BIOS6643. L9 Specifying LMM through G_i and R_i structures

Dental study

Description

- The orthodontic study data of Potthoff and Roy (1964).
- World famous data set that is used to introduce features of longitudinal data modeling and analysis
- A study was conducted involving 27 children, 16 boys and 11 girls
- For each child, the distance (mm) from the center of the pituitary to the pterygomaxillary fissure was measured at ages 8, 10, 12, and 14 years of age
- The pterygomaxillary fissure is a vertical opening in the human skull.
- **Objectives** of the study included:
  - Determine if distances over time are larger on average for boys than for girls
  - Determine if the rate of change of distance over time is different for boys and girls.

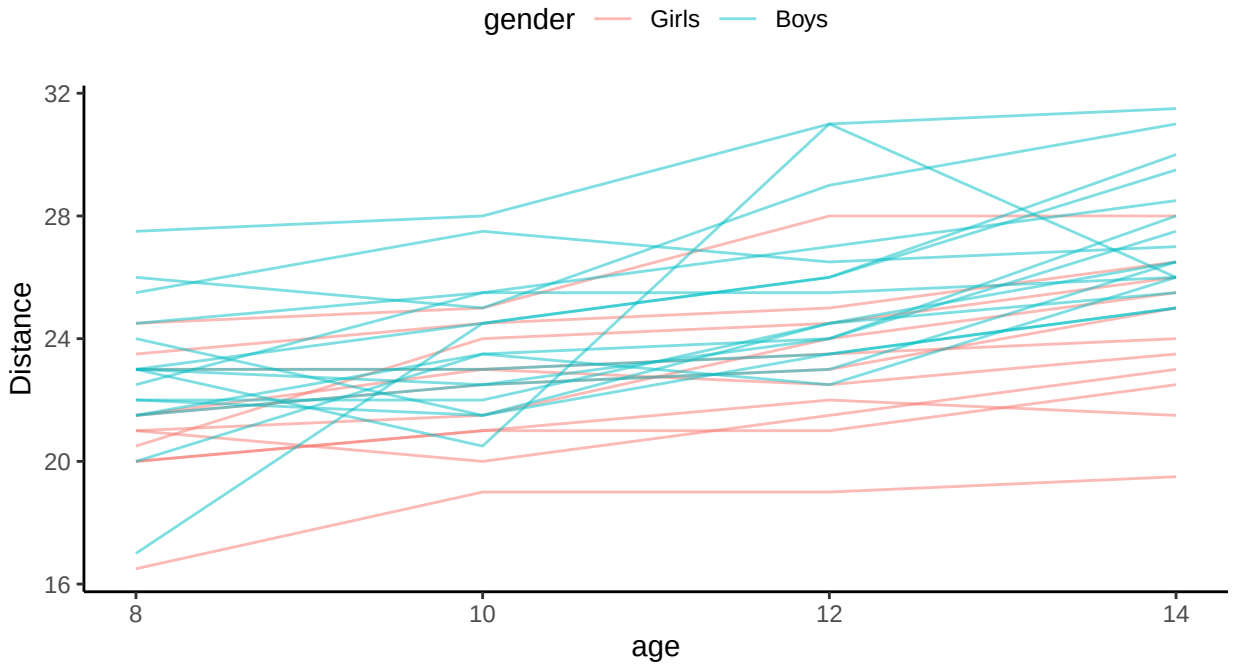
```
#  Read in the data
dat.den <- read.table("/Users/juarezce/Documents/OneDrive - The University of Colorado Denver/BIOS6643/
dat.den <- dat.den[,2:5]      #  remove the first column
colnames(dat.den) <- c("id","age","distance","gender")

#  Total number of individuals
m <- max(dat.den$id)

head(dat.den, 2)
```

```
##   id age distance gender
## 1  1   8       21      0
## 2  1  10       20      0
```

```
table(dat.den$gender)
```

```
##
##  0  1
## 44 64
```

Some descriptives:



a. Common G matrix for both genders with random intercept and slopes

```
#  (a) Common G matrix for both genders, default diagonal within-child
#      covariance matrix R_i with same variance sigma^2 for each
#      gender.

dental.lme.a <- lme(distance ~ -1 + gender + age:gender,data=dat.den,
                    random = ~ age | id,method="ML")
summary(dental.lme.a)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat.den
##       AIC     BIC   logLik
##   443.806 465.263 -213.903
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 2.134688 (Intr)
## age         0.154139 -0.603
## Residual    1.310040
##
## Fixed effects:  distance ~ -1 + gender + age:gender
##                     Value Std.Error DF   t-value p-value
## genderGirls     17.372727 1.2045404 25 14.422702       0
## genderBoys      16.340625 0.9987521 25 16.361042       0
## genderGirls:age  0.479545 0.1017051 80  4.715058       0
## genderBoys:age   0.784375 0.0843294 80  9.301321       0
```

```
##  Correlation:
##                  gndrGr gndrBy gndrG:
## genderBoys        0.00
## genderGirls:age -0.88   0.00
## genderBoys:age   0.00  -0.88   0.00
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.33603363 -0.41539842  0.01039175  0.49169519  3.85819292
##
## Number of Observations: 108
## Number of Groups: 27
```

```
beta.a <- fixed.effects(dental.lme.a)  #  beta, also fixef(dental.lme.a)
b.a <- random.effects(dental.lme.a)    #  posterior modes bi, also ranef(dental.lme.a)
sebeta.a <- summary(dental.lme.a)$tTable[,"Std.Error"]
## Recall we can get the var-cov of fixed coefficients using 'varFix'
## dental.lme.a$varFix

G.a <- getVarCov(dental.lme.a, type="random.effects")   #  G matrix
sigma2.a <- dental.lme.a$sigma^2  # sigma^2
V.a <- getVarCov(dental.lme.a, type="marginal", individual=1)   # V_i
R.a <- getVarCov(dental.lme.a, type="conditional", individual=1)   # R_i
G.a
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
## (Intercept)     4.55690 -0.198250
## age            -0.19825  0.023759
##   Standard Deviations: 2.1347 0.15414
```

```
V.a
```

```
## id 1
## Marginal variance covariance matrix
##        1      2      3      4
## 1 4.6216 2.8891 2.8727 2.8563
## 2 2.8891 4.6839 3.0464 3.1251
## 3 2.8727 3.0464 4.9363 3.3938
## 4 2.8563 3.1251 3.3938 5.3788
##   Standard Deviations: 2.1498 2.1642 2.2218 2.3192
```

```
R.a
```

```
## id 1
## Conditional variance covariance matrix
##        1      2      3      4
## 1 1.7162 0.0000 0.0000 0.0000
## 2 0.0000 1.7162 0.0000 0.0000
## 3 0.0000 0.0000 1.7162 0.0000
## 4 0.0000 0.0000 0.0000 1.7162
##   Standard Deviations: 1.31 1.31 1.31 1.31
```

b. Common G matrix with random intercepts and slopes, diagonal within-child covariance matrix R_i with different variance for each gender

```
dental.lme.b <- lme(distance ~ -1 + gender + age:gender,data=dat.den,
                    random = ~ age | id, weights = varIdent(form = ~ 1 | gender),
                    method="ML")
beta.b <- fixed.effects(dental.lme.b)  #  beta
sebeta.model.b <-  sqrt(diag(dental.lme.b$varFix))
b.b <- random.effects(dental.lme.b)    #  posterior modes bi
G.b <- getVarCov(dental.lme.b, type="random.effects")   #  G matrix
R.b.1 <- getVarCov(dental.lme.b,type="conditional",individual=1)   # R_1; first girl
R.b.12 <- getVarCov(dental.lme.b,type="conditional",individual=12)   # R_12; first boy
G.b
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
## (Intercept)     3.19860 -0.110360
## age            -0.11036  0.019766
##   Standard Deviations: 1.7885 0.14059
```

```
R.b.1
```

```
## id 1
## Conditional variance covariance matrix
##         1       2       3       4
## 1 0.44491 0.00000 0.00000 0.00000
## 2 0.00000 0.44491 0.00000 0.00000
## 3 0.00000 0.00000 0.44491 0.00000
## 4 0.00000 0.00000 0.00000 0.44491
##   Standard Deviations: 0.66702 0.66702 0.66702 0.66702
```

```
R.b.12
```

```
## id 12
## Conditional variance covariance matrix
##        1      2      3      4
## 1 2.6294 0.0000 0.0000 0.0000
## 2 0.0000 2.6294 0.0000 0.0000
## 3 0.0000 0.0000 2.6294 0.0000
## 4 0.0000 0.0000 0.0000 2.6294
##   Standard Deviations: 1.6215 1.6215 1.6215 1.6215
```

```
#  As with gls(), when a weight statement is used, the standard
#  deviation for the first group is sigma, and those for other groups
#  are parameterized as this standard deviation x a factor. To get
#  sigma^2_B, we must extract that factor from the varStruct object:
#
dental.lme.b$modelStruct$varStruct
```

```
## Variance function structure of class varIdent representing
##    Girls     Boys
## 1.000000 2.431012
```

```
sigma2vec.b <- matrix((1/unique(attributes(dental.lme.b$modelStruct$varStruct)$weights)*dental.lme.b$sig
                      nrow=1,byrow=TRUE)
colnames(sigma2vec.b) <- c("sigma2.b.G","sigma2.b.B")
sigma2vec.b
```

```
##      sigma2.b.G sigma2.b.B
## [1,]  0.4449132   2.629357
```

```
V.b.1 <- getVarCov(dental.lme.b,type="marginal",individual=1)   # V_1 (girl)
V.b.12 <- getVarCov(dental.lme.b,type="marginal",individual=12)   # V_12 (boy)
V.b.1
```

```
## id 1
## Marginal variance covariance matrix
##        1      2      3      4
## 1 3.1428 2.7934 2.8889 2.9845
## 2 2.7934 3.4129 3.1426 3.3172
## 3 2.8889 3.1426 3.8412 3.6499
## 4 2.9845 3.3172 3.6499 4.4275
##   Standard Deviations: 1.7728 1.8474 1.9599 2.1042
```

```
V.b.12
```

```
## id 12
## Marginal variance covariance matrix
##        1      2      3      4
## 1 5.3272 2.7934 2.8889 2.9845
## 2 2.7934 5.5973 3.1426 3.3172
## 3 2.8889 3.1426 6.0256 3.6499
## 4 2.9845 3.3172 3.6499 6.6120
##   Standard Deviations: 2.3081 2.3659 2.4547 2.5714
```

c. Common G matrix for both genders, and common within-child AR(1)

```
## Note that we cannot fit a model with separate G matrices for each gender

dental.lme.c <- lme(distance ~ -1 + gender + age:gender, data=dat.den,
                    random = ~ age | id,
                    correlation=corAR1(form = ~ age | id),
                    method="ML")
summary(dental.lme.c)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat.den
##       AIC      BIC   logLik
##   445.806 469.9451 -213.903
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 2.134688 (Intr)
## age         0.154139 -0.603
## Residual    1.310040
##
## Correlation Structure: ARMA(1,0)
##  Formula: ~age | id
##  Parameter estimate(s):
## Phi1
##    0
```

```
## Fixed effects:  distance ~ -1 + gender + age:gender
##                     Value Std.Error DF   t-value p-value
## genderGirls     17.372727 1.2045404 25 14.422702       0
## genderBoys      16.340625 0.9987521 25 16.361042       0
## genderGirls:age  0.479545 0.1017051 80  4.715058       0
## genderBoys:age   0.784375 0.0843294 80  9.301321       0
##  Correlation:
##                 gndrGr gndrBy gndrG:
## genderBoys       0.00
## genderGirls:age -0.88   0.00
## genderBoys:age   0.00  -0.88   0.00
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.33603363 -0.41539842  0.01039175  0.49169520  3.85819292
##
## Number of Observations: 108
## Number of Groups: 27
```

```
beta.c <- fixed.effects(dental.lme.c)
b.c <- random.effects(dental.lme.c)
sebeta.d <- sqrt(diag(dental.lme.c$varFix))
G.c <- getVarCov(dental.lme.c, type="random.effects")   #  G
sigma2.c <- dental.lme.c$sigma^2  # sigma^2
V.c <- getVarCov(dental.lme.c,type="marginal", individual=1)   # V_i
R.c <- getVarCov(dental.lme.c,type="conditional", individual=1)   # R_i
G.c
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
## (Intercept)     4.55690 -0.198250
## age            -0.19825  0.023759
##   Standard Deviations: 2.1347 0.15414
```

```
V.c
```

```
## id 1
## Marginal variance covariance matrix
##        1      2      3      4
## 1 4.6216 2.8891 2.8727 2.8563
## 2 2.8891 4.6839 3.0464 3.1251
## 3 2.8727 3.0464 4.9363 3.3938
## 4 2.8563 3.1251 3.3938 5.3788
##   Standard Deviations: 2.1498 2.1642 2.2218 2.3192
```

```
R.c
```

```
## id 1
## Conditional variance covariance matrix
##        1      2      3      4
## 1 1.7162 0.0000 0.0000 0.0000
## 2 0.0000 1.7162 0.0000 0.0000
## 3 0.0000 0.0000 1.7162 0.0000
## 4 0.0000 0.0000 0.0000 1.7162
##   Standard Deviations: 1.31 1.31 1.31 1.31
```

```
##  The correlation parameter alpha, which is constrained to be |alpha|<=1,
##  is estimated to be 0 (= Phi1).  So this fit is the same as (a)
```

d. Common G matrix for both genders, common within-child AR(1) (exponential correlation)

```
dental.lme.d <- lme(distance ~ -1 + gender + age:gender, data=dat.den,
                    random = ~ age | id,
                    correlation=corExp(form = ~ age | id),
                    method="ML")
summary(dental.lme.d)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: dat.den
##       AIC      BIC   logLik
##   445.806 469.9452 -213.903
##
## Random effects:
##  Formula: ~age | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 2.135114 (Intr)
## age         0.154165 -0.603
## Residual    1.310033
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##     range
## 0.1495162
## Fixed effects:  distance ~ -1 + gender + age:gender
##                     Value Std.Error DF   t-value p-value
## genderGirls     17.372727 1.2046075 25 14.421898       0
## genderBoys      16.340626 0.9988078 25 16.360130       0
## genderGirls:age  0.479545 0.1017084 80  4.714903       0
## genderBoys:age   0.784375 0.0843322 80  9.301015       0
##  Correlation:
##                 gndrGr gndrBy gndrG:
## genderBoys       0.00
## genderGirls:age -0.88   0.00
## genderBoys:age   0.00  -0.88   0.00
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.33579191 -0.41535636  0.01040691  0.49167626  3.85821320
##
## Number of Observations: 108
## Number of Groups: 27
```

```
beta.d <- fixed.effects(dental.lme.d)
b.d <- random.effects(dental.lme.d)
sebeta.model.d <-  sqrt(diag(dental.lme.d$varFix))
```

```
G.d <- getVarCov(dental.lme.d, type="random.effects")   #  G
sigma2.d <- dental.lme.d$sigma^2  # sigma^2
R.d <- getVarCov(dental.lme.d,type="conditional",individual=1)  #  R_i
V.d <- getVarCov(dental.lme.d,type="marginal",individual=1)   # V_i
R.d
```

```
## id 1
## Conditional variance covariance matrix
##            1          2          3          4
## 1 1.7162e+00 2.6622e-06 4.1296e-12 6.4058e-18
## 2 2.6622e-06 1.7162e+00 2.6622e-06 4.1296e-12
## 3 4.1296e-12 2.6622e-06 1.7162e+00 2.6622e-06
## 4 6.4058e-18 4.1296e-12 2.6622e-06 1.7162e+00
##   Standard Deviations: 1.31 1.31 1.31 1.31
```

```
V.d
```

```
## id 1
## Marginal variance covariance matrix
##        1      2      3      4
## 1 4.6218 2.8891 2.8726 2.8561
## 2 2.8891 4.6838 3.0462 3.1248
## 3 2.8726 3.0462 4.9360 3.3935
## 4 2.8561 3.1248 3.3935 5.3783
##   Standard Deviations: 2.1498 2.1642 2.2217 2.3191
```

Comparison of models

```
## Compare the fitted models via AIC and BIC; model (b), with
#  diagonal R_i with gender-specific within-child variances and common
#  random intercept and slopes covariance matrix G is preferred

anova(dental.lme.a, dental.lme.b, dental.lme.c, dental.lme.d)
```

```
##              Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## dental.lme.a     1  8 443.8060 465.2630 -213.9030
## dental.lme.b     2  9 424.0424 448.1816 -203.0212 1 vs 2 21.76356  <.0001
## dental.lme.c     3  9 445.8060 469.9451 -213.9030
## dental.lme.d     4  9 445.8060 469.9452 -213.9030
```

Refitting final model

```
##  Refit model (b) using REML and get
dental.lme.b.reml <- lme(distance ~ -1 + gender + age:gender, data=dat.den,
                  random = ~ age | id, weights = varIdent(form = ~ 1 | gender))
beta.b.reml <- fixed.effects(dental.lme.b.reml)  #  beta
sebeta.model.b.reml <-  sqrt(diag(dental.lme.b$varFix))
b.b.reml <- random.effects(dental.lme.b.reml)    #  posterior modes bi
G.b.reml <- getVarCov(dental.lme.b.reml, type="random.effects")   #  G
G.b.reml
```

```
## Random effects variance covariance matrix
##             (Intercept)       age
```

```
## (Intercept)     3.89140 -0.155210
## age            -0.15521  0.024498
##   Standard Deviations: 1.9727 0.15652
```

```
##  Random effects empirical Bayes estimates bhat_i for first 5 girls
b.b.reml[1:5,]
```

```
##   (Intercept)         age
## 1  -0.4492183 -0.07172941
## 2  -1.4036605  0.16096191
## 3  -1.0781035  0.19761263
## 4   1.7689177  0.03476773
## 5   1.0543227 -0.09938676
```

```
## PA predicted values $X_i \times betahat$ are produced by level=0;
## SS predicted values X_i betahat + Z_i bhat_i are produced by level =  1;
## both are gotten by level = 0:1

fitted(dental.lme.b.reml,level=0:1)[1:5,]
```

```
##      fixed       id
## 1 21.20909 20.18604
## 2 22.16818 21.00167
## 3 23.12727 21.81730
## 4 24.08636 22.63293
## 5 21.20909 21.09313
```

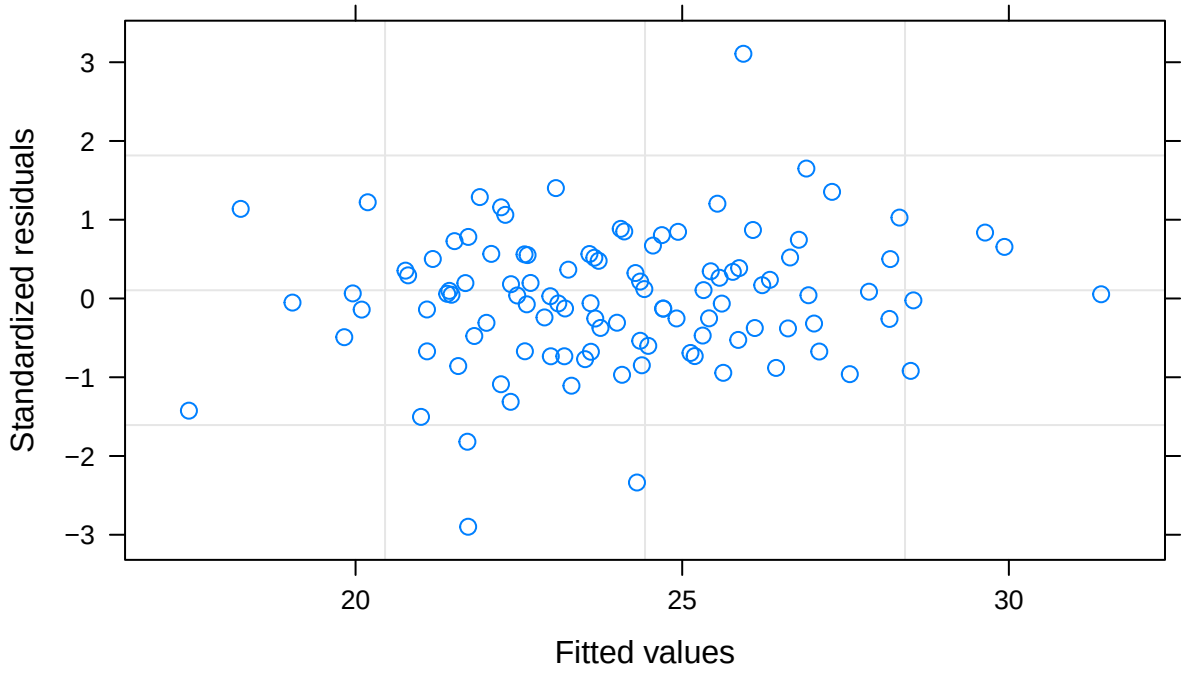
```
## Can also extract both marginal (PA) residuals  Y_i-X_i betahat,
## and conditional (SS) residuals Y_i-X_i betahat - Z_i bhat_i
## level = 0 gives PA, level = 1 gives SS (default), and level = 0:1 gives both;
## type = default is "raw" residuals; can get standardized "pearson" (type=pearson)

residuals(dental.lme.b.reml, level=0:1)[1:5, ]
```

```
##        fixed          id
## 1 -0.2090909  0.81396273
## 2 -2.1681818 -1.00166935
## 3 -1.6272727 -0.31730143
## 4 -1.0863636  0.36706649
## 5 -0.2090909 -0.09312569
```

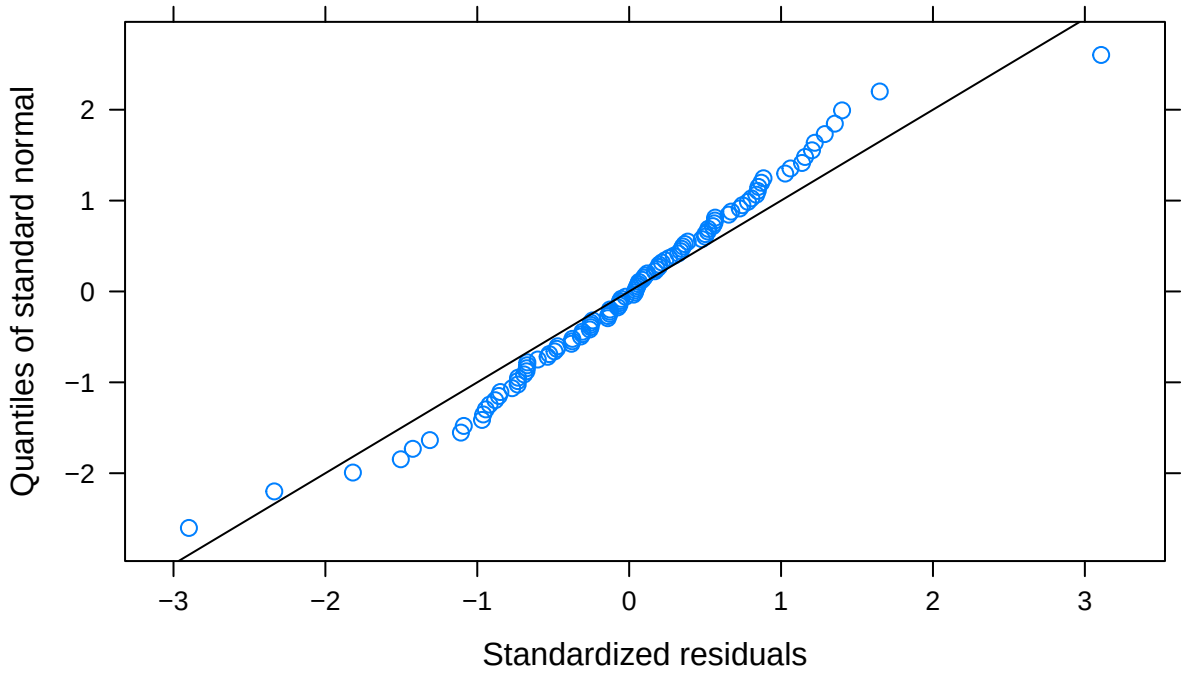
```
###   PA and SS pearson residuals -- the PA residuals differ from those
##   obtained from PROC MIXED apparently due to different standardization
residuals(dental.lme.b.reml, level=0:1, type="pearson")[1:5, ]
```

```
##        fixed         id
## 1 -0.3138325  1.2217076
## 2 -3.2543064 -1.5034436
## 3 -2.4424354 -0.4762498
## 4 -1.6305644  0.5509440
## 5 -0.3138325 -0.1397759
```

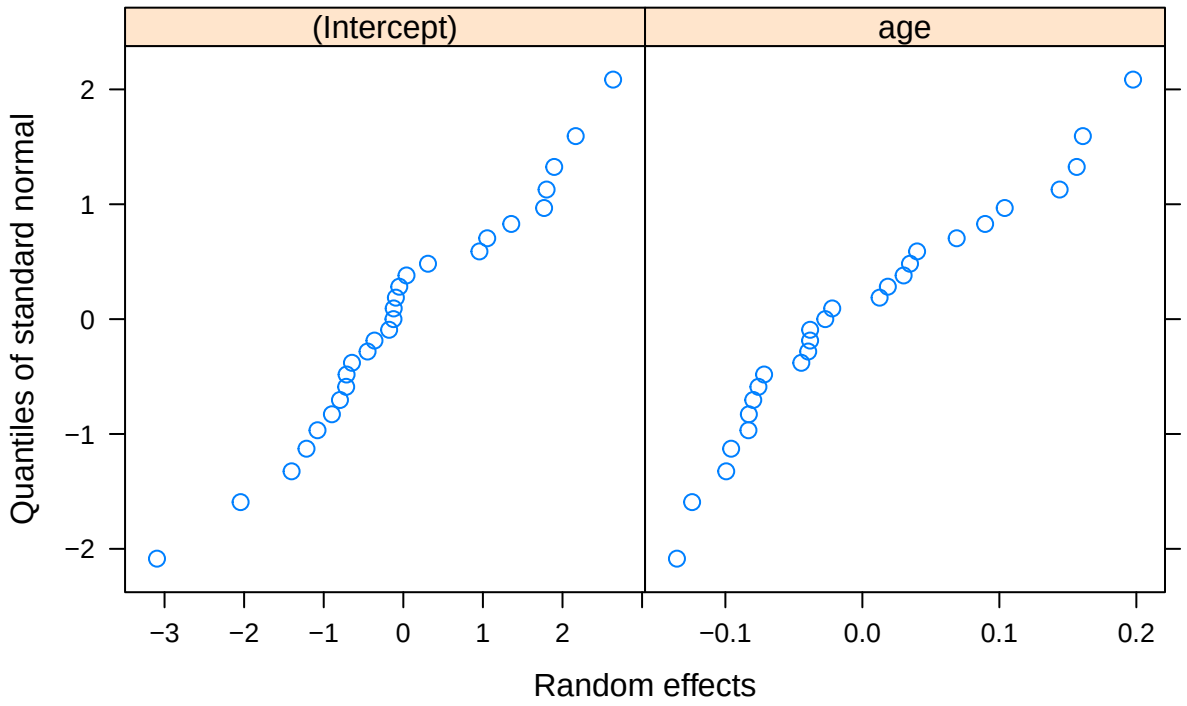
```
##  Plot SS residuals vs. predicted values

plot(dental.lme.b.reml, resid(. , type="p",level=1) ~ fitted(.,level=1) )
```



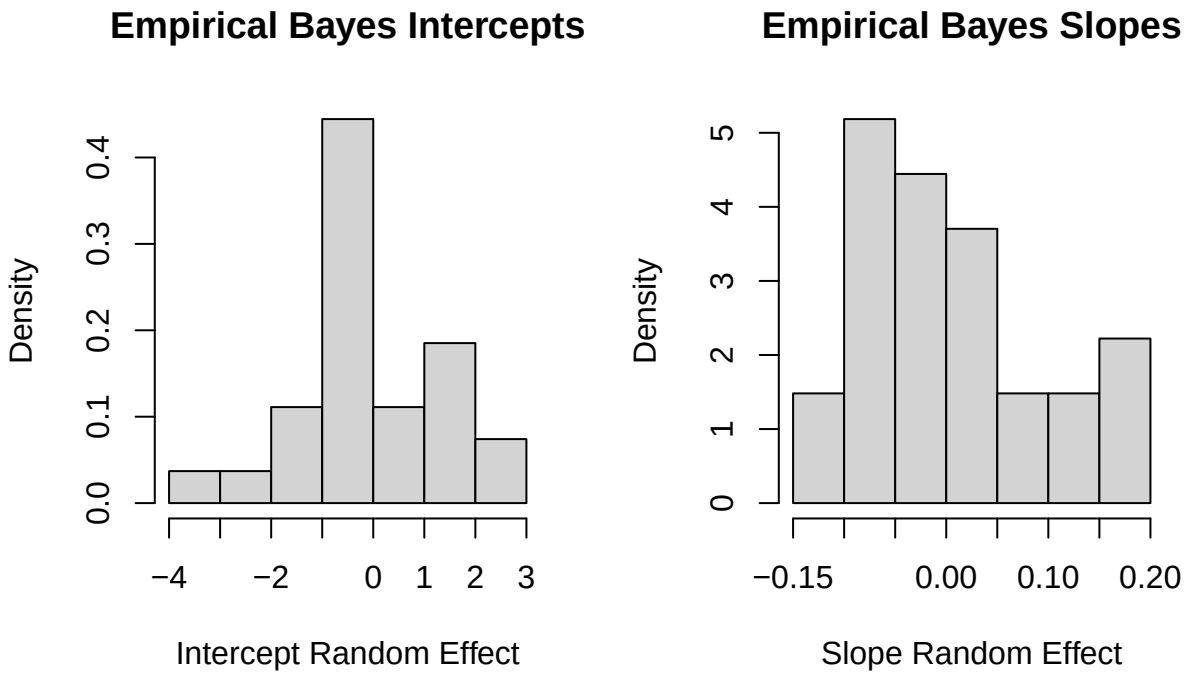

```
#  QQ plot of SS residuals
qqnorm(dental.lme.b.reml, ~ resid(. , type="p",level=1),abline=c(0,1))
```




```
#  One can also make QQ plots and histograms of the bhat_i themselves
#  to assess the normality of the random effects, but remember that
#  these are "shrunken" so could be misleading.

qqnorm(dental.lme.b.reml, ~ ranef(.))
```




```
##  histograms of random effects
par(mfrow=c(1,2))
hist(b.b.reml[,1],xlab="Intercept Random Effect",main="Empirical Bayes Intercepts",freq=FALSE)
hist(b.b.reml[,2],xlab="Slope Random Effect",main="Empirical Bayes Slopes",freq=FALSE)
```



Fitting models using lme4:lmer

```
#  We fit only model (a), Common G matrix for both genders, default
#  diagonal wintin-child covariance matrix  R_i with same variance
#  sigma^2 for each gender.  We use ML as in the lme() and SAS programs.

#  The random effects structure is specified in parentheses -- here,
#  we allow for random intercept and slope that are correlated

dental.lmer.a <- lmer(distance ~ -1 + gender + age:gender + (1 + age | id),
                      REML=FALSE, data=dat.den)

summary(dental.lmer.a)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: distance ~ -1 + gender + age:gender + (1 + age | id)
##    Data: dat.den
##
##      AIC      BIC   logLik deviance df.resid
##    443.8    465.3   -213.9    427.8      100
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3361 -0.4154  0.0104  0.4917  3.8582
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 4.55642  2.1346
##           age         0.02376  0.1541   -0.60
##  Residual             1.71622  1.3100
## Number of obs: 108, groups:  id, 27
##
## Fixed effects:
##                 Estimate Std. Error t value
## genderGirls     17.37273    1.18201  14.698
## genderBoys      16.34062    0.98007  16.673
## genderGirls:age  0.47955    0.09980   4.805
## genderBoys:age   0.78438    0.08275   9.479
##
## Correlation of Fixed Effects:
##             gndrGr gndrBy gndrG:
## genderBoys   0.000
## gendrGrls:g -0.880  0.000
## genderBys:g  0.000 -0.880  0.000
```

```
## sebeta.model.a
##     gender0     gender1 gender0:age gender1:age
##  1.18202362  0.98008221  0.09980390  0.08275303

beta.lmer.a <- fixef(dental.lmer.a)
beta.lmer.a
```

```
##     genderGirls      genderBoys genderGirls:age  genderBoys:age
##      17.3727273      16.3406250       0.4795455       0.7843750
```

```r
b.lmer.a <- ranef(dental.lmer.a)
sigma2.lmer.a <- sigma(dental.lmer.a)^2

#  It is pretty unwieldy to extract the covariance matrix D of the
#  random effects.  We can look at the variances and correlation with
vc.a <- VarCorr(dental.lmer.a)

print(vc.a,comp="Variance")
```

```
##  Groups   Name        Variance Corr
##  id       (Intercept) 4.556417
##           age         0.023758 -0.602
##  Residual             1.716221
```

```r
#  All the covariance matrix stuff can be put in a data frame, from
#  which it can be extracted to form the matrices G, R_i, and V_i
vc.da <- as.data.frame(vc.a, order="lower.tri")
vc.da
```

```
##        grp        var1 var2        vcov      sdcor
## 1       id (Intercept) <NA>  4.55641657  2.1345764
## 2       id (Intercept)  age -0.19822663 -0.6024875
## 3       id         age <NA>  0.02375771  0.1541353
## 4 Residual        <NA> <NA>  1.71622105  1.3100462
```

```r
G.lmer.a <- matrix(c(vc.da[1,4],vc.da[2,4],vc.da[2,4],vc.da[3,4]),2,2,
                   byrow=TRUE)
G.lmer.a
```

```
##            [,1]        [,2]
## [1,]  4.5564166 -0.19822663
## [2,] -0.1982266  0.02375771
```