3. Linear Mixed Effects (LME) Models

1. Overview
2. Simple cases:
   a) Random intercept
   b) Random intercept and slope.
3. The general LME model formulation.
4. Variance and covariance structures.
5. Estimation.
6. Model selection.
7. Residual Analysis and Goodness of Fit.
8. Implementation in R and Phoenix.

3.1 Overview

Overview
LME Models

- Not very common for pharmacokinetic data; however, they lead to a better grasp of the nonlinear mixed effects case.

- Mixed effects are introduced when measurements within subjects are not assumed independent.

- Data is longitudinal: multiple measurements are made on the same subject over time, e.g., blood pressure, drug concentration.

Overview

- Responses may be:
  - Unequally spaced (times between measurements may vary).
  - With an unequal number of observations per subject
  - Often correlated within a subject.
- Overall are more flexible than multiple regression models, since they:
  - Allow for greater control over the sources of variability
  - Incorporate patient-specific characteristics
  - Allow for covariates to vary over time

Mixed effects: Fixed + Random effects

Fixed effects

- Represent variables whose levels were chosen/ controlled, e.g., drug doses or time points for blood draws.

- Their levels represent a set of all possible levels, e.g., gender

Random effects

- Represent variables whose levels do not constitute the set of all possible levels (e.g., countries, in multinational level analysis).
- Often represent nuisance variables: effect is not of interest but rather the variability induced by the variable.
- Are arbitrary samples from a larger pool of other equally possible samples.
- Most common example: the subjects used in an experiment
  - There is often no specific interest in the particular set of subjects, but in generalizing the results to a population at-large.

Simple cases of random and mixed effects models

- Linear random effects model:
  - Random intercept
  - Random intercept and slope

- Linear mixed effects model (random + fixed):
  - Random intercept and slope by groups

Example Random Effects – No Covariates
Orthodontic Study on Maxillary Distance

(Pinheiro, 2000)

- Measurements of the distance from the pituitary gland to the pterygo-maxillary fissure (abbrev. “maxillary distance”).
- Data collected from x-rays of children’s skulls.
- Taken every two years
- From 8 -14 years of age
- Sample of 27 children- 16 males and 11 females.
- Available in R with name “Orthodont” (ISwR package)

Example Random Effects, Orthodontic Study on Maxillary Distance (Pihneiro, 2000)

Response variable:

$$Y_{ij} = \text{Maxillary distance (mm)},$$

where

i indexes subjects ($i=1,\ldots,n=27$),
j indexes occasions ($j=1,\ldots,m=4$)
representing 8,10,12 and 14 years of age.

Data structure
(4 subjects)

Subject i	Maxillary distance measurements (mm) Y_{ij} *i=subject, j=occasion*				Subject Mean $\bar{Y}_i$
	j=1 8 yrs	*j=2* 10 yrs	*j=3* 12 yrs	*j=4* 14 yrs	
1	26	25	29	31	27.75
2	21.5	22.5	23	26.5	23.38
3	23	22.5	24	27.5	24.25
4	25.5	27.5	26.5	27	26.63

Grand Mean $\bar{Y} = 24.02$
(27 subjects)

Note: The arrangement of the data in R consists of multiple lines per subject.

Random effects model with no covariates

$$Y_{ij} = \mu + u_i + \varepsilon_{ij}; \quad j = 1,\ldots,m_i,\ i = 1,\ldots,n.$$

$$= \mu_i + \varepsilon_{ij};$$

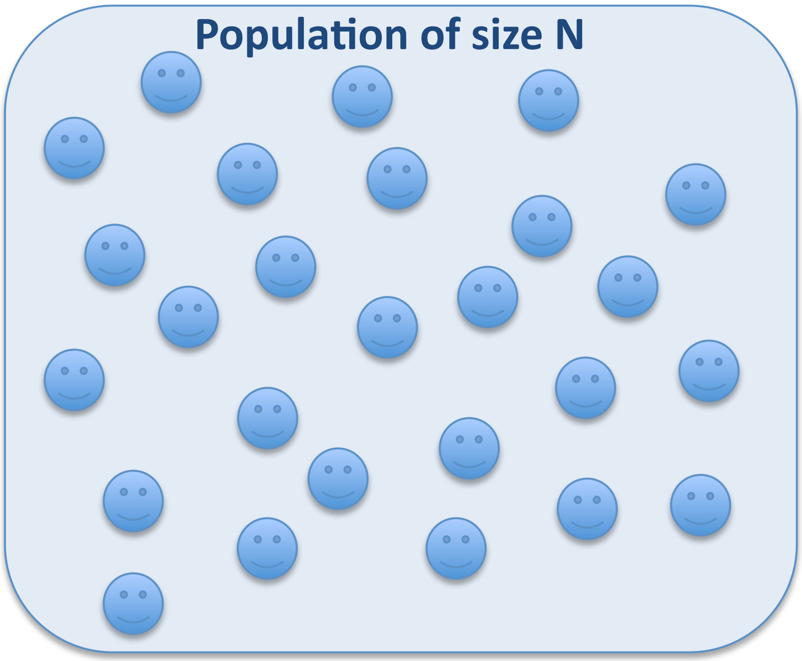
In terms of regression, think of μ as β_0.
and $\mu_i=\beta_0+u_i$ only here there's no line!

In our example,

- $n=4$, $m_i=4$ for all subjects $i=1,,\ldots,n$.
- Y_{ij} is the maxillary distance (mm) taken from subject i at occasion j.
- μ is the population overall mean maxillary distance.
- μ_i is the true (population) subject-specific mean.
- u_i is the random deviation of μ_i from μ
- ε_{ij} is the random deviation for j-th measurement on subject i from μ_i.

Random effects model with no covariates

Population: all children aged 8-14



Take subject i:

Y_{i1}
Y_{i2}
Y_{i3}
Y_{i4}

Set of **true** maxillary distances at ages 8,10,12,14 years.

(Unknown, not measured yet)

True (population) subject-specific mean:
(Unknown)

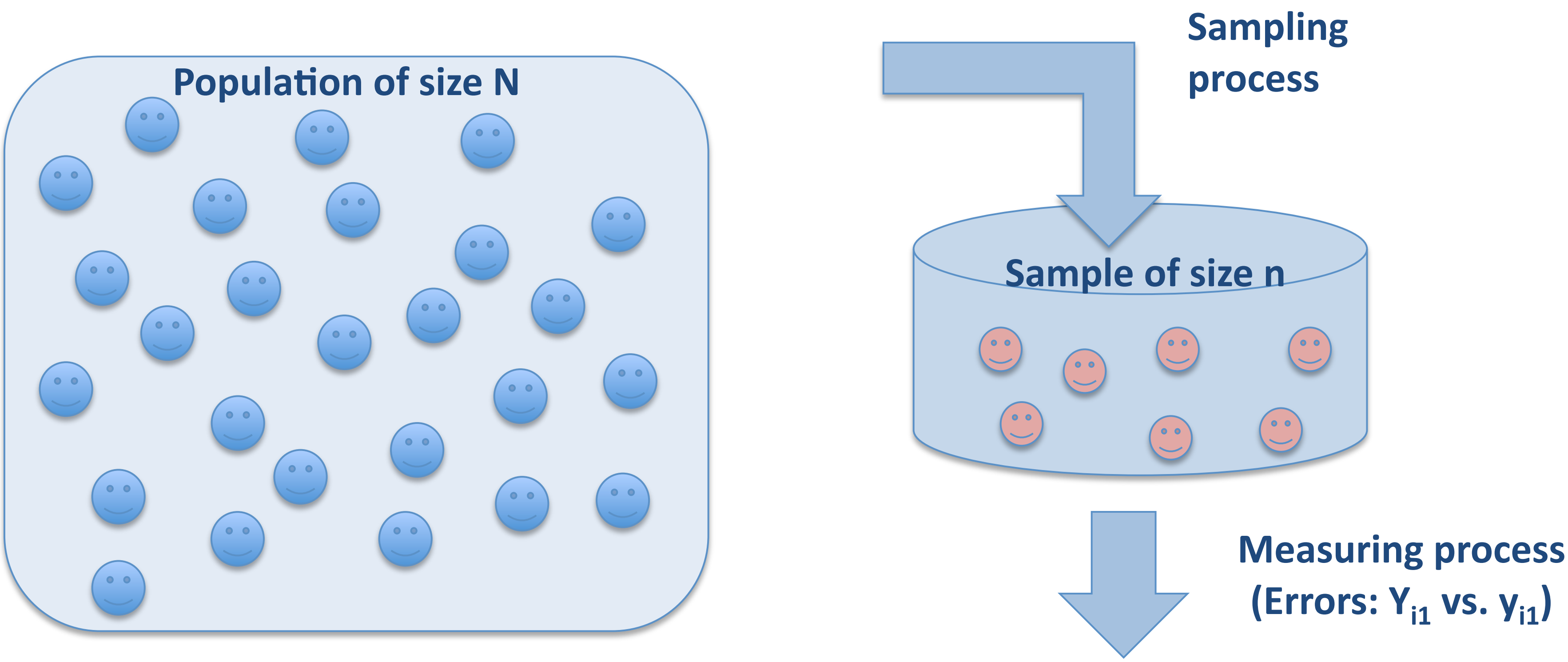
$$\mu_i = \frac{1}{4}\sum_{j=1}^{4} Y_{ij}$$

$$u_i = \mu - \mu_i$$

True population overall mean:
(Unknown)

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \mu_i$$

Random effects model with no covariates



Take subject i from the sample:

y_{i1}	Set of
y_{i2}	**observed**
y_{i3}	maxillary
y_{i4}	distances

Deviation of Y_{i1} from μ_i
\+ measurement error

$$
\begin{aligned}
\varepsilon_{i1} &= (Y_{i1} - \mu_i) + (y_{i1} - Y_{i1}) \\
&= y_{i1} - \mu_i
\end{aligned}
$$

Example Random Effects, Orthodontic Study on Maxillary Distance (Pinheiro, 2000)

Residuals: Deviations from subject means: $\hat{\varepsilon}_{ij} = y_{ij} - \hat{\mu}_i = y_{ij} - \bar{Y}_{i.}$

Subject i	Maxillary distance measurements (mm) Y_{ij} *i=subject, j=occasion*				Subject Mean	Deviation from overall mean
	j=1 8 yrs	*j=2* 10 yrs	*j=3* 12 yrs	*j=4* 14 yrs	$\bar{Y}_i$	$\hat{u}_i$ $(\bar{Y}_{i.} - \bar{Y})$
1	26	25	29	31	27.75	1.98
2	21.5	22.5	23	26.5	23.38	-0.64
3	23	22.5	24	27.5	24.25	0.23
4	25.5	27.5	26.5	27	26.63	2.61

Estimate of subject-specific mean μ_i

Grand Mean (n=27) $\bar{Y} = 24.02$

Estimate of population mean μ

$$\hat{\varepsilon}_{11} = 26 - 27.75 = -1.75$$

Subject i	Residuals			
	8 yrs	10 yrs	12 yrs	14 yrs
1	-1.75	-2.75	1.25	3.25
2	-1.88	-0.88	-0.38	3.13
3	-1.25	-1.75	-0.25	3.25
4	-1.13	0.88	-0.13	0.38

3.2 Simple cases
a) Random Intercept Model

- Model specification
- Marginal and conditional means
- E(Y), Var(Y)
- E(Y|u), Var(Y|u)

Specific learning objectives:

1. Write the random intercept model and its assumptions.
2. Explain the meaning of the random effects.
3. Derive the marginal E(Y), Var(Y) and conditional E(Y|u), Var(Y|u).
4. Explain the marginal and conditional means.
5. Fit a random intercept model in R and identify the estimated model parameters in the output.

Random Intercept Model

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

μ_{ij} is the subject-specific mean at occasion j.

$$= \underbrace{\beta_0 + \beta_1 t_{ij}}_{\text{Fixed}} + \underbrace{u_{i1} + \varepsilon_{ij}}_{\text{Random}}; \quad j = 1,\ldots,m_i; \quad i = 1,\ldots,n.$$

Re-arranging,

$$Y_{ij} = \underbrace{\left(\beta_0 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\beta_1 t_{ij}}_{\text{Slope}} + \varepsilon_{ij};$$

Random Intercept Model

Population subject specific mean: $\mu_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i1}$

Population overall mean: $\beta_0 + \beta_1 t_{ij}$

- t_{ij} is the time variable associated with measurement Y_{ij}
- u_{i1} is the random deviation of the true population subject specific mean intercept ($\beta_0 + u_{i1}$) and overall intercept β_0.
- This translates in the deviation of the true population subject specific mean μ_{ij} vs. the overall population mean $\beta_0 + \beta_1 t_{ij}$.
- ε_{ij} is the random deviation of the ij-th response from from subject specific mean.

Sources of Variability
Random Intercept Model

$$u_{i1} \sim N(0, \sigma_{u_1}{}^2), \qquad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon{}^2).$$

1. $\sigma_u{}^2$: Between subject variability (BSV): reflects the subject specific variability around the overall population mean.

2. $\sigma_\varepsilon{}^2$: Within subject variability (WSV): includes within subject variability and measurement error.
   - Represents variability that cannot be explained with available information.
   - It is impossible to determine if the deviation of Y_{ij} from the subject specific mean μ_{ij} is really measurement error or due to true random variability within a subject.

Model Assumptions

Random Intercept Model

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i1} + \varepsilon_{ij}; \quad j = 1,\ldots,m_i;\; i = 1,\ldots,n.$$

$$u_{i1} \sim N(0,\sigma_u^{\ 2}); \qquad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^{\ 2});$$

$$\varepsilon_{ij} \text{ independent of } \varepsilon_{ik}, \quad u_{i1} \text{ independent of } u_{j1}$$

$$u_{i1} \text{ independent of } \varepsilon_{ij}.$$

1. Linear relationship of Y with respect to parameters β_0, β_1.
2. Normally, independently distributed residual errors ε_{ij}.
3. Normally, independently distributed random effects u_{i1}.
4. Random effects and residuals errors are independent.

Random effects model with a time covariate
Random intercept.

Example: Maxillary Distance

$$Y_{ij} = \beta_0 + \beta_1 Age_{ij} + u_{i1} + \varepsilon_{ij}; \quad j = 1,\ldots,4, \ \ i = 1,\ldots,27.$$

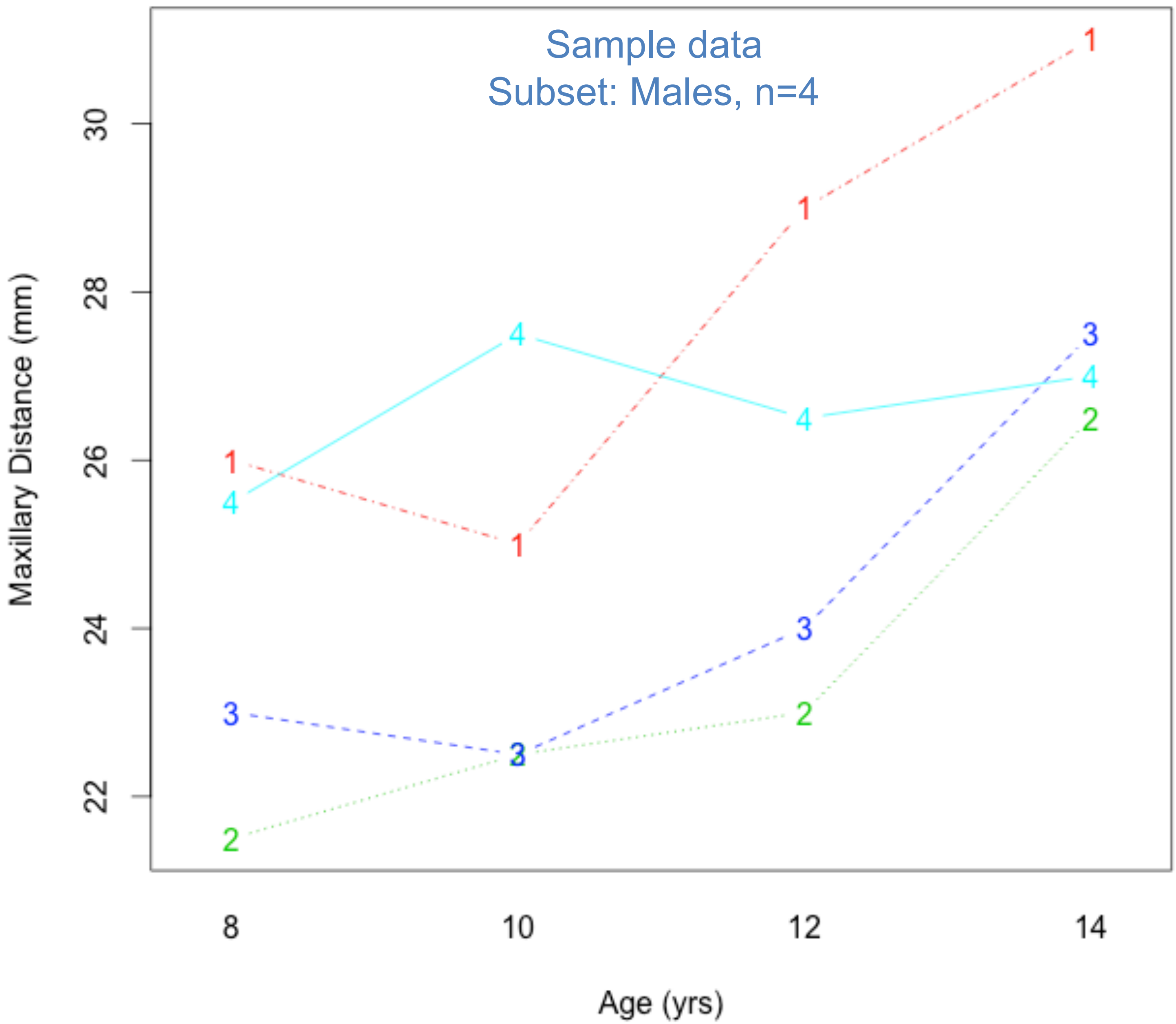
The subject-specific mean is:

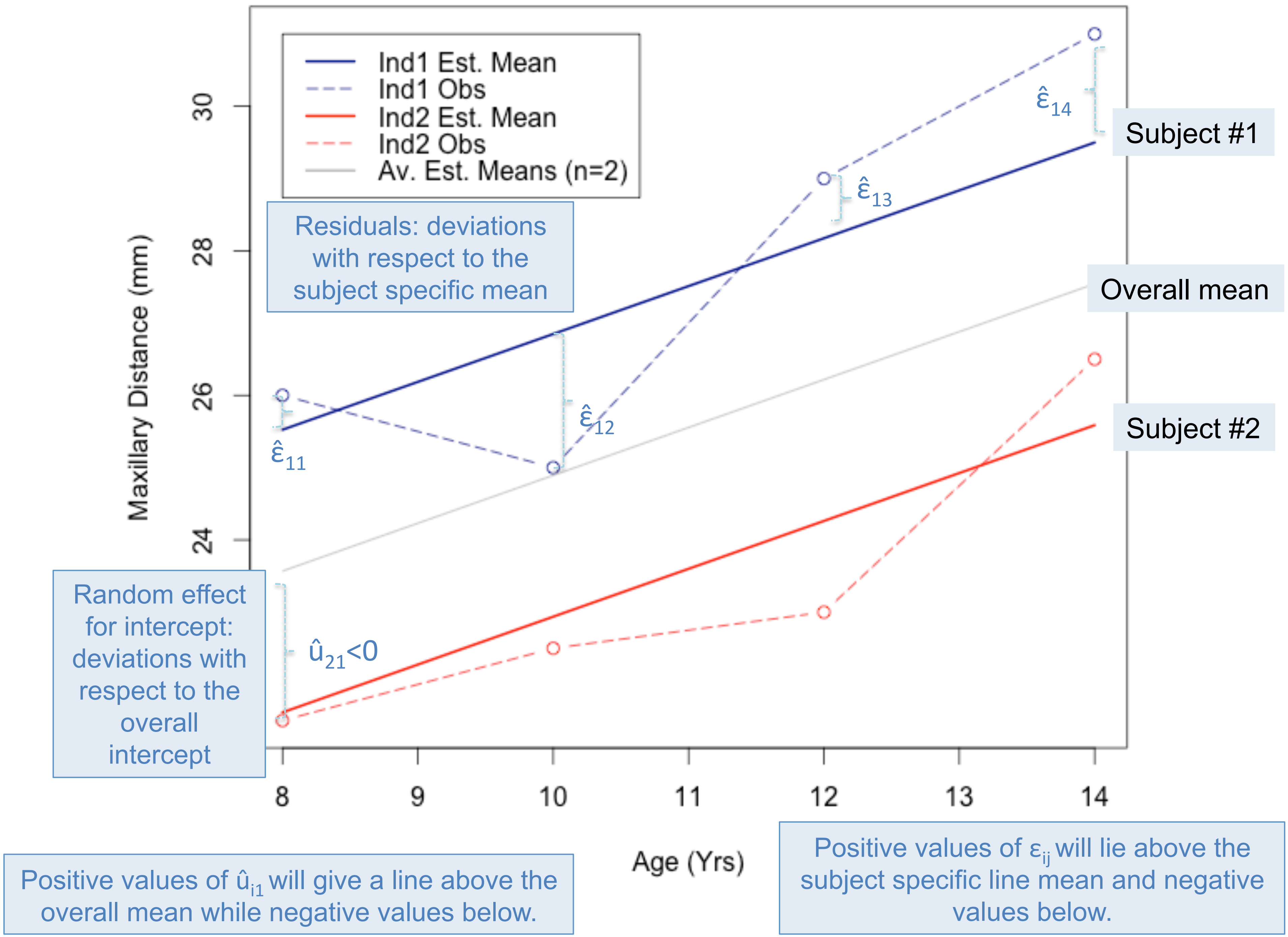
$$\mu_{ij} = \beta_0 + \beta_1 Age_{ij} + u_{i1}$$
$$= \left(\beta_0 + u_{i1}\right) + \beta_1 Age_{ij}$$

The subject specific mean is the regression line.

- Y_{ij} is the maxillary distance (mm) taken from subject i at occasion j.
- μ_{ij} is the true (population) subject-specific mean.
- (β_0+u_{i1}), β_1 are the intercept (random) and slope
- u_{i1} is the random deviation of the subject-specific intercept from the overall intercept β_0.
- ε_{ij} is the random deviation for j-th measurement on subject i from μ_i.

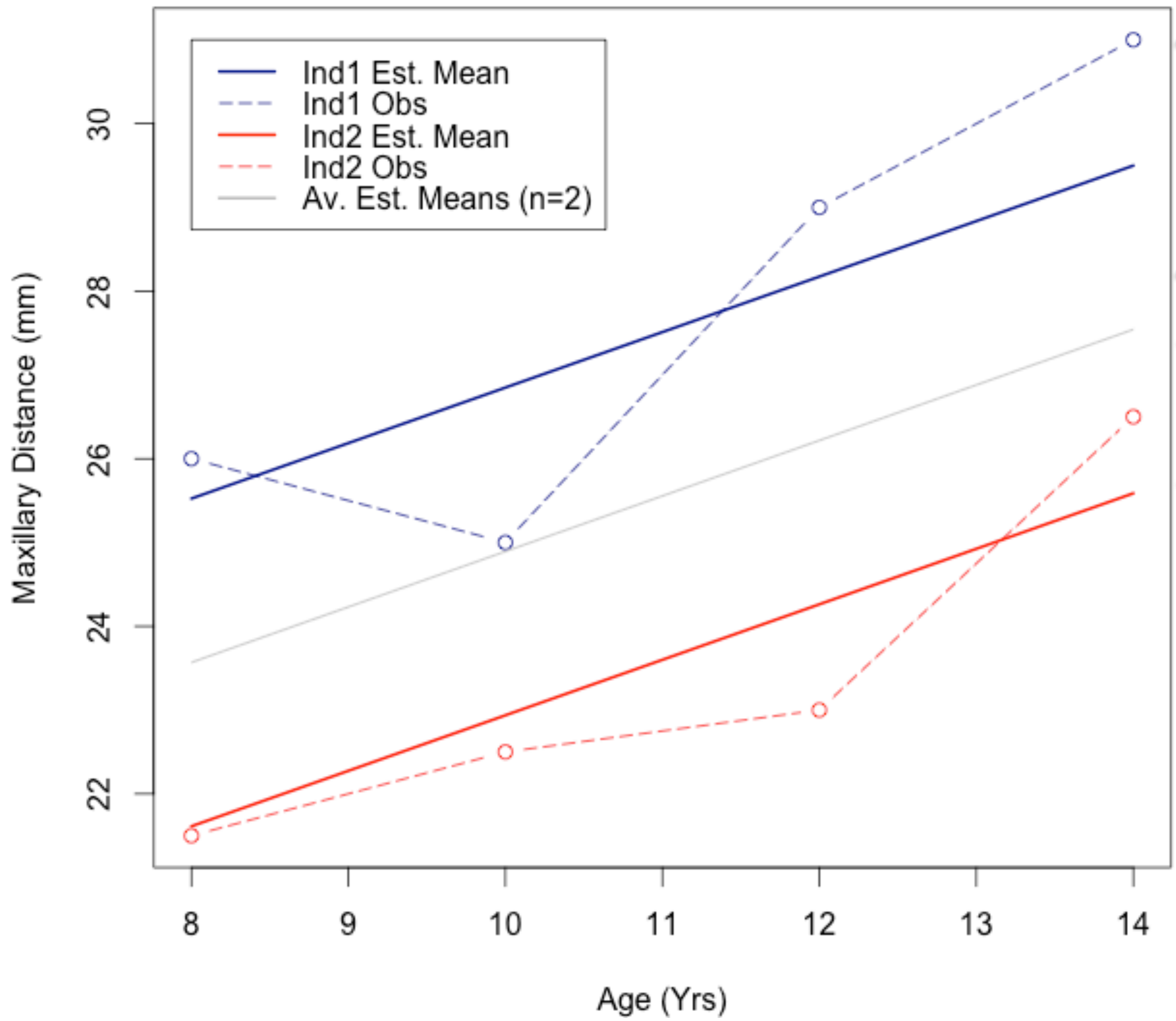
Example Random Effects, Orthodontic Study on Maxillary Distance (Pinheiro, 2000)





Ind1 Est. Mean
Ind1 Obs
Ind2 Est. Mean
Ind2 Obs
Av. Est. Means (n=2)
Residuals: deviations with respect to the subject specific mean
$\hat{\varepsilon}_{11}$
$\hat{\varepsilon}_{12}$
$\hat{\varepsilon}_{13}$
$\hat{\varepsilon}_{14}$
Subject #1
Overall mean
Subject #2
Random effect for intercept: deviations with respect to the overall intercept
$\hat{u}_{21}<0$
Maxillary Distance (mm)
24
26
28
30
8
9
10
11
12
13
14
Age (Yrs)
Positive values of $\hat{u}_{i1}$ will give a line above the overall mean while negative values below.
Positive values of ε_{ij} will lie above the subject specific line mean and negative values below.

Two random sources of variability



Within Subjects (WSV)
With respect to individual mean

Between Subjects (BSV)
With respect to the overall mean

Fitting a Random Intercept Model
Maxillary Distance Example

$$Y_{ij} = \beta_0 + \beta_1 Age^*_{ij} + u_{i1} + \varepsilon_{ij}\,.$$

$$u_{i1} \sim N(0,\sigma_u^{\,2}); \qquad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^{\,2});$$

$$\varepsilon_{ij} \text{ independent of } \varepsilon_{ik}, \quad u_{i1} \text{ independent of } \varepsilon_{ij}\,.$$

Age is centered: Age^*=Age-11

Typical call:

```
lme (fixed , data , random )
```

```
> library(nlme)
> dat <- Orthodont

> # random intercept
> fitt1 <- lme(distance ~ I(age-11),
                     data = dat, random = ~ 1|Subject)
```

```
> summary(fitt1)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  455.0025 465.6563 -223.5013

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    2.114724 1.431592

Fixed effects: distance ~ I(age - 11)
                Value Std.Error DF  t-value p-value
(Intercept) 24.023148 0.4296605 80 55.91193       0
I(age - 11)  0.660185 0.0616059 80 10.71626       0
 Correlation:
            (Intr)
I(age - 11) 0

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.66453932 -0.53507984 -0.01289591  0.48742859  3.72178465

Number of Observations: 108
Number of Groups: 27
```

$\hat{\sigma}_u, \hat{\sigma}_\varepsilon$

$\hat{\beta}_0, \hat{\beta}_1$

Number of subjects

Accessing fitted values

```
> fitt1$fitted
       fixed  Subject
1   22.04259 25.38635
2   23.36296 26.70672
3   24.68333 28.02709
4   26.00370 29.34746
5   22.04259 21.46107
6   23.36296 22.78144
7   24.68333 24.10181
8   26.00370 25.42218 . . .
```

Note that the first two subjects are male, have the same overall mean.

```
> fitt1$coef
$fixed
(Intercept) I(age - 11)
 24.0231481   0.6601852
```

$\boxed{\hat{\beta}_0, \hat{\beta}_1}$

```
$random
$random$Subject
    (Intercept)
M16  -0.9179756
M05  -0.9179756
M02  -0.5815230
M11  -0.3572213
. . .
M01   3.3437571
. . .
```

$\boxed{\hat{u}_{i1}\text{'s}}$

"fixed": $\boxed{\hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij}}$

```
> 24.0231481 +  0.6601852 *(c(8,10,12,14)-11)
[1] 22.04259 23.36296 24.68333 26.00370
```

"Subject": $\boxed{\hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij} + \hat{u}_{i1}}$

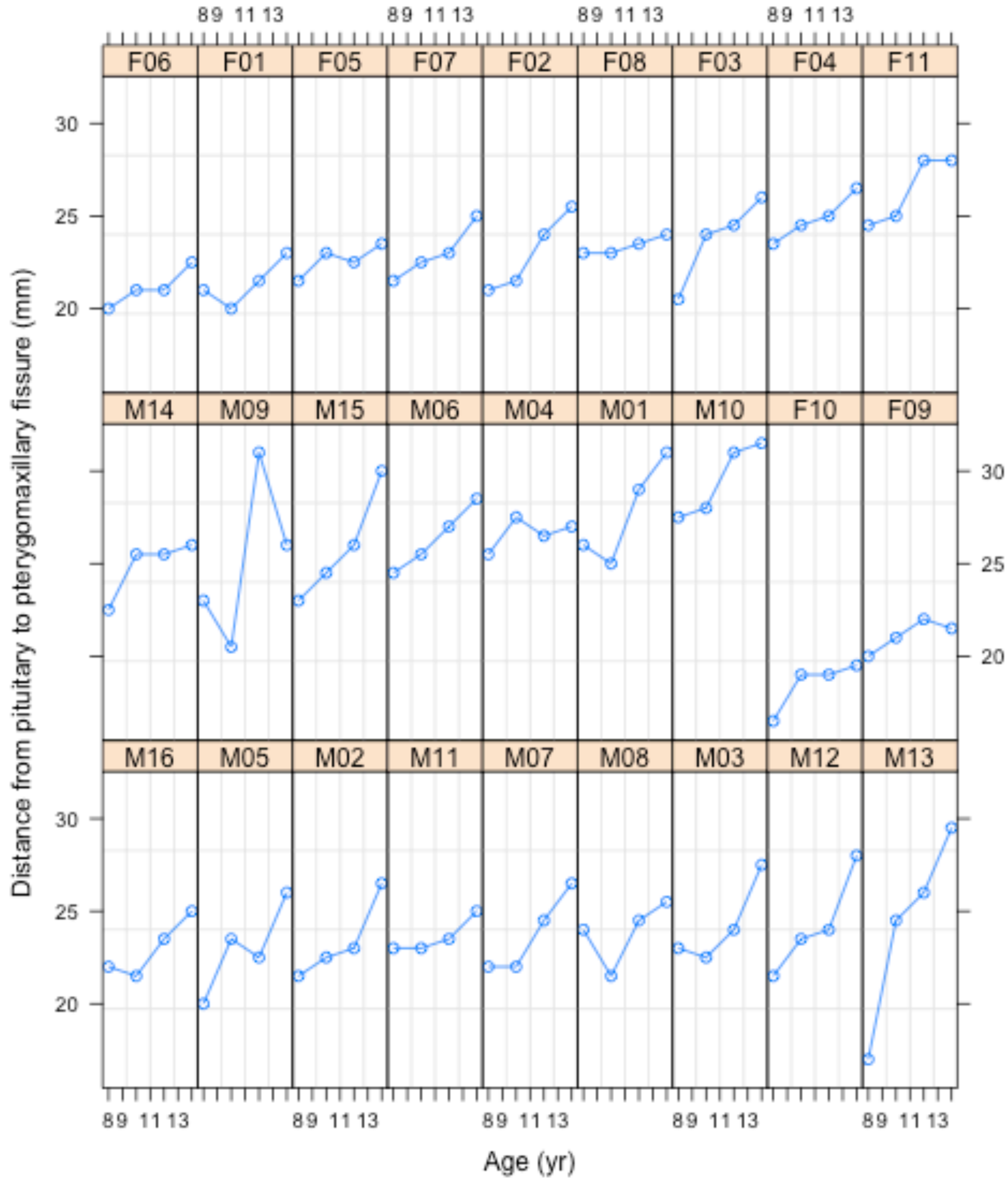
```
> 24.0231481 +  0.6601852 *(c(8,10,12,14)-11) + 3.3437571
[1] 25.38635 26.70672 28.02709 29.34746
```

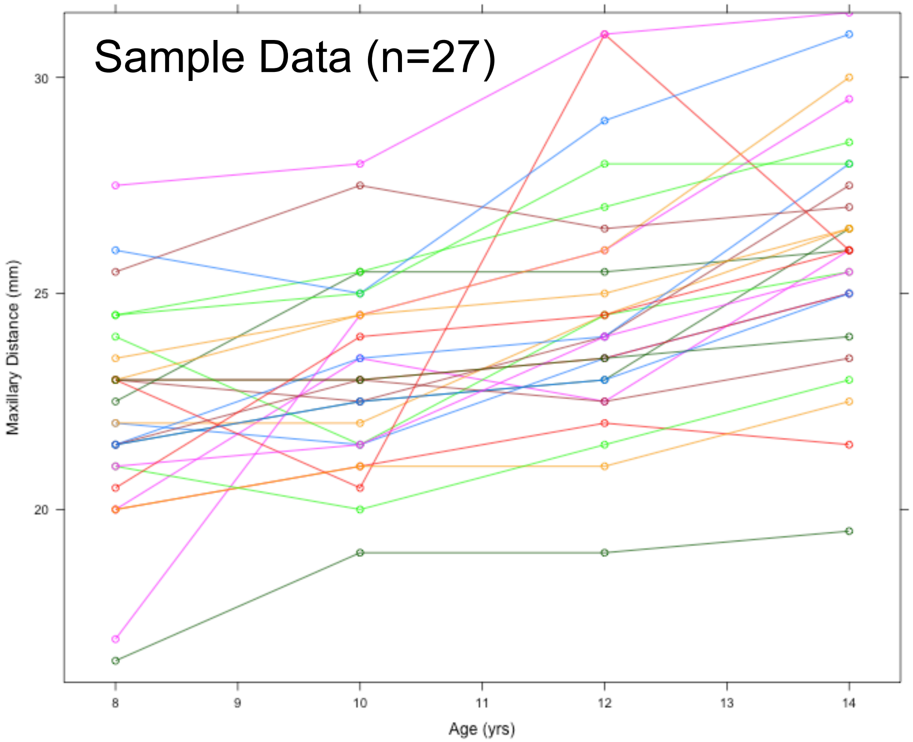
Accessing fitted values

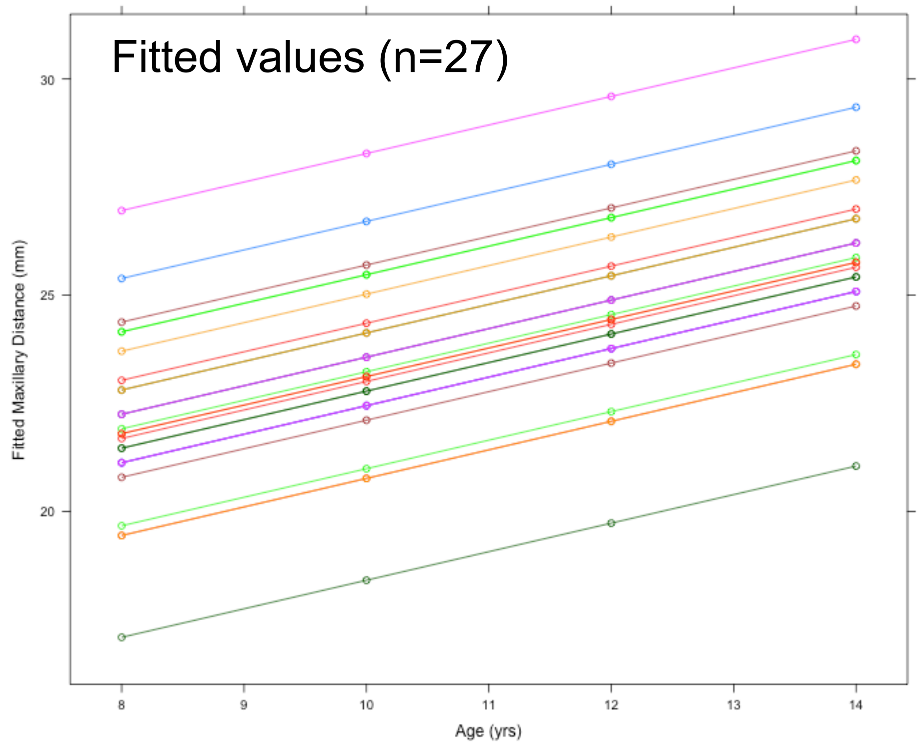
More convenient when plotting…

```
> fitted(fitt1)
M01      M01      M01      M01      M02      M02      M02      M02
25.38635 26.70672 28.02709 29.34746 21.46107 22.78144 24.10181 25.42218
     M03      M03      M03      M03      M04      M04      M04      M04
22.24613 23.56650 24.88687 26.20724 24.37699 25.69736 27.01773 28.33810
```

Individual Maxillary Growth (ages 8-14)







$$\hat{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij} + \hat{u}_{i1}$$

Marginal and Conditional Means

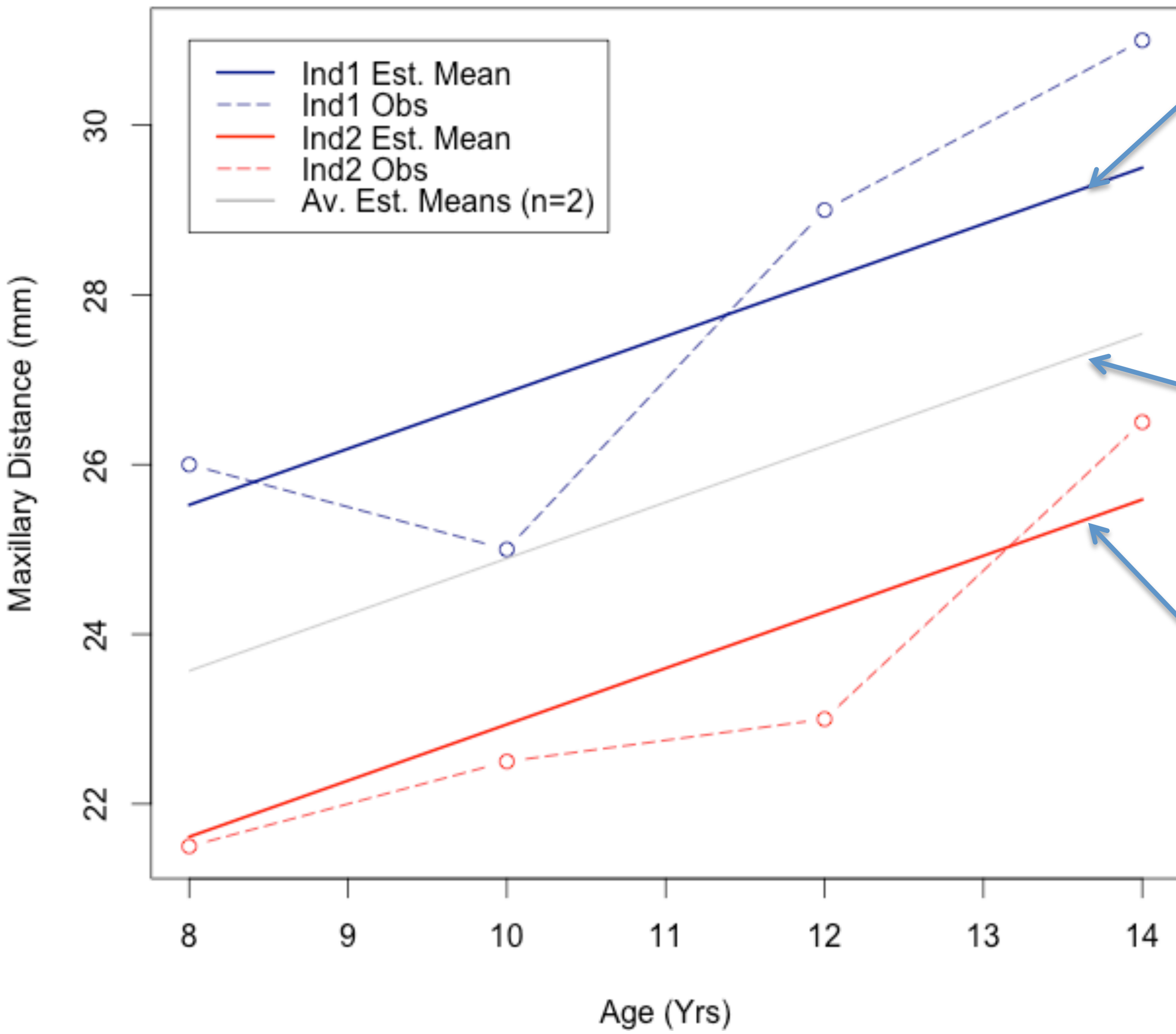
Marginal

Conditional
(u_{i1}'s as fixed)

$$Y_{ij} \sim \left(\mu_{ij}\ ,\ \sigma_u^{\ 2} + \sigma_\varepsilon^{\ 2}\right) \quad \text{and} \quad Y_{ij} \mid u_{i1} \sim \left(\mu_{ij} + u_{i1}\ ,\ \sigma_\varepsilon^{\ 2}\right)$$

- $E(Y)$: expected value of an individual randomly sampled from Y
- $E(Y|u_{i1})$: expected value for a particular individual
- If the expected value of individuals in their most general sense is of interest, then $E(Y)$ is of interest.
- But as soon as the discussion moves to particular subjects, $E(Y|u_{i1})$ becomes of interest.
- $Var(Y)$: total variability.
- $Var(Y|u_{i1})$: variability within a subject.

Two random sources of variability



Conditional mean, Subject 1

Marginal Mean ("averages" all u_i's)

Conditional mean, Subject 2

Marginal E(Y) and Var(Y)

Random intercept model

$$Y_{ij} = \beta_1 + \beta_2 t_{ij} + u_{i1} + \varepsilon_{ij}; \quad j = 1,,\ldots,m_i, i = 1,\ldots,n.$$

- The random errors ε induce variability into Y.

- The random effects u induce:
  - ✓ Variability into Y **AND**
  - ✓ Correlation within individual observations, i.e., $Cov(Y_{ij},Y_{ik})>0$.

How can we see this?

Marginal E(Y) and Var(Y)
Random intercept model

$$E(Y_{ij}) = E(\beta_1 + \beta_2 t_{ij} + u_{i1} + \varepsilon_{ij})$$
$$= \beta_1 + \beta_2 t_{ij} + E(u_{i1}) + E(\varepsilon_{ij})$$
$$= \beta_1 + \beta_2 t_{ij}.$$

Recall E(constant)=constant

The random effects induce variability into Y:

$$Var(Y_{ij}) = Var(\beta_1 + \beta_2 t_{ij} + u_{i1} + \varepsilon_{ij})$$
$$= Var(u_{i1} + \varepsilon_{ij})$$
$$= Var(u_{i1}) + Var(\varepsilon_{ij}) + 2Cov(u_{i1}, \varepsilon_{ij})$$
$$= \sigma_u{}^2 + \sigma_\varepsilon{}^2.$$

Recall Var(constant)=0

0

Independence assumptions u_1's vs. ε's, and within ε's

Marginal E(Y) and Var(Y)
Random intercept model

The random effects induce correlation into subject-specific Y's:

$$
\begin{aligned}
Cov(Y_{ij},Y_{ik}) &= Cov(\beta_1 + \beta_2 t_{ij} + u_{i1} + \varepsilon_{ij}, \beta_1 + \beta_2 t_{ik} + u_{i1} + \varepsilon_{ik}) \\
&= Cov(u_{i1} + \varepsilon_{ij}, u_{i1} + \varepsilon_{ik}) \\
&= Var(u_{i1}) + \underbrace{Cov(u_{i1},\varepsilon_{ik})}_{0} + \underbrace{Cov(\varepsilon_{ij},u_{i1})}_{0} + \underbrace{Cov(\varepsilon_{ij},\varepsilon_{ik})}_{0} \\
&= \sigma_u^{\,2}.
\end{aligned}
$$

0's due to independence assumptions u_1's vs. ε's, and within ε's

$$
Corr(Y_{ij},Y_{ik}) = \frac{Cov(Y_{ij},Y_{ik})}{\sqrt{Var(Y_{ij})}\sqrt{Var(Y_{ik})}} = \frac{\sigma_u^{\,2}}{\sigma_u^{\,2} + \sigma_\varepsilon^{\,2}}.
$$

Marginal E(Y) and Var(Y)
Random intercept model

In summary E(Y) and Var(Y) are:

$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}$$

$$Var(Y_{ij}) = \sigma_u^{\ 2} + \sigma_\varepsilon^{\ 2}.$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_u^{\ 2} \neq 0.$$

$$Corr(Y_{ij}, Y_{ik}) = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij})}\sqrt{Var(Y_{ik})}} = \frac{\sigma_u^{\ 2}}{\sigma_u^{\ 2} + \sigma_\varepsilon^{\ 2}}.$$

Note non-zero covariance implies non-zero correlation, this means that the LME model accounts for within-individual correlation

Conditional E(Y|u) and Var(Y|u)

Random Intercept Model

$$Y_{ij} = \mu_{ij} + u_{i1} + \varepsilon_{ij}; \quad j = 1,\ldots,m_i, \quad i = 1,\ldots,n$$

$$u_{i1} \sim (0, \sigma_u^{\ 2}) \qquad \varepsilon_{\text{ij}} \sim (0, \sigma_\varepsilon^{\ 2})$$

- When u_{i1} is stated as "given" (i.e., "$|u_{i1}$"), it can be considered fixed, then Y's ONLY inherit the distributional properties of the random component ε_{ij}.

$$E(Y_{ij} \mid u_{i1}) = \beta_0 + \beta_1 t_{ij} + u_{i1}$$

$$Var(Y_{ij} \mid u_{i1}) = \sigma_\varepsilon^{\ 2}.$$

$$\text{So} \qquad Y_{ij} \mid u_{i1} \sim \left(\beta_0 + \beta_1 t_{ij}\ ,\ \sigma_\varepsilon^{\ 2}\right)$$

3.2 Simple cases

b) Random Intercept & Slope:

- Model specification
- Variance & Covariance of Y

Specific learning objectives:

1. Write the random intercept & slope model and its assumptions.
2. State the features of the LME models that result from the random intercept & slope.
3. Fit a random intercept & slope model in R and identify the estimated model parameters in the output.
4. Construct graphs in R for longitudinal data.

Random Intercept and Slope Model
A Mixed Effects Model

+ random
effect for slope

$$Y_{ij} = \overbrace{\beta_0 + \beta_1 t_{ij}}^{\text{Fixed}} + \overbrace{u_{i1} + \boxed{u_{i2}} t_{ij}}^{\text{Random}} + \varepsilon_{ij}; \quad j = 1,2,3,4, \quad i = 1,\ldots,26.$$

$$Y_{ij} = \mu_i + \varepsilon_{ij};$$

$$\text{where } \mu_i = \beta_0 + \beta_1 t_{ij} + u_{i1} + u_{i2} t_{ij}.$$

Rearranging,

$$Y_{ij} = \underbrace{\left(\beta_0 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\left(\beta_1 + u_{i2}\right)}_{\text{Slope}} t_{ij} + \varepsilon_{ij};$$

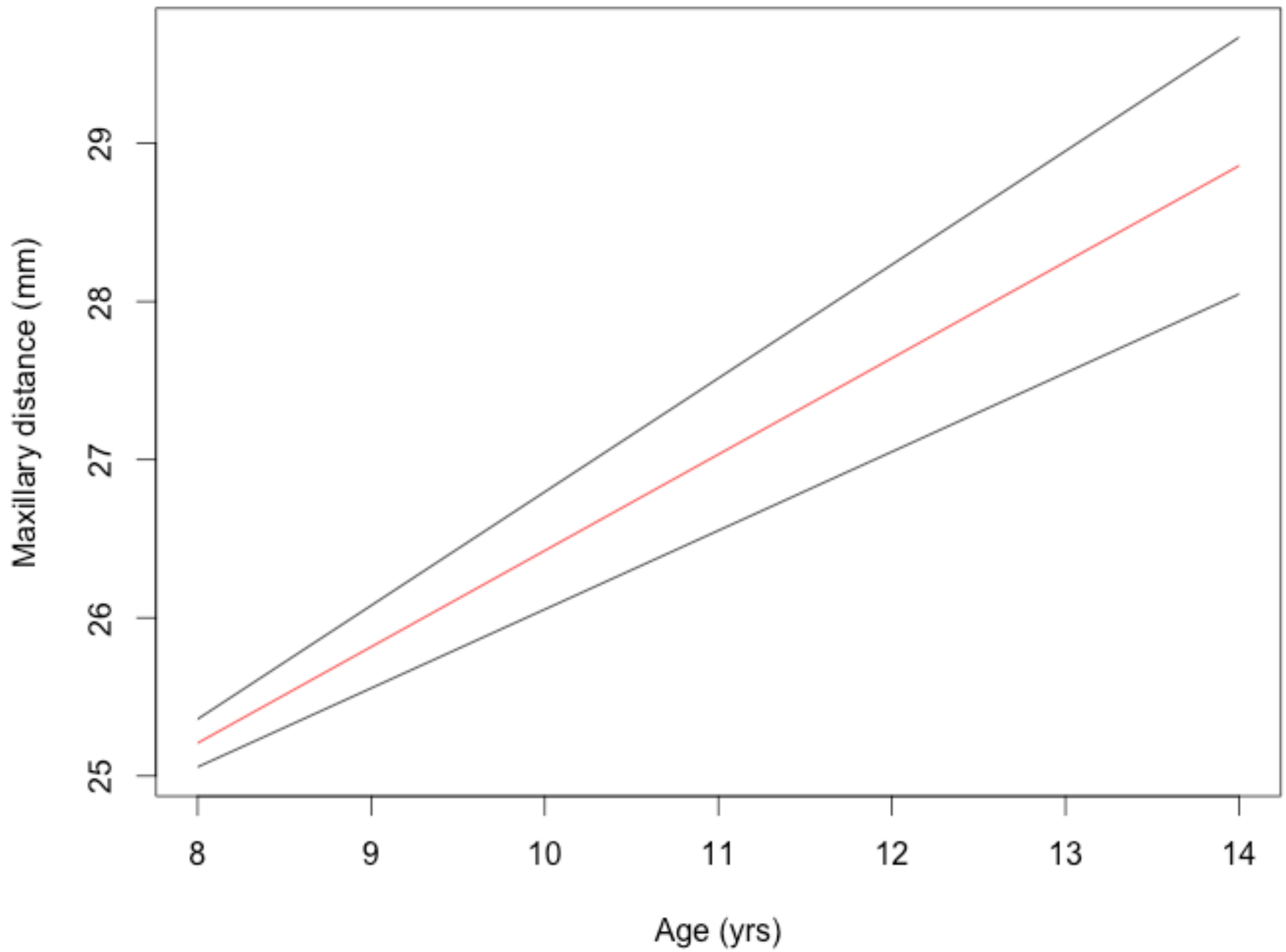
This is a mixed effects model because we are allowing a fixed effect (t_{ij}) to vary randomly.

Random Intercept and Slope Model

- t_{ij} is the time variable associated with measurement Y_{ij}
- u_{i1} is the random deviation of true population subject specific intercept and overall population intercept β_0.
- u_{i2} is the random deviation of true population subject specific slope and overall population slope β_1.
- ε_{ij} is the random deviation of the ij-th response from from subject specific mean.

$$Y_{ij} = \underbrace{\left(\beta_0 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\left(\beta_1 + u_{i2}\right)}_{\text{Slope}} t_{ij} + \varepsilon_{ij};$$

Random Intercept and Slope Model



Subject #1 has a higher intercept (baseline level) β_0+u_{11} than the population average β_0 and thus $u_{11}>0$.

Subject #2 has a lower intercept and thus $u_{21}<0$.

Subject #1 has a steeper rate of increase over time β_1+u_{21}than the population average β_1.

Subject #2 has a less steep rate of increase over time than the population average, $u_{22}<0$.

Fitzmaurice , Laird & Ware, 2nd Ed. 2011.

Random Intercept and Slope Model

Model Assumptions

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i1} + u_{i1} t_{ij} + \varepsilon_{ij}; \quad j = 1,\ldots,m_i;\ i = 1,\ldots,n.$$

$$u_{i1} \sim N(0, g_{11}); \qquad u_{i2} \sim N(0, g_{22});$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^{\ 2});$$

$$\varepsilon_{ij} \text{ independent of } \varepsilon_{ik}, \quad u_{i1} \text{ independent of } u_{j1}$$

$$u_{i2} \text{ independent of } u_{j2}, \quad Cov(u_{i1}, u_{i2}) = g_{12} \neq 0$$

$$u_{i1} \text{ independent of } \varepsilon_{ij}.$$

1. Linear relationship of Y with respect to parameters β_0, β_1.
2. Normally, independently distributed residual errors ε_{ij}.
3. Normally, independently distributed random effects u_{i1} and u_{i2}.
4. Correlated random effects u_{i1} and u_{i2}.
5. Random effects and residuals errors are independent.

Random Intercept and Slope Model

Variance and correlation of Y

Recall the variance and covariance of the random effects are denoted as follows:

- $Var(u_{i1}) = g_{11}$
- $Var(u_{i2}) = g_{12}$
- $Cov(u_{i1}, u_{i2}) = g_{12}$

It can be shown that (see Fitzmaurice, Ch.8):

$$Var(Y_{ij}) = Var(\beta_0 + \beta_1 t_{ij} + u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij})$$

$$= g_{11} + 2t_{ij} g_{12} + t_{ij}^{\ 2} g_{22} + \sigma_{\varepsilon}^{\ 2}.$$

Var(Y) = between variance + within variance

Var(Y) increases when g12>0
Var(Y) decreases when g12<0

$$Cov(Y_{ij}, Y_{ik}) = g_{11} + (t_{ij} + t_{ik}) g_{12} + t_{ij} t_{ik} g_{22}.$$

Cov(Y when age=2, Y when age=4)
≠
Cov(Y when age=2, Y when age=8)

Features of LME models

1. Unlike other models, LME models explicitly distinguish between subject-specific and within-subject sources of variability.
2. Covariance for Y_{ij}, Y_{ik} can be expressed as a function of time, therefore, the time spacing between measurements doesn't have to be uniform between subjects.
3. Magnitude of covariance between a pair of responses Y_{ij}, Y_{ik} depends on the time separation between them.
4. Variance of Y_{ij} increases over time when cov(u_{i1}, u_{i2}) ≥0 but decreases when cov(u_{i1}, u_{i2}) <0. Allows for heteroscedasticity (or non-constant variance).
5. Do not require a balanced longitudinal design: since Cov(Y_{ij}, Y_{ik}) is expressed as an explicit function of t_{ij}'s, each subject can have a unique sequence of measurement times.

Fitzmaurice, Laird & Ware, 2nd Ed. 2011

Fitting a Random Intercept and Slope Model

$$Y_{ij} = \beta_0 + \beta_1 Age^*{}_{ij} + u_{i1} + u_{i2} Age^*{}_{ij} + \varepsilon_{ij}.$$

$$u_{i1} \sim N(0, \sigma_{u_1}{}^2); \quad u_{i2} \sim N(0, \sigma_{u_2}{}^2); \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon{}^2);$$

u_{i1} not independent of u_{i2}, ε_{ij} independent of ε_{ik},

u_{i1} independent of ε_{ij},

```
# library(nlme)


# random intercept and slope
fitt2 <- lme(distance ~ I(age-11),
                      data = dat, random = ~ I(age-11)|Subject)
```

```
> summary(fitt2)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  454.6367 470.6173 -221.3183

Random effects:
 Formula: ~I(age - 11) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 2.1343289 (Intr)
I(age - 11) 0.2264278 0.503
Residual    1.3100402

Fixed effects: distance ~ I(age - 11)
                Value Std.Error DF  t-value p-value
(Intercept) 24.023148 0.4296601 80 55.91198       0
I(age - 11)  0.660185 0.0712533 80  9.26533       0
 Correlation:
            (Intr)
I(age - 11) 0.294

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-3.223106888 -0.493760896  0.007316481  0.472151221  3.916031742
Number of Observations: 108
Number of Groups: 27
```

$\hat{\sigma}_{u_1}, \hat{\sigma}_{u_2}, \hat{\sigma}_{\varepsilon}$

$Corr(u_{i1}, u_{i2})$

$\hat{\beta}_0, \hat{\beta}_1$

Number of subjects

Accessing fitted values

```
> fitt2$fitted
      fixed  Subject
1  22.04259 24.81965
2  23.36296 26.57139
3  24.68333 28.32313
4  26.00370 30.07487
```

```
> fitt2$coef
$fixed
(Intercept) I(age - 11)
 24.0231481   0.6601852
```

$\boxed{\hat{\beta}_0, \hat{\beta}_1}$

```
$random
$random$Subject
    (Intercept) I(age - 11)
M16  -0.9451479 -0.06885385
M05  -0.8950636  0.02560005
M02  -0.5679151  0.01450765
. . .
M01   3.4241123 0.21568501. . .
```

$\boxed{\hat{u}_{i1}, \hat{u}_{i2}'s}$

$$\boxed{\hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij}}$$

```
> 24.0231481 +  0.6601852 *(c(8,10,12,14)-11)
[1] 22.04259 23.36296 24.68333 26.00370
```

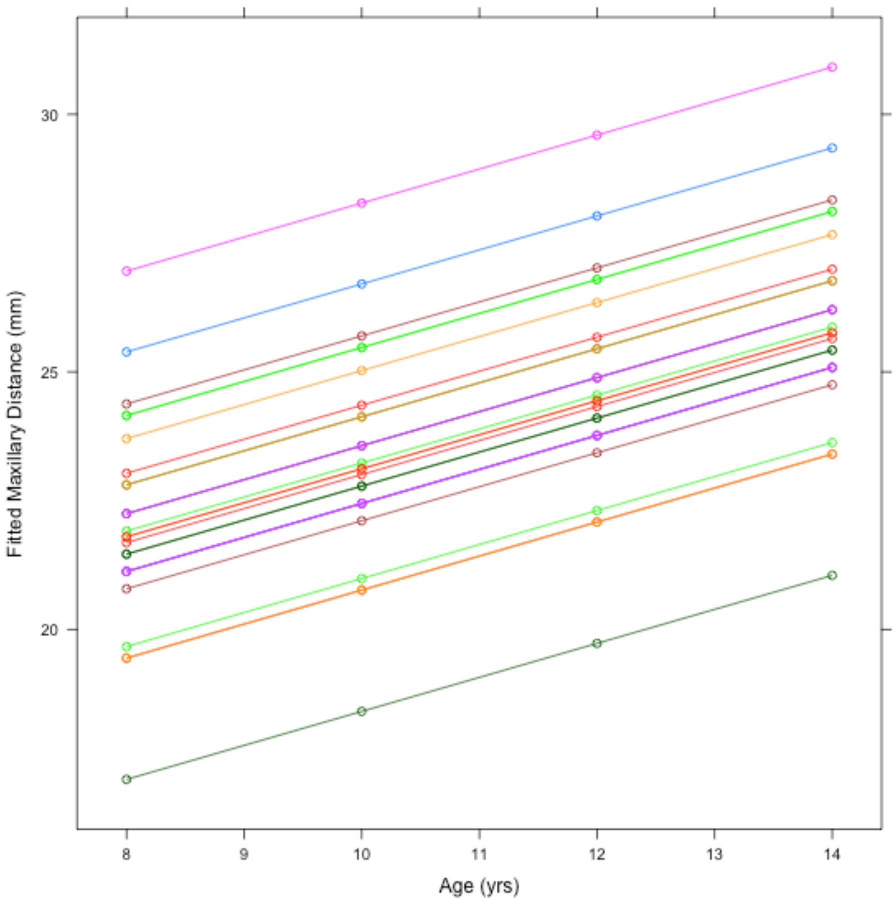
$$\boxed{\hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij} + \hat{u}_{i1} + \hat{u}_{i2} Age^*_{ij}}$$

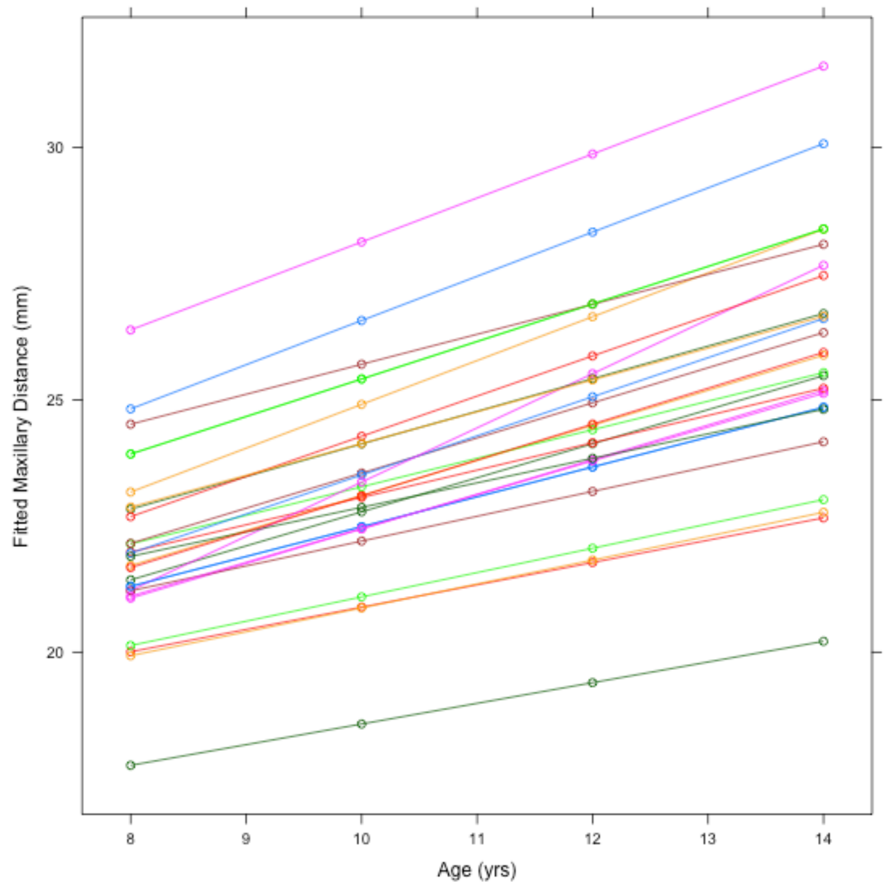
```
> 24.0231481 + 0.6601852*(c(8,10,12,14)-11)
                    + 3.4241123 + 0.21568501*(c(8,10,12,14)-11)
[1] 24.81965 26.57139 28.32313 30.07487
```

Random intercept vs. random intercept & slope
Fitted models
(n=27)

$$\hat{Y}_{ij} = \left(\hat{\beta}_0 + \hat{u}_{i1}\right) + \hat{\beta}_1 Age^*{}_{ij}$$

$$\hat{Y}_{ij} = \left(\hat{\beta}_0 + \hat{u}_{i1}\right) + \left(\hat{\beta}_1 + \hat{u}_{i2}\right) Age^*{}_{ij}$$





Mixed Effects Model
Random Intercept & Slope Model
Adding Categorical Covariate

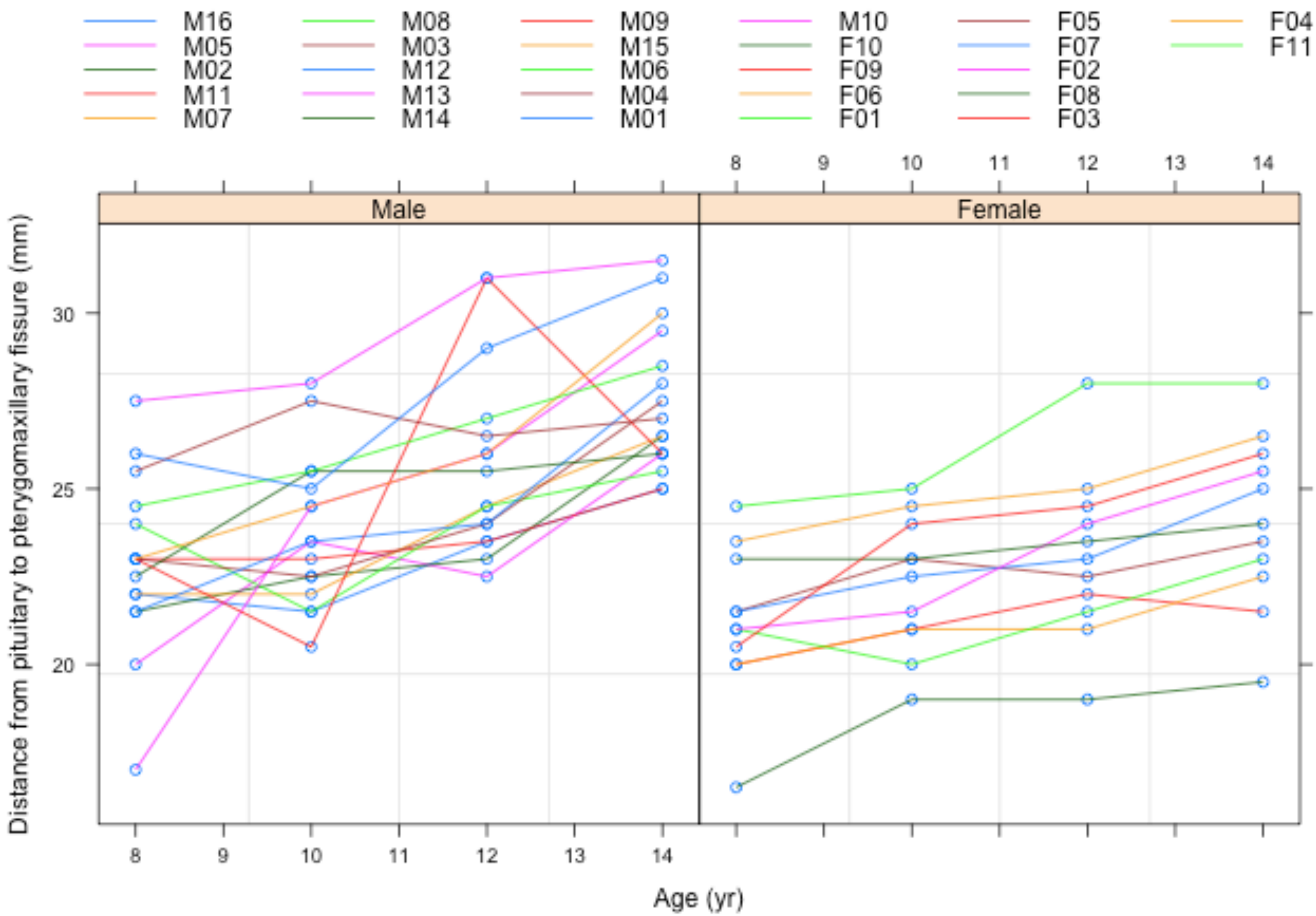
Random Intercept and Slope + Categorical Covariate

Data structure

Subject i	Sex (1=M,0=F)	Maxillary distance measurements (mm) Y_{ij} *i=subject, j=occasion*				Subject Mean
		8 yrs	10 yrs	12 yrs	14 yrs	
1	1	26	25	29	31	27.75
2	1	21.5	22.5	23	26.5	23.38
3	1	23	22.5	24	27.5	24.25
4	1	25.5	27.5	26.5	27	26.63
5	0	21	20	21.5	23	21.38
6	0	21	21.5	24	25.5	23.00
7	0	20.5	24	24.5	26	23.75
8	0	23.5	24.5	25	26.5	24.88

Random Intercept and Slope + Categorical Covariate

Sample data, n=27, 11 female, 16 male.



Random Intercept and Slope + Categorical Covariate

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 Sex_i}_{\text{Fixed}} + \underbrace{u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij}}_{\text{Random effects for intercept and slope + error (Stays the same)}}; \quad j = 1, \ldots, m_i .$$

Rearranging….

When Sex_i=1 *(Males)*

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 + u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij}$$

$$= \underbrace{\left(\beta_0 + \beta_2 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\left(\beta_1 + u_{i2}\right)}_{\text{Slope}} t_{ij} + \varepsilon_{ij}; \quad j = 1, \ldots, m_i .$$

When Sex_i=0 *(Females)*

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij}$$

$$= \underbrace{\left(\beta_0 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\left(\beta_1 + u_{i2}\right)}_{\text{Slope}} t_{ij} + \varepsilon_{ij}; \quad j = 1, \ldots, m_i .$$

Random Intercept and Slope + Categorical Covariate

```
> summary(fitt.sex)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  449.2339 467.8116 -217.6169

Random effects:
 Formula: ~I(age - 11) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.8320242 (Intr)
I(age - 11) 0.2264279 0.19
Residual    1.3100396

Fixed effects: distance ~ I(age - 11) + Sex
                Value Std.Error DF  t-value p-value
(Intercept) 24.897236 0.4852090 80 51.31239   0.000
I(age - 11)  0.660185 0.0712533 80  9.26533   0.000
SexFemale   -2.145489 0.7574536 25 -2.83250   0.009
 Correlation:
            (Intr) I(-11)
I(age - 11)  0.085
SexFemale   -0.636  0.000

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.08141614 -0.45675578  0.01552687  0.44704106  3.89437718

Number of Observations: 108
Number of Groups: 27
```

Random Intercept and Slope Model, Adding Categorical Variable + Interaction

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij} + \beta_2 Sex_i + \beta_3 \left(t_{ij} \times Sex_i\right)}_{\text{Fixed}} + \underbrace{u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij}}_{\text{Random}}; \quad j = 1, \ldots, m_i.$$

Rearranging….

When $Sex_i=1$ *(Males)*

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 + \beta_3 t_{ij} + u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij}$$

$$= \underbrace{\left(\beta_0 + \beta_2 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\left(\beta_1 + \beta_3 + u_{i2}\right)}_{\text{Slope}} t_{ij} + \varepsilon_{ij}; \quad j = 1, \ldots, m_i.$$

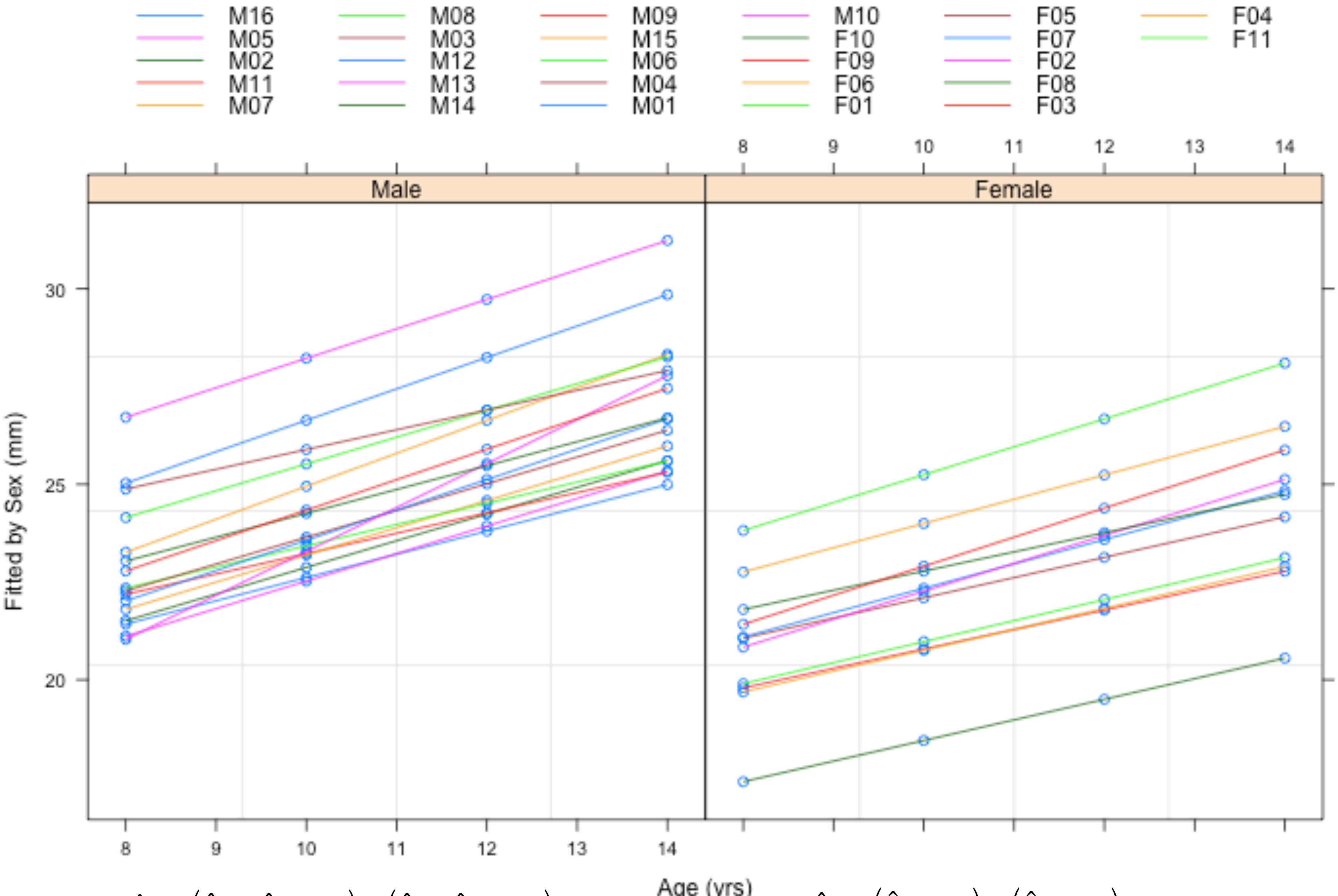
When $Sex_i=0$ *(Females)*

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + u_{i1} + u_{i2} t_{ij} + \varepsilon_{ij}$$

$$= \underbrace{\left(\beta_0 + u_{i1}\right)}_{\text{Intercept}} + \underbrace{\left(\beta_1 + u_{i2}\right)}_{\text{Slope}} t_{ij} + \varepsilon_{ij}; \quad j = 1, \ldots, m_i.$$

```
> summary(fitt.sexage)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  448.5817 469.7368 -216.2908

Random effects:
 Formula: ~I(age - 11) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.8303267 (Intr)
I(age - 11) 0.1803454 0.206
Residual    1.3100397

Fixed effects: distance ~ I(age - 11) + Sex + Sex * I(age - 11)
                          Value Std.Error DF  t-value p-value
(Intercept)           24.968750 0.4860007 79 51.37596  0.0000
I(age - 11)            0.784375 0.0859995 79  9.12069  0.0000
SexFemale             -2.321023 0.7614168 25 -3.04829  0.0054
I(age - 11):SexFemale -0.304830 0.1347353 79 -2.26243  0.0264
 Correlation:
                      (Intr) I(g-11) SexFml
I(age - 11)            0.102
SexFemale             -0.638 -0.065
I(age - 11):SexFemale -0.065 -0.638   0.102

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-3.168078484 -0.385939134  0.007103929  0.445154686  3.849463230

Number of Observations: 108
Number of Groups: 27
```

Random intercept and slope, adding a categorical variable (fixed)



Models: $\hat{Y}_{ij} = \left(\hat{\beta}_0 + \hat{\beta}_2 + \hat{u}_{i1}\right) + \left(\hat{\beta}_1 + \hat{\beta}_3 + \hat{u}_{i2}\right) Age_{ij}$ $\quad$ $\hat{Y}_{ij} = \left(\hat{\beta}_0 + \hat{u}_{i1}\right) + \left(\hat{\beta}_1 + \hat{u}_{i2}\right) Age_{ij}$

Data Structure in R

- Data sets can be stored as data frames or as groupedData objects.
- Multiple lines per subject
- In R, the variable Subject will be labeled as a "grouping" variable, used for plotting and LME analysis.

```
> head(Orthodont)
Grouped Data: distance ~ age | Subject
  distance age Subject  Sex
1     26.0   8     M01 Male
2     25.0  10     M01 Male
3     29.0  12     M01 Male
4     31.0  14     M01 Male
5     21.5   8     M02 Male
6     22.5  10     M02 Male
```

Data Structure in R

The groupedData object:

- Contains data stored as a data frame
- Designates special roles for some variables:
  - Response
  - Primary covariate
  - Grouping factor

response ~ primary | grouping

- R stores a formula with the data, which can be accessed through the formula() function

```
> formula(Orthodont)
distance ~ age | Subject
```

Also useful when fitting models

Data Structure in R

- Since there's multiple lines per subject, `dim(dat)` or `nrow(dat)` does not longer work to obtain the number of subjects.
- Also, it may be of interest to check the balance of the data.

```
> female.dat <- dat[Sex=="Female",]
> # no. repeated measurements per subject:
> table(female.dat$Subject)

F10 F09 F06 F01 F05 F07 F02 F08 F03 F04 F11
  4   4   4   4   4   4   4   4   4   4   4
> # no. of subjects:
> length(table(female.dat$Subject))
[1] 11
```

Alternatively, the `getGroups()` function can be used if the data has the groupedData format:

```
> table(getGroups(female.dat))

F10 F09 F06 F01 F05 F07 F02 F08 F03 F04 F11
  4   4   4   4   4   4   4   4   4   4   4
```

Data Structure in R

- To check whether the data is balanced with respect of a covariate, e.g. Age in the Maxillary Distance data:

```
> table(getCovariate(female.dat),getGroups(female.dat))

    F10 F09 F06 F01 F05 F07 F02 F08 F03 F04 F11
 8    1   1   1   1   1   1   1   1   1   1   1
 10   1   1   1   1   1   1   1   1   1   1   1
 12   1   1   1   1   1   1   1   1   1   1   1
 14   1   1   1   1   1   1   1   1   1   1   1
```

Plotting longitudinal data in R

Some ways to plot involve:

- The groupedData structure.
- The `interaction.plot()` function
- The lattice library (and the “lattice” R package).

Plotting longitudinal data in R
Plots using a groupedData object

R Code to create a groupedData object:

```
newData <-
  groupedData( response ~ primary | grouping,
               data = as.data.frame( dataset ),
               outer = ~ covariate,
               labels = list( x = "primary",
                              y = "response" ),
               units = list( x = "(units)", y = "(units)") )
```

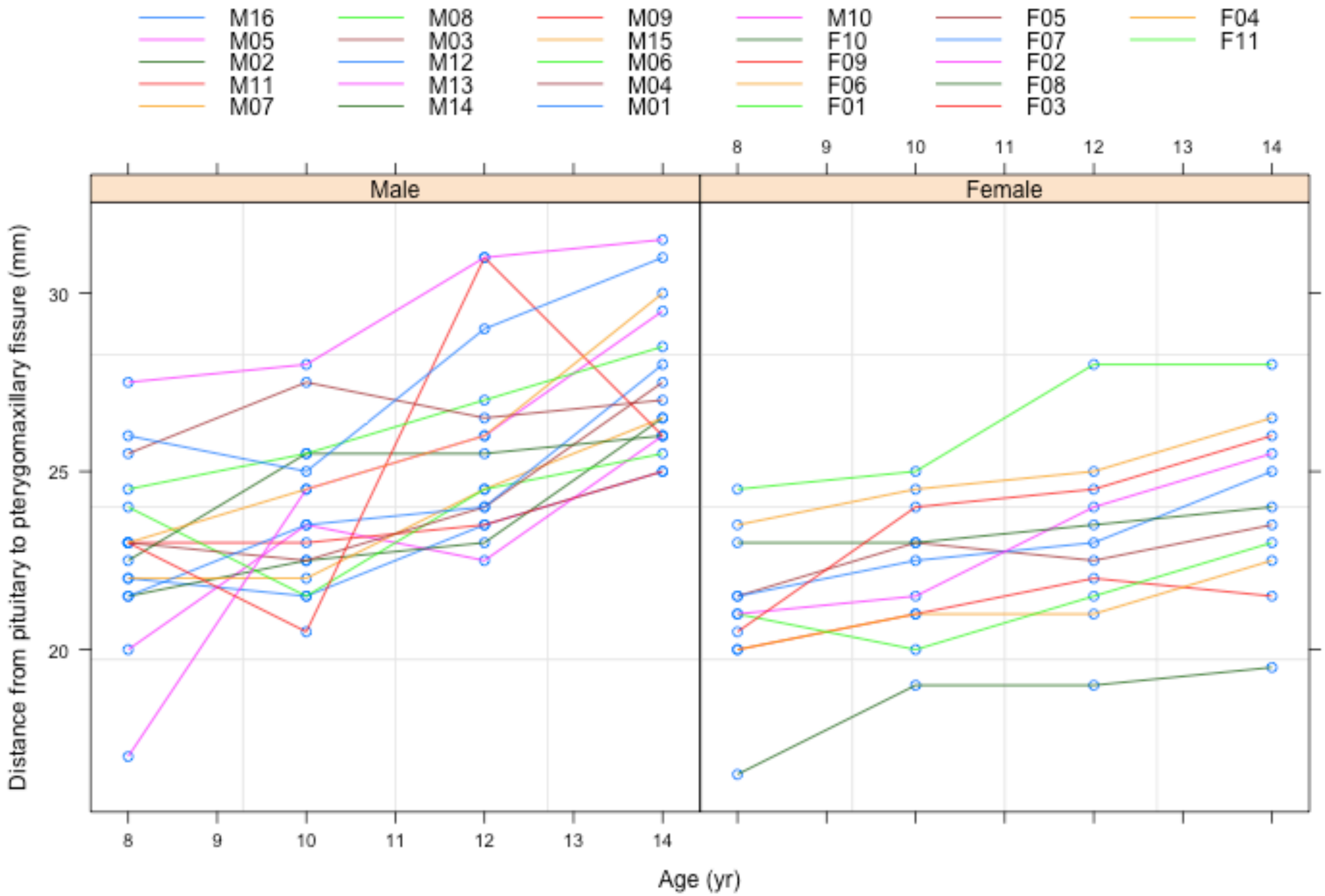
- The “data” option can also be used together with a function to read data, such as the `read.table()` function:

```
          data = read.table(dataset,header=T)
```

- The “outer” option is designated to variables that do not vary within a subject, e.g., sex.

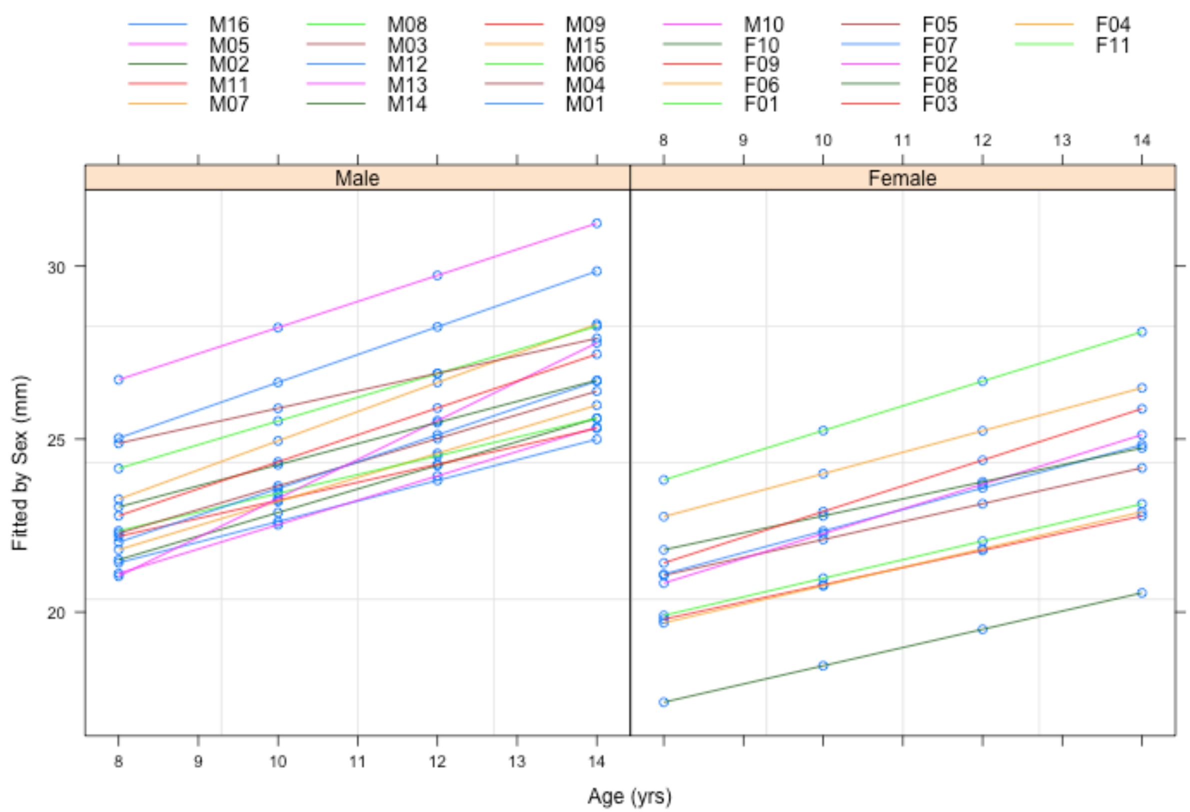
Plotting longitudinal data in R

```
plot(Orthodont, outer=~Sex, aspect=1)
```



Plotting longitudinal data in R

```
> fitt.sex <- lme(. . ., data=dat)
> dat$fitted.sex <- fitted(fitt.sex)

> formula(dat.sex)
distance ~ age | Subject

> dat.sex <- update( dat , fitted.sex ~ age | Subject)

> plot(dat.sex, outer=~Sex, aspect=1)
```



Plotting longitudinal data in R

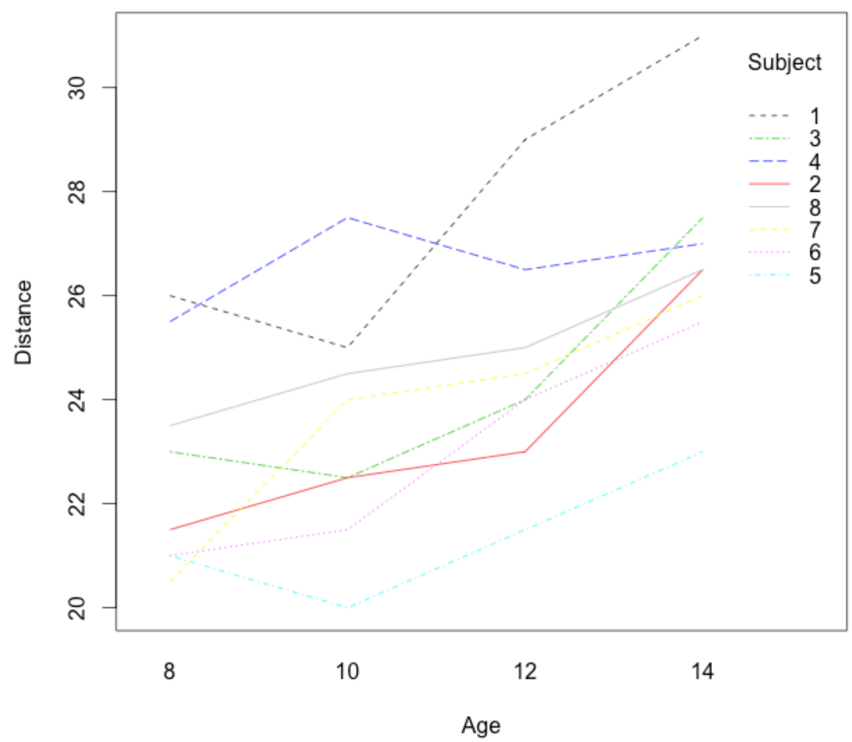
`interaction.plot()`

Plots trajectories ("traces") in one single panel – useful when not too many:

```
interaction.plot(x.factor,trace.factor,response)
```

E.g.,

```
interaction.plot(age,Subject,distance,col=1:8)
```



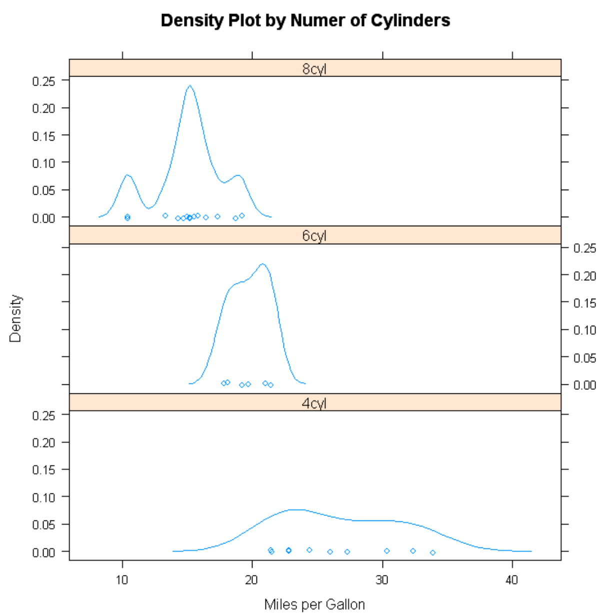
Plotting longitudinal data in R

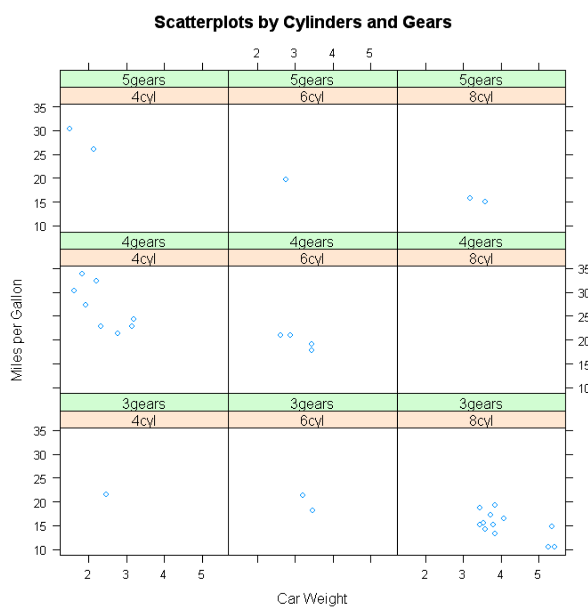
The lattice library (lattice R package)

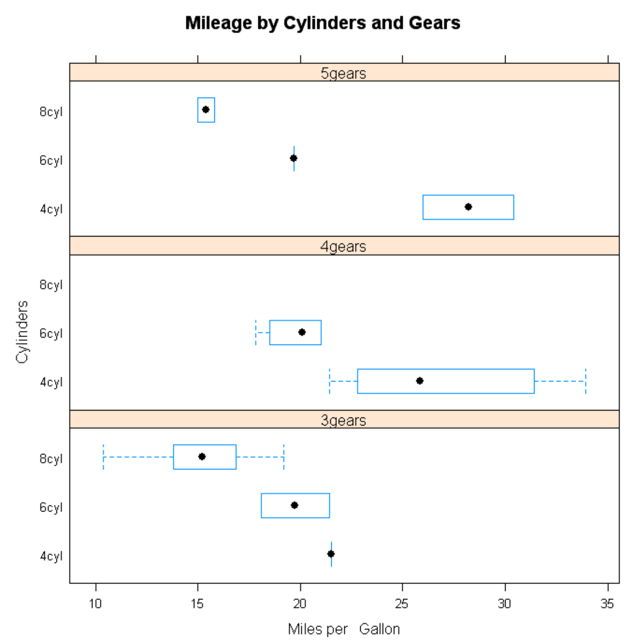
- Used to plot information in multiple panels: to display multivariate relationships.
- The typical format is:

```
graph_type(formula, data=)
```

- In the examples below, suppose x,y are continuous variables and A,B are factors.





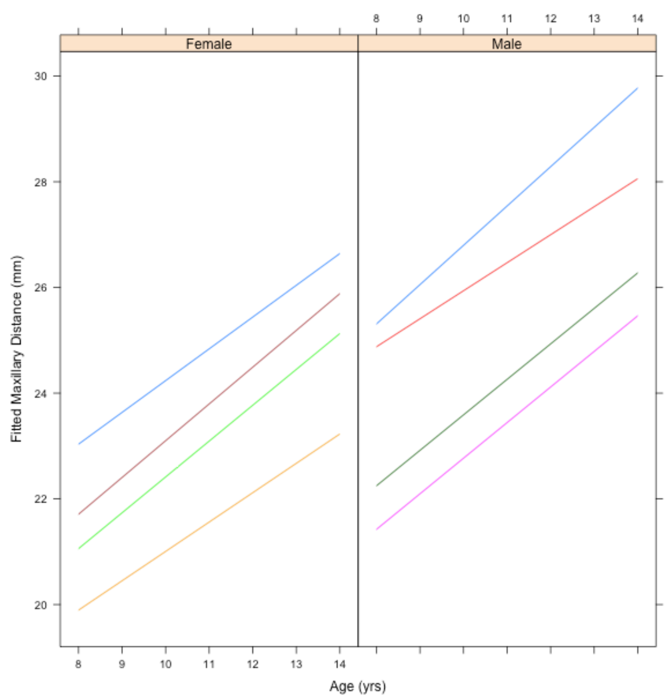


densityplot()	xyplot()	bwplot()
x|A*B	y|x|A	X~A or A~x

See http://www.statmethods.net/advgraphs/trellis.html

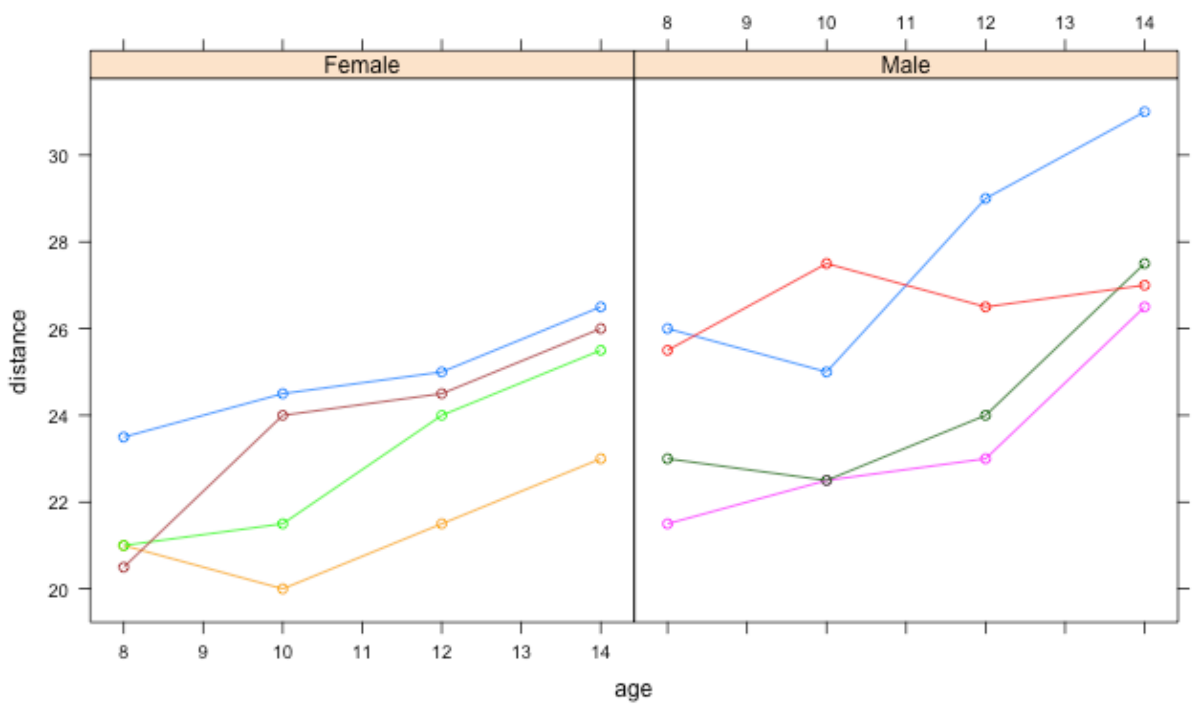
Plotting longitudinal data in R
The lattice library

```
xyplot(fitted~age| Sex,
            groups=Subject,
            data=dat0,
            type="l",
            ylab="Fitted Maxillary Distance (mm)",
            xlab="Age (yrs)",aspect=2)
```



Plotting longitudinal data in R
The lattice library

```
xyplot(distance~age|Sex,
            groups=Subject,
            data=dat0,
            type="b",
            aspect=1)
```



Simplification of lme() with groupedData

Since

- the data set used has a groupedData structure
- by default the random effects have the same form as the fixed effects,

… the following two fits give the same results:

```
# fit with random intercept & slope

fitt2 <- lme(distance ~ I(age-11),
                  data = dat,
                  random = ~ I(age-11)|Subject)

fitt2.same <- lme(distance ~ I(age-11), data = dat)
```