

Problem: **Orthodontic Growth Curves in Girls.** We will use the R dataset `Orthodont` (in the `nlme` library or package). Use the `plot` command in the help file. It appears that there are qualitative differences between boys and girls, so we will just model the data from the female subjects by

```
> OrthoF <- Orthodont[Orthodont$Sex == "Female",]
```

(a) Plot the data using,

```
> plot(OrthoF)
> OrthoF <- Orthodont[Orthodont$Sex == "Female",]
> postscript("hw8plot3.ps", horizontal=F)
> par(pty="s")
> plot(OrthoF)
> dev.off()
```

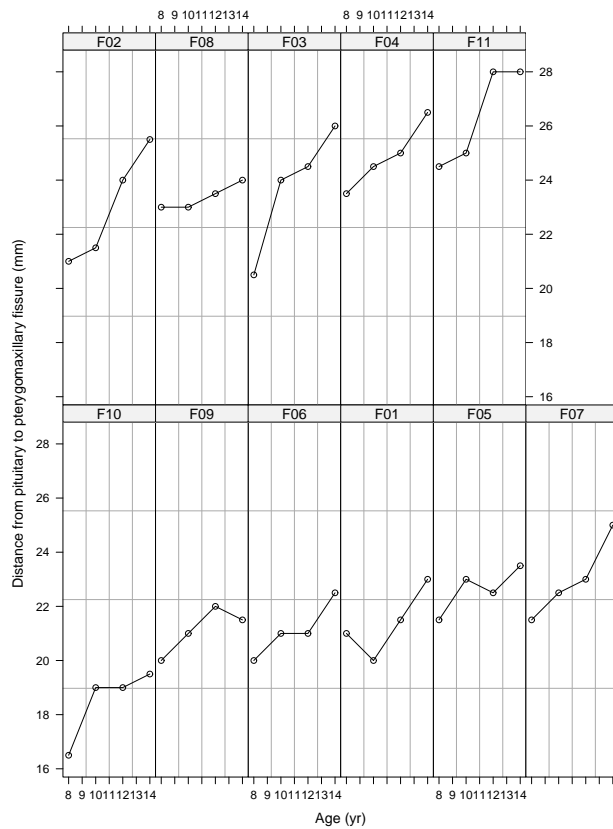


Figure 1:

(b) The model we want to fit in matrix notation is:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 11$$

with $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. The following matrices are the same for all $i = 1, \dots, 11$

$$\mathbf{X}_i = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}$$

$$\mathbf{Z}_i = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The two dimensional fixed effects vector $\boldsymbol{\beta}$ consists of a mean intercept and common slope or growth rate. The one-dimensional random effects vector \mathbf{u}_i describes a shift in the intercept for each subject i . The matrix $\mathbf{D} = \sigma_u^2$ is a 1 by 1 matrix in this case. Use the R `lme` function to fit this model. There are four parameters in this model. What are the parameter estimates? The summary function is useful.

```
> fit1 <- lme(distance~age, data=OrthoF, random=~1 | Subject)
```

```
> summary(fit1)
```

```
Linear mixed-effects model fit by REML
```

```
Data: OrthoF
```

```
      AIC      BIC    logLik
149.2183 156.169 -70.60916
```

```
Random effects:
```

```
Formula: ~1 | Subject
```

```
(Intercept) Residual
```

```
StdDev:      2.06847 0.7800331
```

```
Fixed effects: distance ~ age
```

```
              Value Std.Error DF   t-value p-value
(Intercept) 17.372727 0.8587419 32 20.230440     0
age           0.479545 0.0525898 32  9.118598     0
```

```
Correlation:
```

```
(Intr)
```

```
age -0.674
```

```
Standardized Within-Group Residuals:
```

```
      Min      Q1      Med      Q3      Max
-2.2736479 -0.7090164  0.1728237  0.4122128  1.6325181
```

```
Number of Observations: 44
```

```
Number of Groups: 11
```

The REML estimates are $\hat{\beta}_0 = 17.37, \hat{\beta}_1 = 0.479, \hat{\sigma}_{intercept}^2 = (2.068)^2 = 4.279, \hat{\sigma}^2 = (0.78)^2 = 0.6085$.

(c) We have made the assumption of a common slope or growth rate for all subjects. To test this we can fit a model with random effects for both the intercept and the slope. The `update` function is useful for this model modification. Compare the two models with the `anova` function. What do you conclude?

```
> fit2 <- update(fit1, random=~age | Subject)
OR
> fit2 <- lme(distance~age, data=OrthoF, random=~age | Subject)
> summary(fit2)
Linear mixed-effects model fit by REML
Data: OrthoF
      AIC      BIC    logLik
149.4287 159.8547 -68.71435

Random effects:
Formula: ~age | Subject
Structure: General positive-definite, Log-Cholesky parametrization
          StdDev   Corr
(Intercept) 1.8841866 (Intr)
age          0.1609278 -0.354
Residual    0.6682746

Fixed effects: distance ~ age
              Value Std.Error DF   t-value p-value
(Intercept) 17.372727 0.7606027 32 22.840737    0
age          0.479545 0.0662140 32  7.242353    0
Correlation:
  (Intr)
age -0.637

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-1.85438224 -0.46784889  0.06779759  0.42976634  1.59215840

Number of Observations: 44
Number of Groups: 11
> anova(fit2, fit1)
      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
fit2     1  6 149.4287 159.8547 -68.71435
fit1     2  4 149.2183 156.1690 -70.60916 1 vs 2 3.789622 0.1503
```

$H_0 : \sigma_{slope}^2 = 0$ against $H_1 : \sigma_{slope}^2 \neq 0$.

From the LRT with a p-value of 0.1503, do not reject the null hypothesis. The “best” model is the simpler model.

How well does the “best” model fit the data? (Note: For model diagnostics, there is a plot function for the lme object.)

Note: The notation in the book follows the paper by Laird and Ware (1982). You can find the link to it off the course calendar.

```
> postscript("hw8plot4.ps", horizontal=F)
> par(pty="s")
> plot(fit1)
> dev.off()
```

The model fits the data well with randomly distributed residuals with respect to the fitted values.

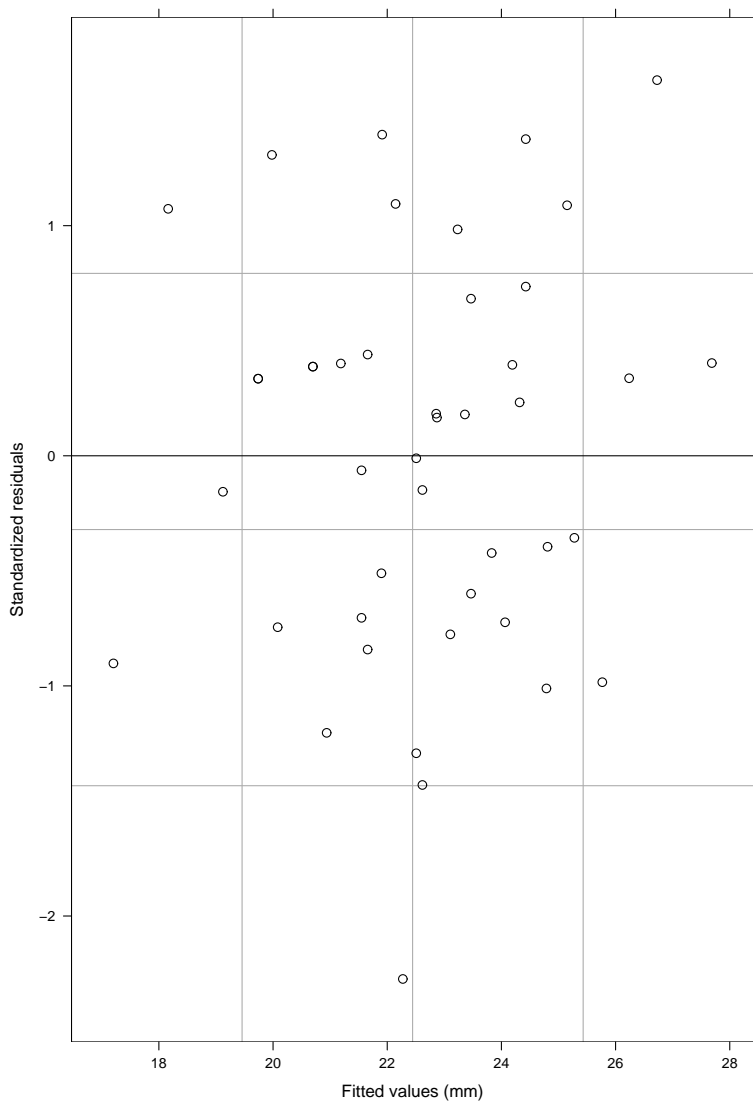


Figure 2:

What about predictions or fitted values?

```
> postscript("hw8plot5.ps", horizontal=F)  
> par(pty="s")  
> plot(augPred(fit1))  
> dev.off()
```

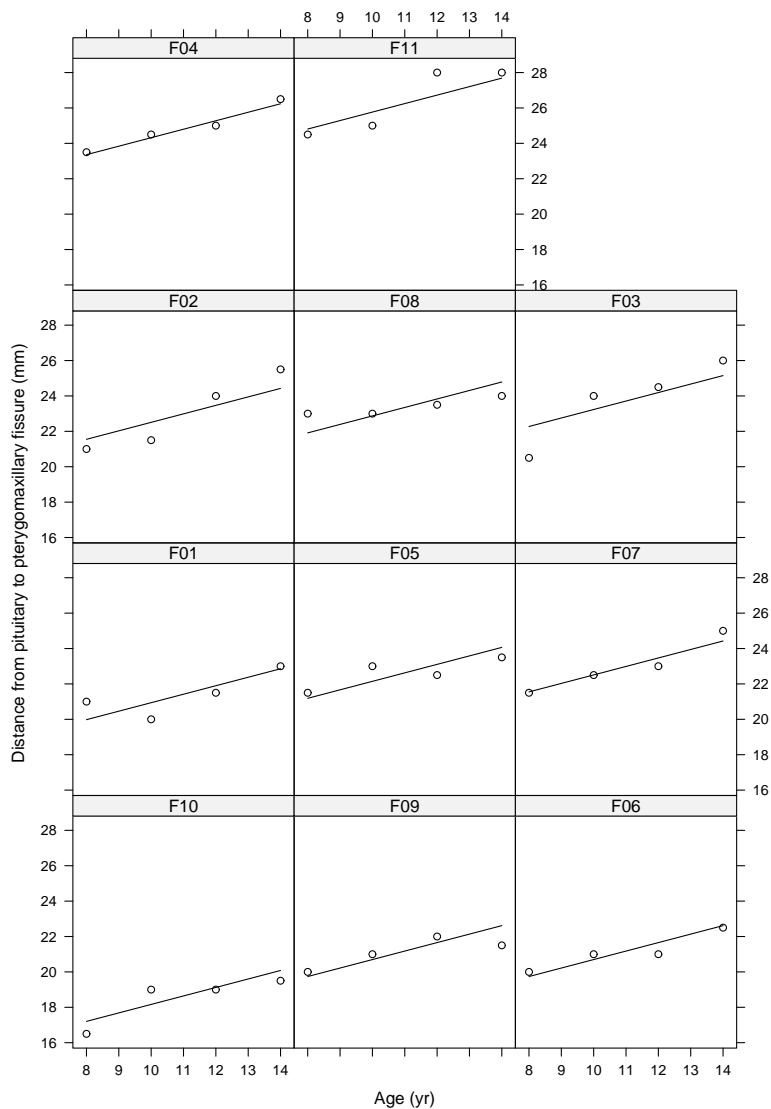


Figure 3: