.
- There is considerable variability between subjects in both intercept and slope but little evidence of a relationship *between* intercept and slope.

A preliminary mixed-effects model

- We can use a linear mixed model to explore these relationships further.
- The fixed effects $[\beta_1, \beta_2]'$ are the representative intercept and slope for the population.
- The random effects $\mathbf{b}_i = [b_{i1}, b_{i2}], i = 1, \ldots, 18$ are the deviations in intercept and slope associated with subject i.
- The random effects vector, $\mathbf{b}$, consists of the 18 intercept effects followed by the 18 slope effects.

Fitting the model

```
(fm1 <- lmer(Reaction ~ Days + (Days | Subject),
+     sleepstudy))

Linear mixed model fit by REML
Formula: Reaction ~ Days + (Days | Subject)
   Data: sleepstudy
  AIC  BIC logLik deviance REMLdev
 1756 1775 -871.8     1752    1744
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 612.095  24.7405
          Days         35.071   5.9221  0.065
 Residual             654.944  25.5919

Number of obs: 180, groups: Subject, 18
Fixed effects:
            Estimate Std. Error t value
(Intercept)  251.405      6.825   36.84
Days          10.467      1.546    6.77
Correlation of Fixed Effects:
     (Intr)
Days -0.138
```

Terms and matrices

- The term `Days` in the formula generates a model matrix $\mathbf{X}$ with two columns, the intercept column and the numeric `Days` column.
- Note that the intercept column is included by default unless you specify otherwise.
- The term `(Days|Subject)` generates a vector-valued random effect (intercept and slope) for each of the 18 levels of the `Subject` factor.

A model with uncorrelated random effects

- The data plots gave little indication of a systematic relationship between a subject's random effect for slope and his/her random effect for the intercept.
- The fitted model shows that the estimated correlation between these effects is quite small (0.065).
- We should therefore consider a model with uncorrelated random effects.

A model with uncorrelated random effects

- In the formula for an `lmer` model, distinct random effects terms are modelled as being independent.
- We therefore specify the model with two distinct random effects terms, each of which has `Subject` as the grouping factor.
- The model matrix for one term is intercept only (`1`) and for the other term is the column for `Days` only, which can be written `0+Days`.
- The expression `Days` generates a column for `Days` and an intercept. To suppress the intercept we add `0+` to the expression (`-1` also works).

A mixed-effects model with independent random effects

```
Linear mixed model fit by REML
  Formula: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
     Data: sleepstudy
    AIC  BIC logLik deviance REMLdev
   1754 1770 -871.8     1752    1744
  Random effects:
   Groups   Name        Variance Std.Dev.
   Subject  (Intercept) 627.577  25.0515
   Subject  Days         35.852   5.9876
   Residual             653.594  25.5655
  Number of obs: 180, groups: Subject, 18
  Fixed effects:
              Estimate Std. Error t value
  (Intercept)  251.405      6.885   36.51
  Days          10.467      1.559    6.71
  Correlation of Fixed Effects:
       (Intr)
Days -0.184
```

Comparing the models

- Model `fm1` can be described as containing model `fm2`.
- If the parameter values for model `fm1` were constrained so as to force the correlation (and hence the covariance) to be zero, we would get model `fm2`.
- The value `0`, to which the correlation is constrained, is not on the boundary of the allowable parameter values.
- In these circumstances a likelihood ratio test and a reference distribution of a χ^2 on 1 degree of freedom is suitable.

Comparing the models

```
anova(fm2, fm1)
Data: sleepstudy
Models:
fm2: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
fm1: Reaction ~ Days + (Days | Subject)
    Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm2  5 1762.05 1778.01 -876.02
fm1  6 1763.99 1783.14 -875.99 0.0609      1      0.805
```

- The large p-value indicates that we would not reject `fm2` in favor of `fm1`.
- We therefore prefer the more parsimonious `fm2`.
- Our conclusion is consistent with the AIC and BIC values for which "smaller is better".

Likelihood ratio tests on variance components

- We now compare the current model to a model without the random slope to see if we can remove another variance component.
- The likelihood ratio is a reasonable test statistic for the comparison but the "asymptotic" reference distribution of a χ^2 does not apply because the parameter value being tested is on the boundary.
- The p-value computed using the χ^2 reference distribution will be conservative (i.e. greater than the p-value that would be obtained through simulation).

Likelihood ratio tests on variance components

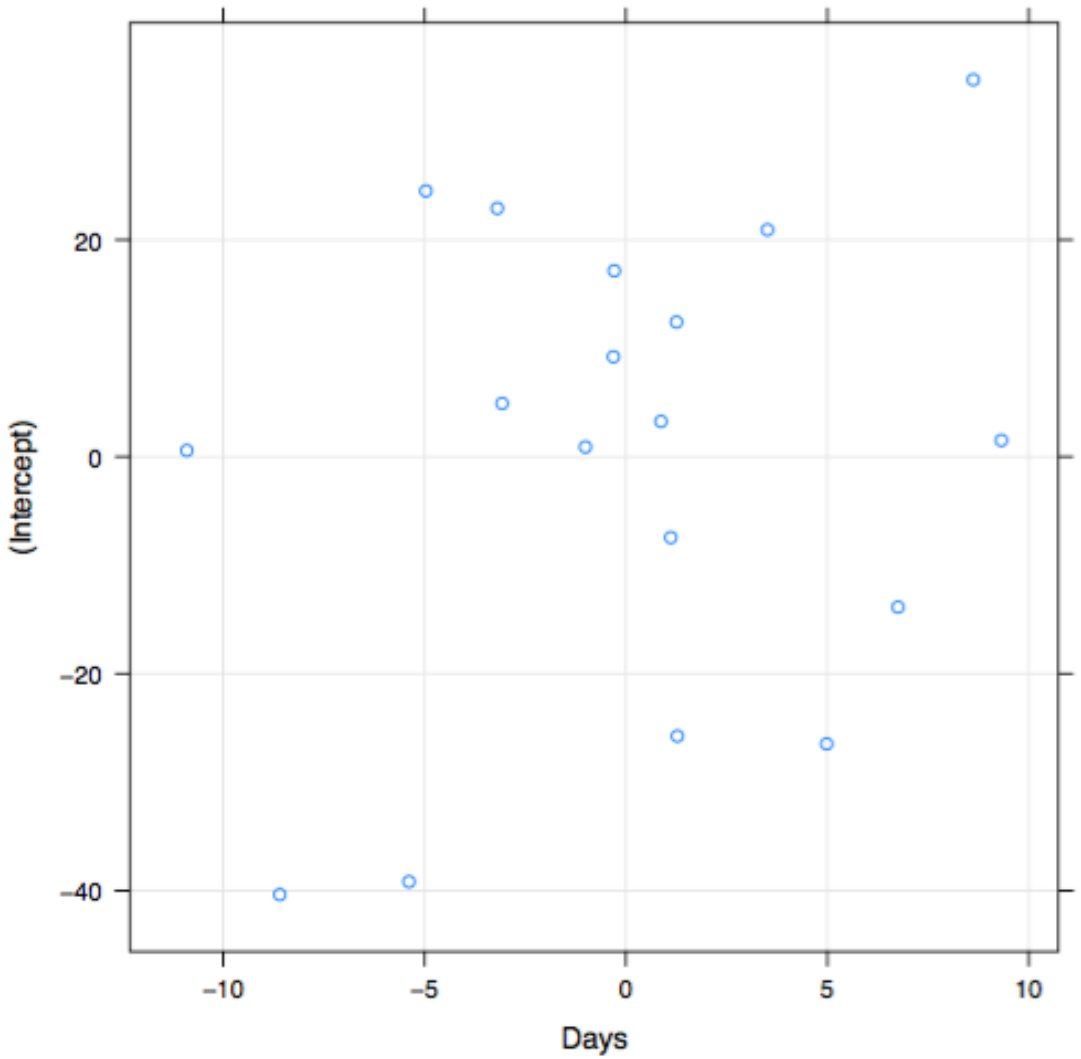
```
fm3 <- lmer(Reaction ~ Days + (1 | Subject), sleepstudy)
anova(fm3, fm2)
Data: sleepstudy
Models:
fm3: Reaction ~ Days + (1 | Subject)
fm2: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
    Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm3  4 1802.10 1814.87 -897.05
fm2  5 1762.05 1778.01 -876.02 42.053      1  8.885e-11
```

- The p-value is $<< 0.05$.
- This suggests that we should keep the random slope in the model.

Conditional means of the random effects

```
> (rr2 <- ranef(fm2))
$Subject
    (Intercept)        Days
308   1.5138200   9.3232135
309 -40.3749105  -8.5989183
310 -39.1816682  -5.3876346
330  24.5182907  -4.9684965
331  22.9140346  -3.1938382
332   9.2219311  -0.3084836
333  17.1560765  -0.2871973
334  -7.4515945   1.1159563
335   0.5774094 -10.9056435
337  34.7689482   8.6273639
349 -25.7541541   1.2806475
350 -13.8642120   6.7561993
351   4.9156063  -3.0750415
352  20.9294539   3.5121076
369   3.2587507   0.8730251
370 -26.4752098   4.9836365
371   0.9055257  -1.0052631
372  12.4219020   1.2583667
```

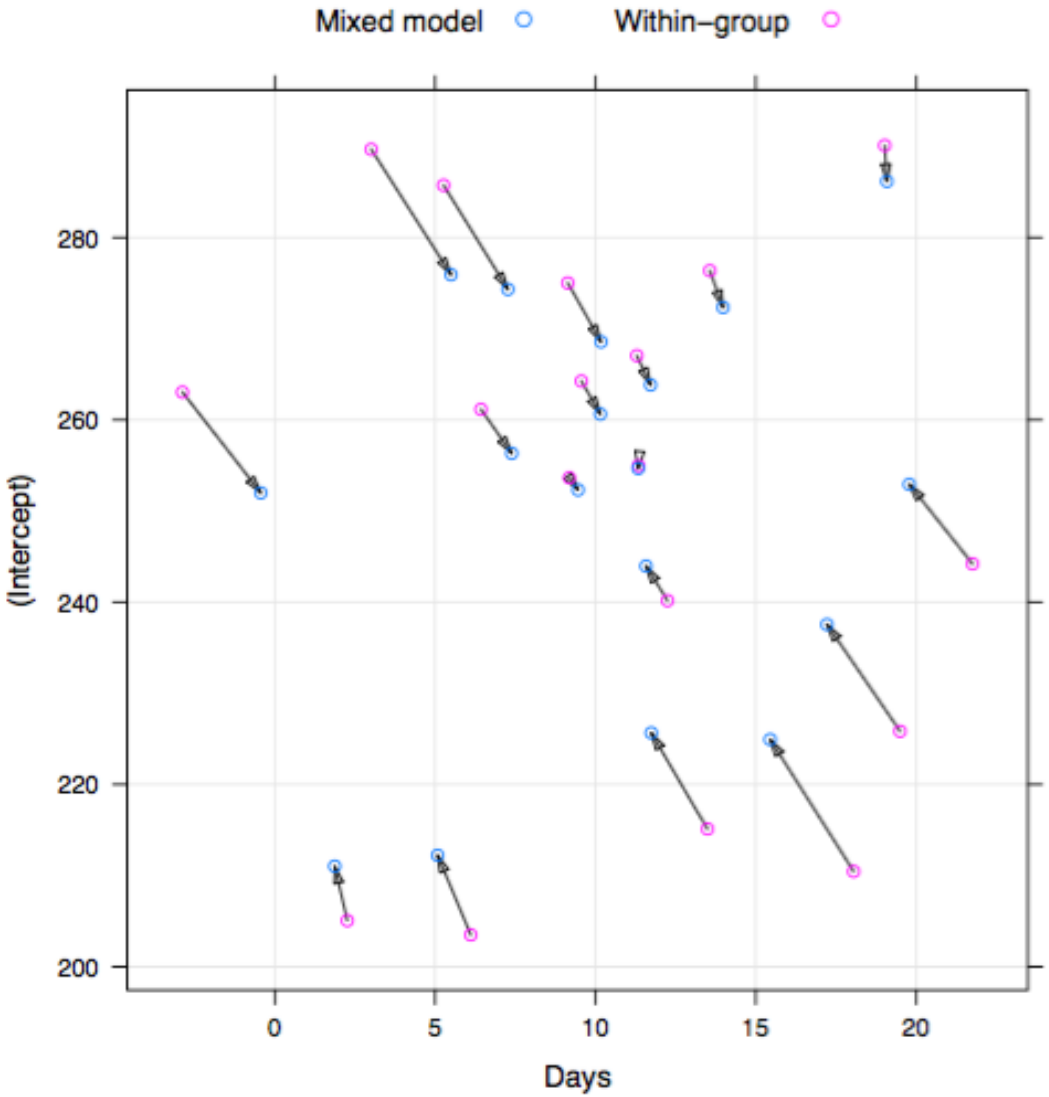
Scatterplot of the conditional means



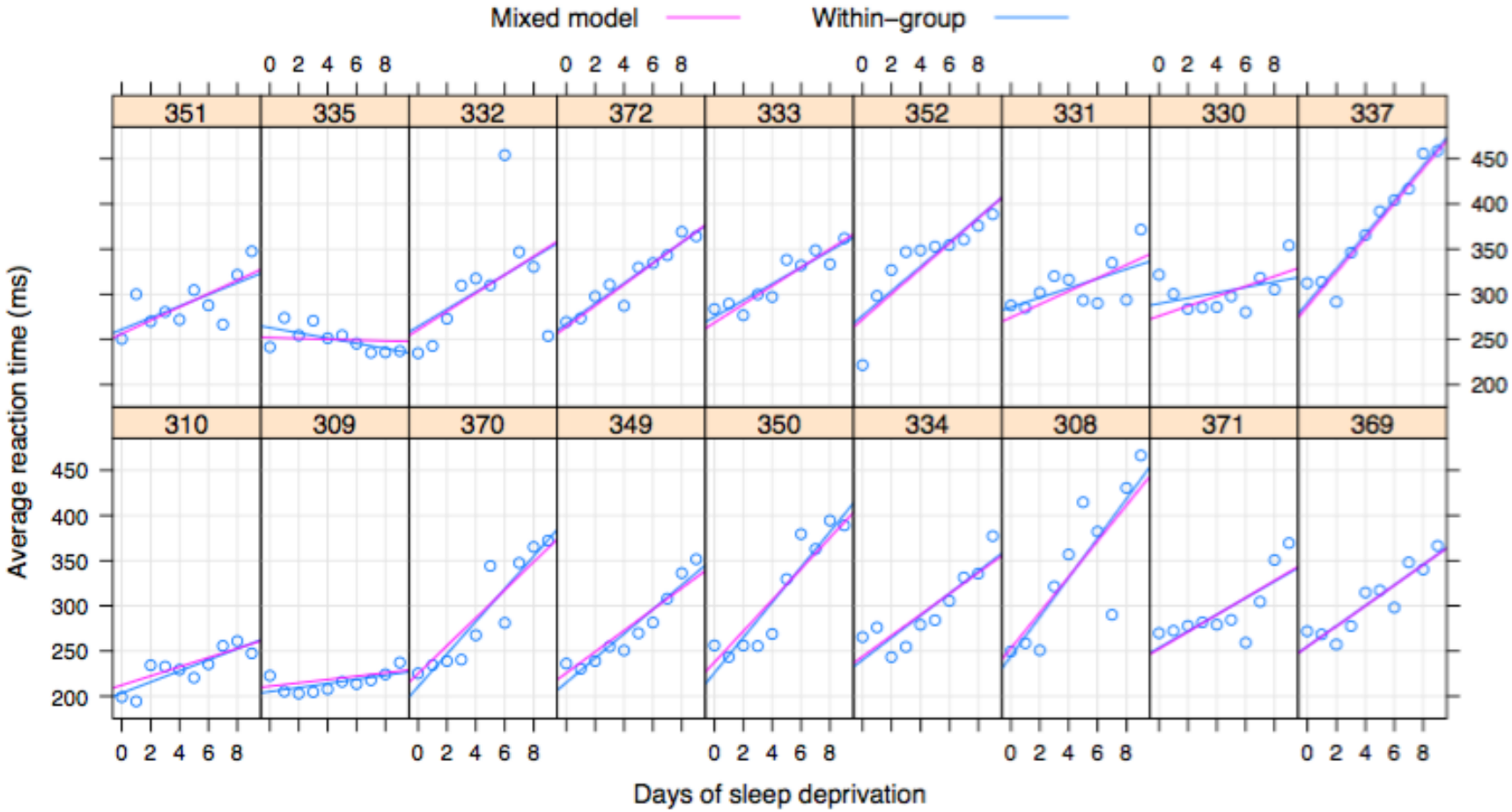
Comparing within-subject coefficients

- We can combine the conditional means of the random effects and the estimates of the fixed effects to get conditional means of the within-subject coefficients.
- These conditional means will be "shrunken" towards the fixed-effects estimates relative to the estimated coefficients from each subject's data.
- John Tukey called this "borrowing strength" between subjects.

Estimated within-group coefficients and BLUPs



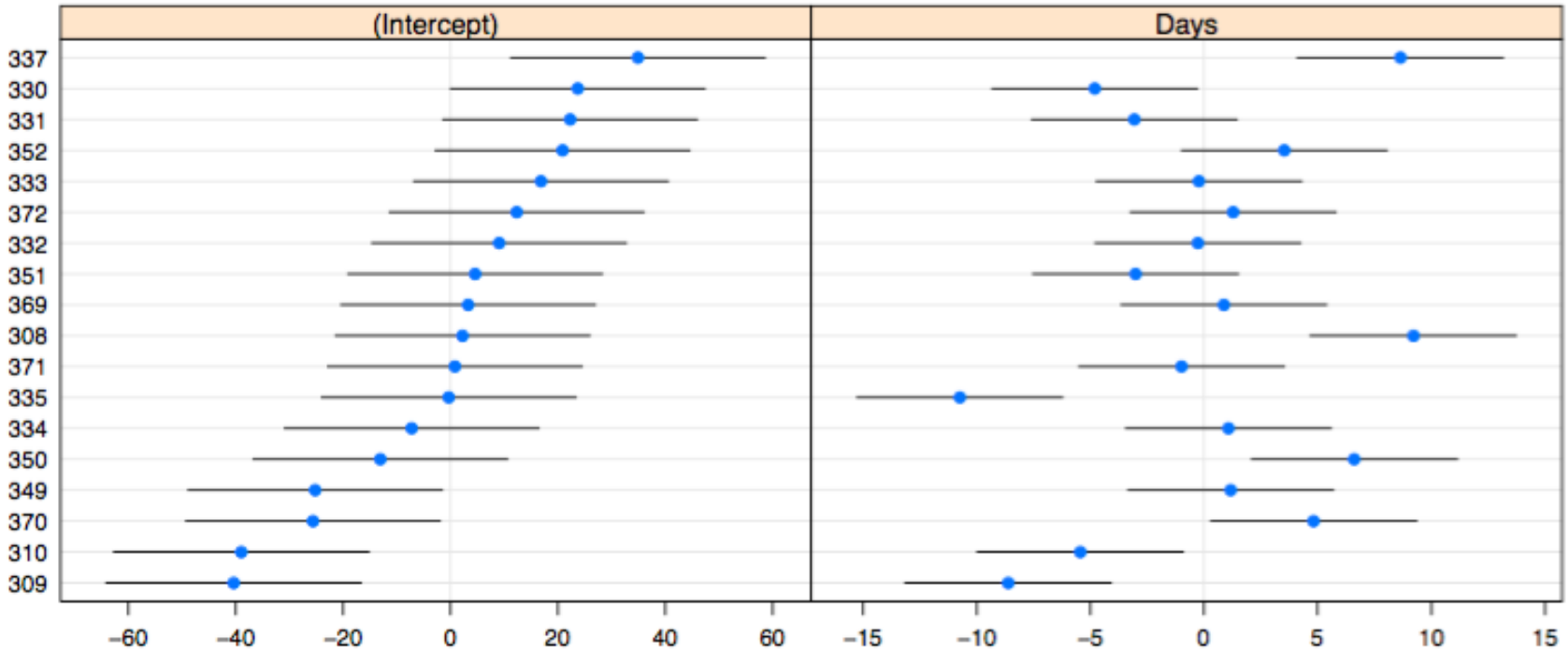
Observed and fitted



Comparing within-subject coefficients

- Plotting the shrinkage of the within-subject coefficients shows that some of the coefficients are considerably shrunken toward the fixed-effects estimates.
- However, comparing the within-group and mixed model fitted lines shows that large changes in coefficients occur in the noisy data.
- Within-group coefficients which were estimated with high precision are not changed substantially.

Plot of prediction intervals for the random effects



- Each set of prediction intervals have constant width because of the balance in the experiment.

Example

Bangladesh Growth Data

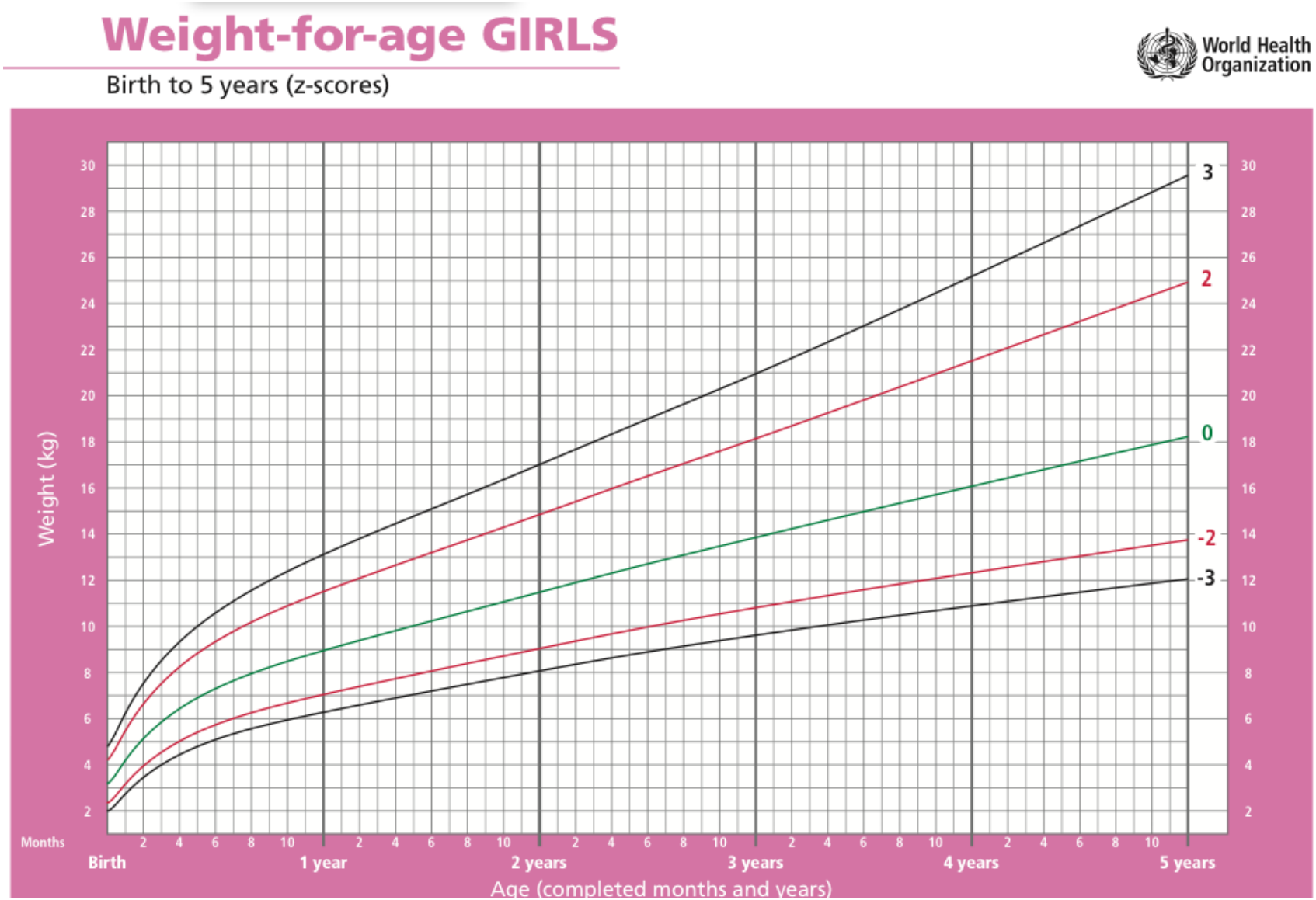
- A study was carried out to monitor the growth of children in Bangladesh.
- 700 children were monitored from birth until age 2.
- There were a total of 15 scheduled healthcare visits for each child over these two years.
- At each visit, the child had their height and weight measured.
- These data are almost balanced, and we will consider them to be so for the purposes of our analysis.

Gates Foundation

Bangladesh Growth Example

- We will look at the height data in today's lecture.
- We have measurements of height (in cm) over time.
- However, we know that all children grow over time - they don't get smaller.
- We are really interested in a child's relative growth.

World Health Organisation Growth Chart



Standardised Z-scores

- We use the WHO growth chart to obtain height-for-age Z-scores (HAZ).
- A HAZ score compares a child's height to the global reference population at the same age.
- For example, a child who has a Z-score of -1 at an age of 50 days is one standard deviation below the global average for children aged 50 days.
- We will model changes in this Z-score over time - an increase suggests a child is growing faster than average, a decrease suggests their growth is slower than average.

Model

$$y_{ij} = b_{i0} + b_{i1}x_{ij} + e_{ij}$$

where

- y_{ij} is the HAZ score for the ith child at the jth measurement, $i = 1, \ldots, 700$ and $j = 1, \ldots, 15$.
- x_{ij} is the age in days of child ith at their jth measurement.
- b_{i0} is the intercept for the ith child. This is a random effect because child is a random effect.
- b_{i1} is the slope for the ith child. This is also a random effect because child is a random effect.
- e_{ij} is the random error, assumed i.i.d. $N(0, \sigma_E^2)$.

Model

- For the random intercept and random slope we assume

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \overset{\text{iid}}{\sim} N\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \begin{bmatrix} \sigma^2_{B_0} & \sigma_{B_0B_1} \\ \sigma_{B_0B_1} & \sigma^2_{B_1} \end{bmatrix} \right)$$

- The fixed effects of the model are the intercept α and the slope β.

- These are the expected values of the intercepts and slopes for the population of children.

Mixed model parameterisation

- We have shown that

$$b_{i0} = \beta_0 + b^*_{i0}$$
$$b_{i1} = \beta_1 + b^*_{i1}$$

- We can therefore rewrite the model as:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b^*_{i0} + b^*_{i1} x_{ij} + e_{ij}$$

where

$$b^*_{i0} \overset{\text{iid}}{\sim} N(0, \sigma^2_{B_0}),$$
$$b^*_{i1} \overset{\text{iid}}{\sim} N(0, \sigma^2_{B_1})$$
$$\text{Cov}\left(b^*_{i0}, b^*_{i1}\right) = \sigma_{B_0 B_1}.$$

A preliminary mixed-effects model

- We can use a linear mixed model to explore these relationships further.
- The fixed effects $[\beta_0, \beta_1]'$ are the representative intercept and slope for the population.
- The random effects $\mathbf{b}_i = [b_{i0}, b_{i1}], i = 1, \ldots, 700$ are the deviations in intercept and slope associated with child i.
- The random effects vector, $\mathbf{b}$, consists of the 700 intercept effects followed by the 700 slope effects.

Fitting the model

```
mod1 <- lmer(HAZ ~ Age + (Age | Subject))
```

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 0.74707  0.8643
          Age         0.02457  0.1568   -0.20
 Residual             0.11358  0.3370
```

```
Fixed effects:
             Estimate Std. Error t value
(Intercept) -0.875850   0.033279  -26.32
age         -0.144476   0.006611  -21.86
```

Fixed Effects

```
Fixed effects:
             Estimate Std. Error t value
(Intercept) -0.875850   0.033279  -26.32
age         -0.144476   0.006611  -21.86
```

- The t-value for age has an absolute value $>> 2$.
- There is a significant relationship between age and HAZ.
- We obtain our estimates for the global parameters: intercept $\hat{\beta}_0 = -0.876$ and slope $\hat{\beta}_1 = -0.144$.

Random Effects

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Subject  (Intercept) 0.74707  0.8643
          Age         0.02457  0.1568   -0.20
 Residual             0.11358  0.3370
```

- The estimate for the variance of the intercept random effect is given by $\hat{\sigma^2_{B_0}} = 0.747$.
- The estimate for the variance of the slope random effect is given by $\hat{\sigma^2_{B_1}} = 0.025$.
- The estimate for the covariance between the intercept and slope random effects is given by $\hat{\sigma_{B_0 B_1}} = -0.20$.
- There appears to be a weak negative correlation between the intercept and slope random effects.

Uncorrelated random effects

- We can fit a model with uncorrelated random effects.

```
mod2 <- lmer(HAZ ~ Age + (1 | Subject) +
               (0 + Age | Subject))
```

```
Random effects:
 Groups    Name        Variance Std.Dev.
 Subject   (Intercept) 0.73453  0.8570
 Subject.1 age         0.02415  0.1554
 Residual              0.11372  0.3372
```

```
Fixed effects:
            Estimate Std. Error t value
(Intercept) -0.87600    0.03301  -26.54
age         -0.14425    0.00657  -21.96
```

Likelihood ratio test

- We can compare these two models using a likelihood ratio test.
- We are testing the hypothesis $\sigma_{B_0 B_1} = 0$: random effects for slope and intercept are uncorrelated.

```
anova(mod1,mod2)

mod2: HAZ ~ Age + (1 | Subject) + (0 + Age | Subject)
mod1: HAZ ~ Age + (Age | Subject)
     Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
mod2  5 10641 10677 -5315.7    10631
mod1  6 10621 10664 -5304.7    10609 22.018      1  2.701e-06 ***
```

- Here, $p << -0.05$ therefore we conclude that the correlation is necessary in the model.

Predicting child-specific equations

- We can predict the growth equations (known as growth curves) for each individual child using our model output.
- The model for each child takes the form

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b^*_{i0} + b^*_{i1} x_{ij} + e_{ij}$$

- We can therefore use the EBLUP to predict the curve for child i as

$$y_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{b}^*_{i0} + \hat{b}^*_{i1} x_{ij}$$

Predicting child-specific equations

- We already obtained estimates for the fixed effects, $\hat{\beta}_0$ and $\hat{\beta}_1$.
- We can obtain the random effect estimates, $\hat{b}^*_{i0}$ and $\hat{b}^*_{i1}$, as follows:

```
ranef(mod1)$Subject

     (Intercept)         age
1005  -0.3778223 0.05397033
1027   1.0658083 0.09923373
1038   0.2668162 0.19635009
1056  -1.0880302 0.02280638
1058  -1.6046882 0.16492661
1061   0.9039936 0.02793797
etc
```

Predicting child-specific equations

- We can now predict the growth curve for child `1005` as follows.

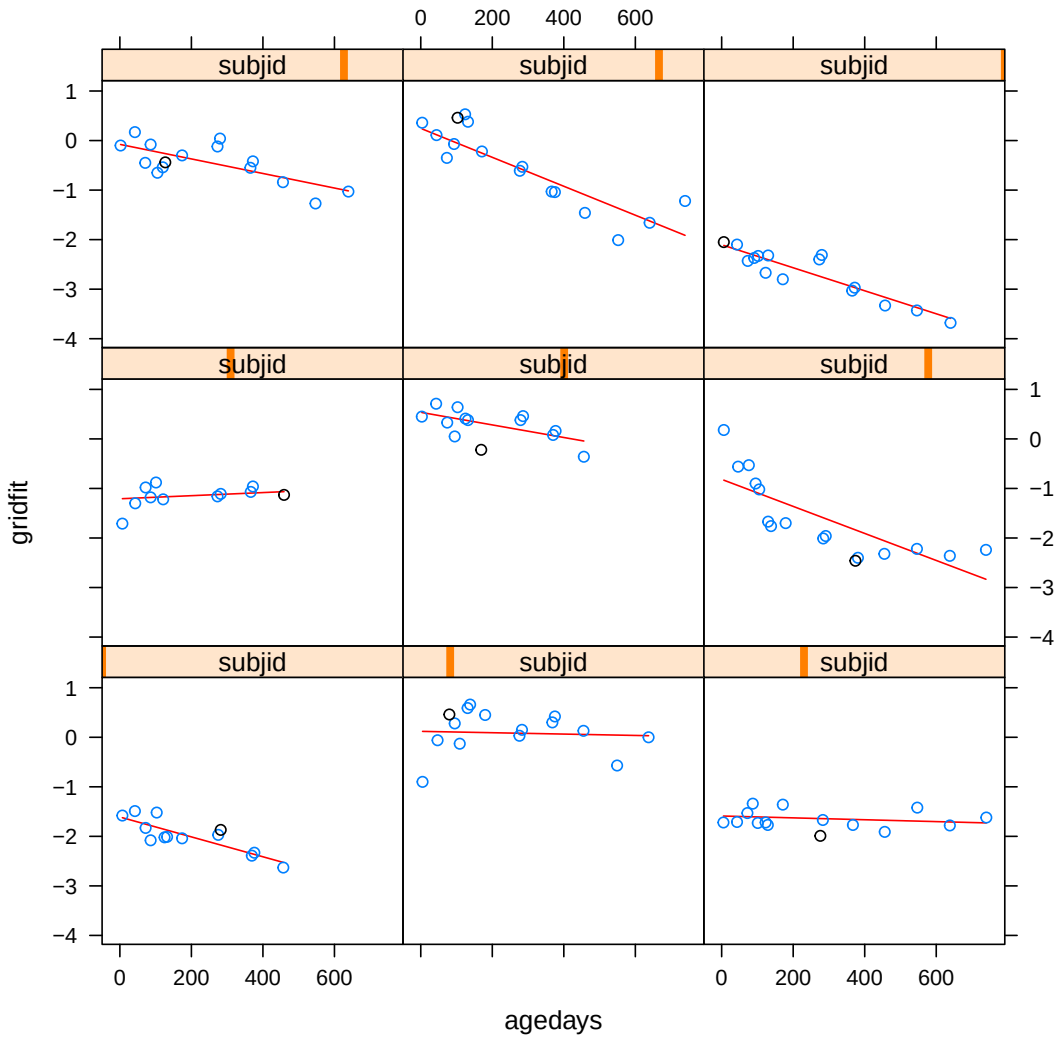
$$y_{1005,j} = \beta_0 + \beta_1 x_{1005,j} + b^*_{1005,0} + b^*_{1005,1} x_{ij}$$

$$= -0.876 + (-0.144)x_j + (-0.378) + 0.054x_j$$

$$= -1.254 - 0.090x_j$$

Fitted Data

- Fitted lines for 9 randomly selected children.
- Each child has a unique growth curve.
- Possibly need non-linear trends for some children.



Limitations of linear mixed models

- Normality assumptions limit general linear mixed models to continuous responses.
- Different methodology must be used when the responses are discrete and non-normal.
- Generalised estimating equations (GEE) are widely used as a way to deal with correlated discrete response data.
- For example, longitudinal data with response variables that are binary or discrete counts can be modelled using this method.

GEE regression models

- Generalised estimating equations (GEE) were developed to accommodate correlated observations within subjects.
- An estimating equation is the equation we solve to calculate the parameter estimates.
- The extra term *generalised* distinguishes GEE as the estimating equations that accommodate the correlation structure of the repeated measurements.

GEE regression models

- GEE models are useful in analysing data that arises from a longitudinal or clustered design.
- They are marginal models that model the effect of the explanatory variables on the population-averaged response.
- GEE models are recommended when our main goal is inference from the regression equation.

Review of Generalised Linear Models

- GEE models are an extension of generalised linear models (GLMs).
- A GLM relates the expected value of the response variable to the linear predictor through a link function:

$$g(E(Y_i)) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

- The variance of the response variable is a specified function of its mean.
- The distribution of the response variable comes from the exponential family of distributions.

Generalised Estimating Equations

GEE regression models extend GLMs by allowing:

1. the correlation of outcomes within an experimental unit to be estimated and taken into account when estimating the regression coefficients and their standard errors
2. the calculation of robust standard errors of the regression coefficients.

Variance-covariance matrix for GEE models

- In GEE regression models the variance-covariance matrix is a block-diagonal matrix.
- The observations within each block (blocks corresponding to subjects) are assumed to be correlated.
- The observations outside of the blocks are assumed to be independent.
- In other words, the subjects are still assumed to be independent of each other and the measurements within each subject are assumed to be correlated.

Quasi-likelihood estimation

- Maximum likelihood estimation requires the specification of the distribution of the response variable.
- For repeated discrete outcomes it may be difficult to specify the distribution.
- GEE regression models use the method of quasi-likelihood estimation.

Quasi-likelihood estimation

- This estimation method does not require the specification of the distribution of the response variable.
- It only requires specification of the relationships between the response mean and covariates and between the response mean and variance.
- The log-likelihood is not calculated for the GEE model.

Quasi-likelihood estimation

- Let $\mathbf{y}_i$ be a vector of random variables representing the responses on a given subject.
- Let $E(\mathbf{y}_i) = \boldsymbol{\mu}_i$, which is linked to the linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$ in some appropriate way.
- Let $\text{Var}\,(\mathbf{y}_i) = \text{Var}\,(\mathbf{y}_i; \boldsymbol{\beta}, \boldsymbol{\alpha})$ where $\boldsymbol{\alpha}$ represents parameters that model the correlation structure within subjects.

Quasi-likelihood estimation

- The parameters $\boldsymbol{\beta}$ are estimated by setting the multivariate score function to zero and solving

$$\sum_i \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)^{\mathsf{T}} \text{Var}\left(\mathbf{y}_i\right)^{-1} \left(\mathbf{y}_i - \boldsymbol{\mu}_i\right) = 0$$

  with a consistent estimate of $\boldsymbol{\alpha}$ substituted into $\text{Var}\left(\mathbf{y}\right)$.

- A similar set of equations can be derived with respect to $\boldsymbol{\alpha}$.

- These are called *generalised estimating equations*.

GEE fitting algorithm

GEE algorithm

1. Fit a generalised linear model assuming independence.
2. Compute the parameter estimates of the working correlation matrix based on the Pearson standardized residuals, the assumed correlation matrix structure and parameter estimates from the mean model.
3. Refit the regression model using an algorithm that incorporates the parameters from the working correlation matrix.
4. Keep alternating between steps 2 and 3 until model convergence is achieved.

Choice of working correlation structure

- The choice of correlation structure may be determined by the nature of the problem.
- If the number of observations is small in a balanced and complete design, unstructured is recommended.
- If repeated measurements are obtained over time, AR(1) or Toeplitz is recommended.
- If repeated measurements are not naturally ordered, compound symmetry is recommended.
- If the number of clusters is large and the number of measurements is small, an independent structure may suffice.

Misspecification of correlation structure

- What if the assumed correlation structure is wrong?
- If the estimation of the regression coefficients is the primary objective and there are a large number of clusters and a small number of time points, then choice of a correlation structure is not that important.
- The loss of efficiency from an incorrect choice of the working correlation structure is inconsequential when the number of subjects is large.

Missing values for GEE models

- Missing values that occur intermixed with non-missing values are called intermittent missing values.
- If these missing values are missing completely at random (MCAR), then the consistency results established by Liang and Zeger (1986) hold.
- MCAR means that missingness is completely independent of any study variable.

Notes on GEE models

- Correct specification of marginal mean and variance is required.
- Missing data cannot depend upon the observed or unobserved responses.
- A moderate to large number of independent subjects is required.
- We can fit them using the `gee` package in R.

Example

Radial keratotomy study

- Radial keratotomy is a form of surgery used to reduce myopia (nearsightedness).
- A longitudinal study of 362 adult myopic patients was conducted to evaluate the long-term (10-year) efficacy and stability of the surgery.
- After surgery, patients were examined at 6 months and then annually each year for 10 years.
- At each visit their refractive error was recorded.
- The concern of the scientists is that the refractive error would continue to change over time and the patients would become less and less nearsighted.

Data

Variables

- `patientid`: patient identification number
- `visit`: time of follow-up visit (1=1 year, 4= 4years, 10=10 years)
- `unstable`: the outcome variable coded as 1 if there is a continuing effect of the surgery and 0 otherwise
- `diameter`: diameter of the clear zone during the surgery (in mm)
- `age`: patient age at baseline in years
- `gender`: patient's gender

The dataset was provided by Azhar Nizam, Senior Associate, Rollins School of Public Health of Emory University, and is stored in the SAS dataset `keratotomy`.

R code for GEE model

- We can fit the model using the `gee` function in the `gee` package.
- R provides the initial regression parameter estimates (which are obtained from a glm).
- These initial parameters are used as are starting values for our GEE.

```
library(gee)

mod1 <- gee(unstable ~ age + diameter + gender + visit, id=patientid,
    family="binomial", corstr="unstructured", data=ker)

 running glm to get initial regression estimate
(Intercept)         age    diameter  genderMale       visit
 1.38439881  0.01049883 -1.19565213  0.56109816  0.31876954
```

GEE Output

- We obtain output using the `summary()` function.

```
summary(mod1)

Coefficients:
               Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  1.11055878 0.88613544  1.253261  0.92785741  1.196907
age          0.01247191 0.01170444  1.065571  0.01123404  1.110189
diameter    -1.13059505 0.21736827 -5.201288  0.22287872 -5.072692
genderMale   0.53868954 0.17345531  3.105639  0.17710700  3.041605
visit        0.32333578 0.02184779 14.799471  0.02255012 14.338540

Working Correlation
             [,1]      [,2]         [,3]
[1,]  1.00000000 0.2518344 -0.07471068
[2,]  0.25183436 1.0000000  0.23906732
[3,] -0.07471068 0.2390673  1.00000000
```

Output: Correlation

```
Working Correlation
             [,1]       [,2]         [,3]
[1,]  1.00000000 0.2518344 -0.07471068
[2,]  0.25183436 1.0000000  0.23906732
[3,] -0.07471068 0.2390673  1.00000000
```

- Since unstructured correlation structure is used, the correlations between time points are all estimated.
- There are a relatively large number of clusters and a relatively small number of time points.
- This means the choice of correlation structure will not substantially affect the results of the GEE model.

Output: SE Estimates

```
Coefficients:
               Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  1.11055878 0.88613544  1.253261  0.92785741  1.196907
age          0.01247191 0.01170444  1.065571  0.01123404  1.110189
diameter    -1.13059505 0.21736827 -5.201288  0.22287872 -5.072692
genderMale   0.53868954 0.17345531  3.105639  0.17710700  3.041605
visit        0.32333578 0.02184779 14.799471  0.02255012 14.338540
```

- We obtain estimates of our fixed effects.
- We also obtain both ‘naive’ and ‘robust’ standard error estimates and z-values.

Inference

- The naive estimate is the standard error under the assumption that the correlation matrix has been correctly specified and estimated.
- Using the robust estimate allows one to draw correct inferences from the data even if the correlation model was incorrectly specified.
- No p-values are given - the reported z-statistics can be treated as standard normal random variables and tests carried out in the usual fashion.
- The term `age` is not significant in the model, but `diameter`, `gender` and `visit` are.

Inference

- Recall that in logistic regression we interpret the parameter estimates in terms of **odds**.
- For every one unit increase in our covariate value, the odds are multiplied by a factor of $\exp(\beta)$.
- For diameter, we have $\beta = -1.13$, therefore the odds of a continuing effect of a surgery are multiplied by $\exp(-1.13) = 0.323$ for a 1mm increase in diameter.
- If we have a wider clear zone during surgery, we are much less likely to see a continuing effect.

Estimating probabilities

- We can also use our model output to estimate probabilities.
- We compute the probability as

$$P(Y = 1) = \frac{\exp(\lambda)}{1 + \exp(\lambda)}$$

where λ is the linear predictor from the model.

- Here, we compute λ from our fitted model as:

$$\lambda = 1.111 + 0.012 * \text{age} + (-1.131) * \text{diameter} + 0.539 * \text{sex} + 0.323 * \text{visit})$$

Estimating probabilities

- For example, suppose we were interested in the probability of a continuing effect of surgery after 10 years for a 49 year old male with a diameter 3mm.
- We compute our linear predictor as

$$\lambda = 1.111 + 0.012 * 49 + (-1.131) * 3 + 0.539 * 1 + 0.323 * 10) = 2.075$$

- Therefore our probability of a continuing effect is given by:

$$P(\text{unstable} = 1) = \frac{\exp(2.075)}{1 + \exp(2.075)} = 0.888$$

Example

Epilepsy

- Data from a clinical trial of 59 epilepsy sufferers.
- Patients were observed for eight weeks and the number of seizures was recorded. This acted as the baseline seizure count.
- The patients were randomized to treatment by the drug Progabide (31 patients) or the placebo (28 patients).
- They were then observed for four two-week periods and the number of seizures was recorded.
- The data are available in the dataset `epilepsy` available from `library(faraway)` in R.

The data

```
library(faraway)
library(gee)
epilepsy[1:10,]
   seizures id treat expind timeadj age
1        11  1     0      0       8  31
2         5  1     0      1       2  31
3         3  1     0      1       2  31
4         3  1     0      1       2  31
5         3  1     0      1       2  31
```

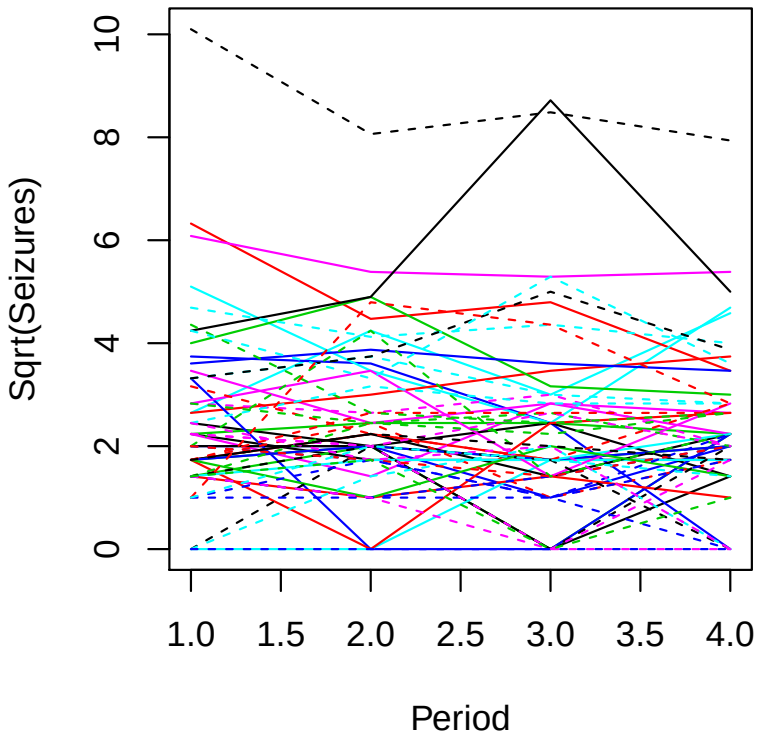
- The first two patients were both in the placebo group (`treat=0`).
- `expind=0` corresponds to the baseline period (with duration `timeadj`= 8 weeks.
- `expind=1` corresponds to the each of the four two-week observations periods (with duration `timeadj`= 2 weeks.

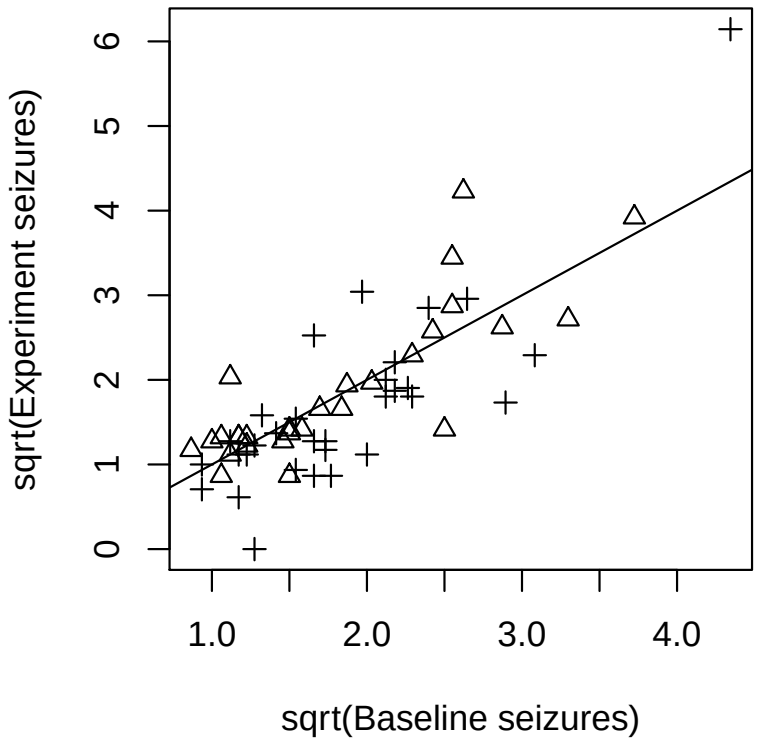
Seizures per week in epilepsy patients

	Baseline	Experiment
Placebo	3.85	4.30
Treatment	3.96	3.98

- The rate of seizures in the treatment group increases during the experiment.
- The rate increases even more for the placebo group.
- It is possible that, whatever is causing this increase, the drug is having a beneficial effect.

Exploratory plots





- Left: square root of seizures per two-week period (solid lines for treatment group).
- Right: mean numbers of seizures per week indicated with + for the treatment group.

Comments on the plots

- There doesn't appear to be much difference between the two groups in the left panel. If there is a treatment effect, it is not obvious.
- The square root transformation is used to stabilise the variance.

Fitting a GEE model

- We use a Poisson distribution with an offset due to the different lengths of the baseline and treatment periods (8 and 2 weeks respectively).
- Patient 49 has unusually high rate of seizures and is excluded from the analysis.
- An AR(1) correlation structure is assumed because measurements taken close to each other are expected to be more highly correlated than measurements taken farther apart from each other.

R code and output

```
> gmod <- gee(seizures ~ offset(log(timeadj))+expind+treat
            +I(expind*treat), family=poisson, corstr="AR-M", Mv=1,
            data=epilepsy, subset=(id!=49))
> summary(gmod)

Coefficients:
                  Estimate Naive S.E.   Naive z Robust S.E. Robust z
(Intercept)      1.32037722 0.1035447 12.7517565 0.1606548  8.2187244
expind           0.14277683 0.1393189  1.0248206 0.1076926  1.3257816
treat           -0.07940229 0.1468169 -0.5408251 0.1971622 -0.4027256
I(expind*treat) -0.37754557 0.2177363 -1.7339576 0.1683892 -2.2421009
```

Inference

- Here the term `I(expind*treat)` is significant under the robust SE.
- We interpret this as a signficant treatment effect that differs for the levels of `expind`.
- There is no difference between treatment and placebo group during the baseline observation period, but a significant difference during the experiment.

Working correlation matrix

```
Working Correlation
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.0000000 0.6185377 0.3825889 0.2366457 0.1463743
[2,] 0.6185377 1.0000000 0.6185377 0.3825889 0.2366457
[3,] 0.3825889 0.6185377 1.0000000 0.6185377 0.3825889
[4,] 0.2366457 0.3825889 0.6185377 1.0000000 0.6185377
[5,] 0.1463743 0.2366457 0.3825889 0.6185377 1.0000000
```

- Adjacent measurements have a correlation of 0.62.

Further Analysis

Further analysis of the seizures data would include

- investigating alternative correlation structures,
- including the age covariate, and
- investigating any trend during the experimental period.

Notes on Generalised Estimating Equations

- GEEs are *marginal* models: they extend GLMs by directly incorporating the within-subject association among the repeated measurements.
- Marginal models require assumptions about the **mean**, its **dependence on the covariates** and the **relationship between the mean and the variance** of the response.
- Assumptions are also made about the **correlations** between repeated observations of the same subject.
- In marginal models the mean response and the covariance are modelled separately, and therefore interpretation of the regression coefficients does not rely on the assumed correlation structure.

Generalised linear mixed models

- Generalised linear mixed models (GLMMs) are another approach for modelling correlated data with a discrete response.
- In a GLMM the linear predictor can contain random effects. (This is not the case in GEE models.)
- The random effects are assumed to be normally distributed.
- The conditional mean relates to the linear predictor through a link function:

$$g(E(\mathbf{y}|\mathbf{u})) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$$

- The conditional distribution (given $\mathbf{u}$) of the data belongs to the exponential family of distributions.

GLMM for a binary response

- Let Y_{ij} be a binary response, taking values 0 or 1.
- Given the random effect u_i, the Y_{ij} are independent Bernoulli random variables with

$$\text{Var}\,(Y_{ij}|u_i) = E(Y_{ij}|u_i)\{1 - E(Y_{ij}|u_i)\}.$$

- The conditional mean of Y_{ij} depends on fixed and random effects via the linear predictor

$$\eta_{ij} = \mathbf{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathsf{T}}\mathbf{u}_i = \mathbf{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} + u_i$$

where $z_{ij} = 1$ for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$.

GLMM for a binary response

- Then, given the logit link

$$\log\left\{\frac{\Pr(Y_{ij}=1|u_i)}{\Pr(Y_{ij}=0|u_i)}\right\} = \eta_{ij} = \mathbf{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} + u_i$$

- The single random effect u_i is assumed to have a normal distribution with mean zero and some variance.

- This example is a simple logistic regression with randomly varying intercepts.

GLMMs vs GEEs

GEE models

Address scientific questions about the impact of covariates on the changes in the mean response over time in the study **population**.

GLMMs

Address scientific questions about the impact of covariates on the changes in the mean response for any **individual**.

Example: Radial keratotomy (revisited)

Variables:

- `patientid`: patient identification number
- `visit`: time of follow-up visit (1=1 year, 4= 4 years, 10=10 years)
- `unstable`: the outcome variable coded as 1 if there is a continuing effect of the surgery and 0 otherwise
- `diameter`: diameter of the clear zone during the surgery (in mm)
- `age`: patient age at baseline in years
- `gender`: patient's gender

Fitting GLMMs in R

- To fit a generalised linear mixed model we can use the `glmer` function in the `lme4` package.
- This is an extension of the `lmer` function - the only difference is that we have to specify a family in addition to the formula and data arguments.
- For binary responses we use the binomial family for which the canonical link function is the logit.
- For count data we use the Poisson family for which the canonical link is the log.

R code for GLMM

```
mod1 <- glmer(unstable ~ age + diameter + gender +
                          visit + ( 1 | patientid ),
              data=ker, family="binomial")

Random effects:
 Groups    Name        Variance Std.Dev.
 patientid (Intercept) 1.31     1.14

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.52166    1.06995   1.422  0.15497
age          0.01449    0.01419   1.021  0.30731
diameter    -1.44060    0.26812  -5.373 7.75e-08 ***
genderMale   0.66822    0.21086   3.169  0.00153 **
visit        0.39948    0.03242  12.322  < 2e-16 ***
```

Output: Fixed Effects

```
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.52166    1.06995   1.422  0.15497
age          0.01449    0.01419   1.021  0.30731
diameter    -1.44060    0.26812  -5.373 7.75e-08 ***
genderMale   0.66822    0.21086   3.169  0.00153 **
visit        0.39948    0.03242  12.322  < 2e-16 ***
```

- Age is not significant, but everything else is.
- This agrees with our conclusions from the GEE, but note that some of the estimates and standard errors are different since it is a different model.

Output: Random Effects

```
Random effects:
 Groups    Name        Variance Std.Dev.
 patientid (Intercept) 1.31     1.145

 ranef(mod1)
$patientid
     (Intercept)
1    1.026738316
2    0.045266159
3    0.552361631
4    0.265157840
...
```

- For a standard mixed model, we can test the significance of the random effects by comparing the models with and without the random effect.
- However, by definition, a GLMM must have random effects, so we cannot take this approach here.

Inference

- Inference for binary GLMMs is similar to that for GEEs.
- Both models are based on logistic regression.
- We therefore interpret the parameter estimates in terms of **odds**.
- For every one unit increase in our covariate value, the odds are multiplied by a factor of $\exp(\beta)$.

Estimating probabilities

- Like for GEEs, we can also use our model output to estimate probabilities.
- We compute the probability as

$$P(Y = 1) = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})}$$

where η_{ij} is the linear predictor from the model.

- Unlike the GEE, these probabilities will be subject-specific since they will incorporate the random effect term.

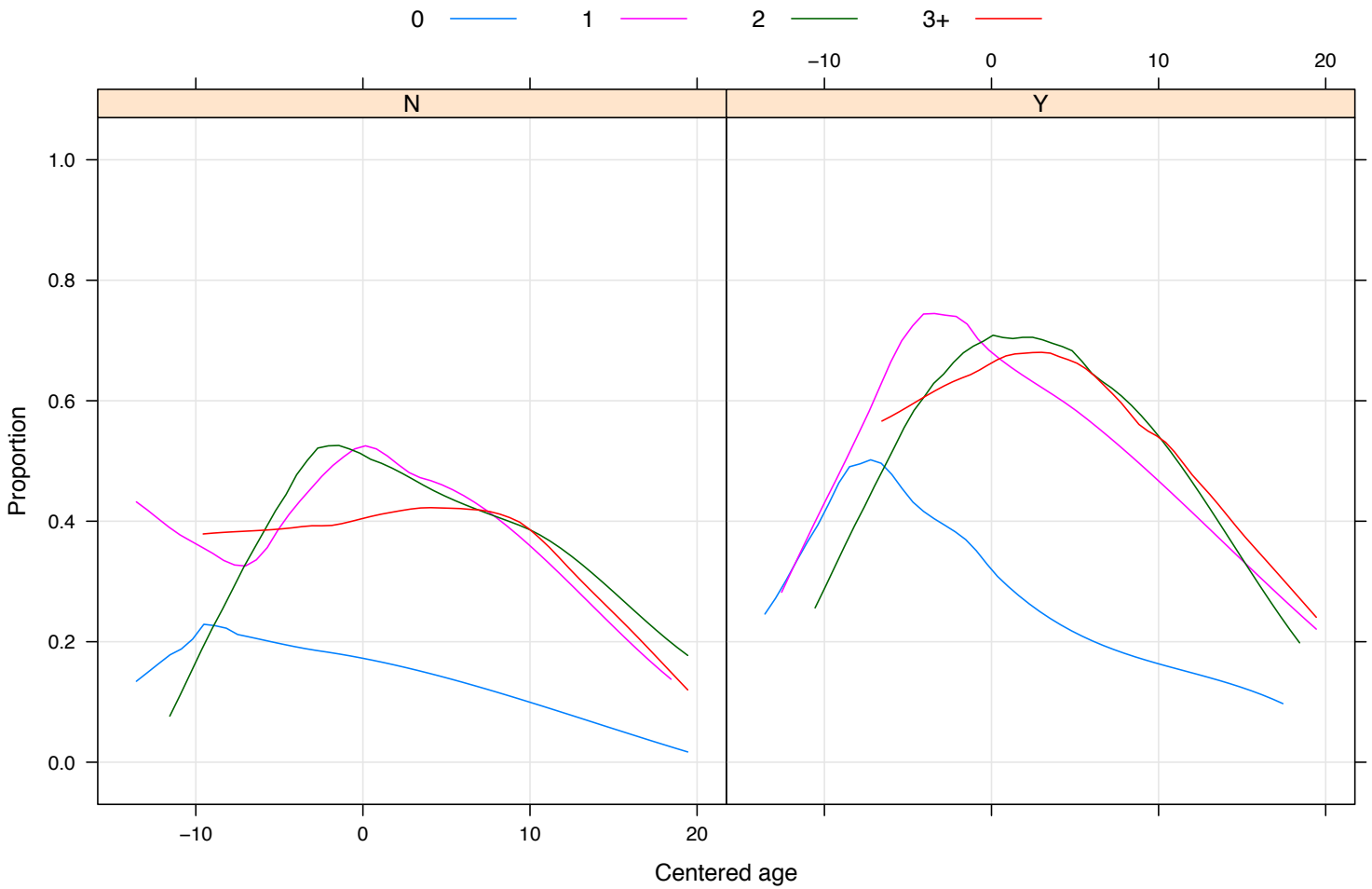
Example: Contraception

- One of the test data sets from the Center for Multilevel Modelling, University of Bristol is derived from the 1989 Bangladesh Fertility Survey (Huq and Cleland, 1990).
- The data are a subsample of responses from 1934 women grouped in 60 districts.
- The data are available in the dataset `Contraception` available from `library(mlmRev)` in R.

The data

```
library(mlmRev)
Contraception[1:10,]
   woman district use livch      age urban
1      1        1   N    3+  18.4400     Y
2      2        1   N     0  -5.5599     Y
3      3        1   N     2   1.4400     Y
4      4        1   N    3+   8.4400     Y
5      5        1   N     0 -13.5590     Y
6      6        1   N     0 -11.5600     Y
7      7        1   N    3+  18.4400     Y
8      8        1   N    3+  -3.5599     Y
9      9        1   N     1  -5.5599     Y
10    10        1   N    3+   1.4400     Y
```

Contraceptive use against age



Fitting the model in R

```
mod1 <- glmer(use ~ 1+age+I(age^2)+urban+livch+(1|district),
              data= Contraception, family= binomial)
summary(mod1)

Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.2258   0.4752
Number of obs: 1934, groups:  district, 60

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0350274  0.1757482  -5.889 3.88e-09 ***
age          0.0035352  0.0092784   0.381    0.703
I(age^2)    -0.0045621  0.0007294  -6.255 3.98e-10 ***
urbanY       0.6972851  0.1208606   5.769 7.96e-09 ***
livch1       0.8149767  0.1631910   4.994 5.91e-07 ***
livch2       0.9164595  0.1863436   4.918 8.74e-07 ***
livch3+      0.9150272  0.1873128   4.885 1.03e-06 ***
```

Further Comments

- The coefficients labelled `livch1`, `livch2` and `livch3+` are all large relative to their standard errors but they are not too different from each other.

- We could reparameterize the model in terms of whether the woman had children or not, using the following code:

```
Contraception <- within(Contraception,
ch <- factor(livch != 0, labels =
c("N", "Y")))
```

Model with binary variable for children

```
mod2 <- glmer(use ~ 1+age+I(age^2)+urban+ch+(1|district),
              data= Contraception, family= binomial)
summary(mod2)

Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.2247   0.474
Number of obs: 1934, groups:  district, 60

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0063737  0.1691107  -5.951 2.67e-09 ***
age          0.0062561  0.0078848   0.793    0.428
I(age^2)    -0.0046353  0.0007207  -6.432 1.26e-10 ***
urbanY       0.6929220  0.1206566   5.743 9.31e-09 ***
chY          0.8603821  0.1483014   5.802 6.57e-09 ***
```

Model with interaction term

```
mod3<- glmer(use ~ 1+age*ch+I(age^2)+urban+(1|district),
                 data= Contraception, family= binomial)
summary(mod3)

Random effects:
 Groups   Name        Variance Std.Dev.
 district (Intercept) 0.223    0.4723
Number of obs: 1934, groups:  district, 60

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.3232984  0.2150587  -6.153 7.59e-10 ***
age         -0.0472945  0.0218116  -2.168  0.03013 *
chY          1.2107566  0.2073394   5.839 5.24e-09 ***
I(age^2)    -0.0057569  0.0008405  -6.849 7.42e-12 ***
urbanY       0.7140073  0.1212596   5.888 3.90e-09 ***
age:chY      0.0683543  0.0254383   2.687  0.00721 **
```

Model comparisons

```
anova(mod2,mod3)
mod2: use ~ 1 + age + I(age^2) + urban + ch + (1 | district)
mod3: use ~ 1 + age * ch + I(age^2) + urban + (1 | district)
     Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
mod2  6 2385.2 2418.6 -1186.6   2373.2
mod3  7 2379.2 2418.2 -1182.6   2365.2 8.0045      1   0.004666 **
```

- The interaction term `age:ch` is useful.
- Contraceptive use does not peak at the same time for women with children and childless women.

Contraceptive use against age