ORIGIN := 0

Mixed Model ANCOVA Designs

ANCOVA models seek to relate a continuous dependent variable (Y) with a continuous covariate (X) plus one or more additional factors (T) with discrete levels. A mixed model ANCOVA typically assigns additional random coefficients in a linear model to reflect covariance that occurs between related observations (termed pseudoreplication). Studies of this kind are useful in longitudinal studies where the same individuals are measured repeatedly through time, such as in growth studies. Individuals, however, are often chosen from very large populations and serve in the study mostly as blocking factors. Each individual j has a growth curve that may be modeled as a linear relationship $Y_i \sim X_i$ for observations i within each j. The whole ANCOVA is noted Y ~ X*T (with interaction) or Y ~ X+T (no interaction) using the standard model formulas employed by R. The random coefficients are given the notation b_j or b_{0j} for intercept and b_{1j} for slope reflecting the potentially unique behavior of the individuals j. The models are fit using using Restricted Maximum Likelihood (REML) or Maximum Liklihood (ML) criteria employing iterative calculations (Pinheiro & Bates 2000 *Mixed-Effects Models in S and S-PLUS* (PB). A major advantage of these calculations is accommodation of unbalanced designs. Following one or more fits, several tests are possible using an extension of standard GLM format employing "full" (FM) and "reduced" or "restricted" (RM) models using likelihood ratios instead of F as the test statistic. The following example is from PB and is extensively utilized by them to display several aspects of software options, graphics, and analysis strategy. Reported here is a more concise synthesis of the basic approach to testing focusing on the use of formulas in the function lme(). Graphic analysis and appraisal of fits are not covered here.

```
#ORTHODONT EXAMPLE FROM Pinheiro & Bates
setwd("c:/DATA/Models")
library(nlme)
O=read.table("Orthodont.txt")
O
```

^ Note: the dataset Orthodont is available in the library of package {nlme} that must be downloaded from a CRAN website. In this library, the datafile is in "grouped data" format, as explained in PB, offering advantages when using the plotting functions in {nlme}. However, "group data" is not necessary and, to a degree obscures the logic of formulas in mixed-models. In this example, the datafile is a standard dataframe downloaded from a text file.

```
# SLOPES AND INTERCEPTS OF REGRESSION LINES FOR FEMALES:
Of=O[O$Sex=="Female",]  # NOTE: ',' needed here!
Of
# distance = dependent variable
# age = covariate
```

^ in ANCOVA, a dependent continuous variable, here distance, is related to a continuous "covariate" (also termed "concomitant") variable, here age. In addition, one or more fixed factors identify tests for difference of means of interest. Here, interest centers on differences in regression lines in the fixed factor Sex between "Male" and "Female" levels. Mixed-models also include one or more random factors where primary interest is control for variance. Here, the Subject random factor defines groups (blocks) within which each regression line is constructed. Subjects may vary in slope, intercept, or both, depending on the specific mixed linear model constructed, as below. To start, following PB, fixed regression lines are fit for each individual female separately. Work with males would be handled in the same way, but not here.

```
> O
    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
7       23.0  12     M02   Male
8       26.5  14     M02   Male
9       23.0   8     M03   Male
10      22.5  10     M03   Male
11      24.0  12     M03   Male
12      27.5  14     M03   Male
13      25.5   8     M04   Male
14      27.5  10     M04   Male
15      26.5  12     M04   Male
...
98      21.0  10     F09 Female
99      22.0  12     F09 Female
100     21.5  14     F09 Female
101     16.5   8     F10 Female
102     19.0  10     F10 Female
103     19.0  12     F10 Female
104     19.5  14     F10 Female
105     24.5   8     F11 Female
106     25.0  10     F11 Female
107     28.0  12     F11 Female
108     28.0  14     F11 Female
```

Fitting a regression line for each level of the Random factor:

```
# FITTING REGRESSIONS FOR EACH Subject USING MEAN CENTERED COVARIATE:
LML=lmList(distance~I(age-mean(age))|Subject,data=Of)
summary(LML)
```

^ Note: mean centering the covariate is the usual method for removing covariance between slope and intercept in fitted linear models, indicated below as age_c.

Model:

$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$

$Y \sim X \mid G$

distance ~ age_c | Subject

where: i j = index for objects i within each level j of the random factor

- Y_{ij} = dependent variable
- X_{ij} = covariate variable
- β_{0j} = intercept coefficient for each level j
- β_{1j} = slope coefficient for each level j
- ε_{ij} = error ~ $N(0,\sigma^2)$

Group (G) factor "Subject" in dataset identifies each female (F01, F02, etc.). These are the levels j within Subject. Regressions are found using multiple measurements i within each j.

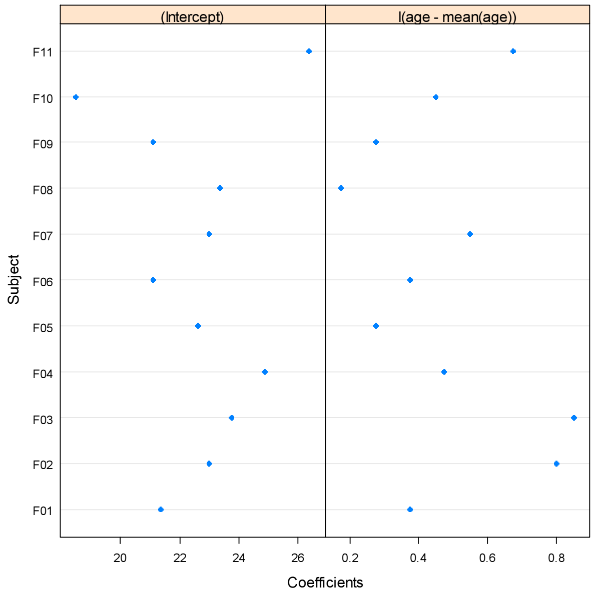
Function LmList() in {nlme} provides very efficient access to estimated coeffficients, and confidence intervals of the coefficients for each regression line. > {nlme} also implements diagnostic trellis graphics in R for simultaneous display of these lines.

```
> summary(LML)
Call:
  Model: distance ~ I(age - mean(age)) | Subject
   Data: Of

Coefficients:
   (Intercept)
    Estimate Std. Error  t value Pr(>|t|)
F01   21.375  0.3341373 63.97071        0
F02   23.000  0.3341373 68.83398        0
F03   23.750  0.3341373 71.07857        0
F04   24.875  0.3341373 74.44545        0
F05   22.625  0.3341373 67.71169        0
F06   21.125  0.3341373 63.22252        0
F07   23.000  0.3341373 68.83398        0
F08   23.375  0.3341373 69.95627        0
F09   21.125  0.3341373 63.22252        0
F10   18.500  0.3341373 55.36646        0
F11   26.375  0.3341373 78.93462        0
   I(age - mean(age))
    Estimate Std. Error  t value     Pr(>|t|)
F01    0.375  0.1494307 2.509524 1.994762e-02
F02    0.800  0.1494307 5.353651 2.247359e-05
F03    0.850  0.1494307 5.688254 1.013337e-05
F04    0.475  0.1494307 3.178730 4.344240e-03
F05    0.275  0.1494307 1.840317 7.925195e-02
F06    0.375  0.1494307 2.509524 1.994762e-02
F07    0.550  0.1494307 3.680635 1.310008e-03
F08    0.175  0.1494307 1.171111 2.540828e-01
F09    0.275  0.1494307 1.840317 7.925195e-02
F10    0.450  0.1494307 3.011429 6.422160e-03
F11    0.675  0.1494307 4.517143 1.705794e-04

Residual standard error: 0.6682746 on 22 degrees of freedom
```

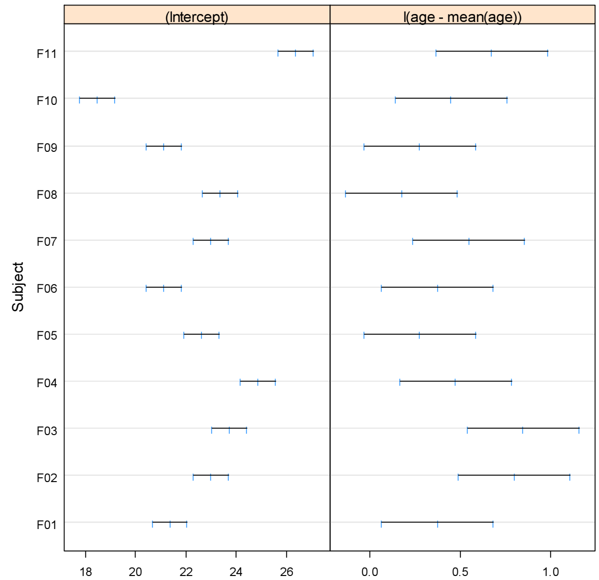
```
coef(LML)          # regression coefficients for each Subject
plot(coef(LML))    # trellis plot of coefficients

intervals(LML,0.95)  # 95% confidence intervals for coefficients
plot(intervals(LML)) # trellis plot of confidence intervals
```

< similar to other R functions for linear models, important information is extracted from the grouped model object LML made by the function call to LmList(). In addition plot() now contains specific methods to handle grouped data objects directly.

> plot(coef(LML))



> plot(intervals(LML))



Fitting a Linear Mixed Model with random intercept and random slope:

Model:

$Y_{ij} = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij} + \varepsilon_{ij}$

$Y \sim X + (1 + X \mid G)$

distance ~ X + (1 + age_c | Subject)

In the random part of the formula, identified by (): 1 specifies b_{0j} the random coefficient for intercept, and age_c (mean centered age) specifies b_{1j} the random coefficient for slope.

where: ij = index for objects i within each level j of the random factor

Y_{ij} = dependent variable

X_{ij} = covariate variable

β_0 = fixed intercept coefficient

b_{0j} = random intercept coefficient for each level of j

β_1 = common slope coefficient for all levels of j

b_{1j} = random slope coefficient for each level of j

$b_{0j} \sim N(0,\sigma_{b0}^2)$ and $b_{1j} \sim N(0,\sigma_{b1}^2)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$

b_{0j} and b_{1j} may be correlated but both are independent of ε_{ij}

```
# FITTING LINEAR MIXED MODEL WITH RANDOM INTERCEPT AND RANDOM SLOPE:
FME1=lme(distance~I(age-mean(age)),data=Of,random= ~1+I(age-mean(age))|Subject,method="REML")
summary(FME1)
FMe1=update(FME1,method="ML")
summary(FMe1)
```

^ The update() function only changes a part of the original specification in lme() for object FME1 and re-runs the fit. The result is put into FMe1 here. REML and ML are alternative estimation procedures in lme() often giving similar results. ML estimation is required when comparing models with different fixed factor structure. The output of anova() using models constructed by lme() will provide warnings about this.

AIC & BIC give *relative* indications of fit. Smaller numbers are preferred. >

Variance components (here as standard deviations) for the random factors and residual error. >

Estimates and marginal tests for the fixed coefficients >

```
> summary(FME1)
Linear mixed-effects model fit by REML
 Data: Of
       AIC      BIC    logLik
  149.4287 159.8547 -68.71435

Random effects:
 Formula: ~1 + I(age - mean(age)) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
                   StdDev    Corr
(Intercept)        2.0782266 (Intr)
I(age - mean(age)) 0.1609277 0.53
Residual           0.6682748

Fixed effects: distance ~ I(age - mean(age))
                       Value Std.Error DF  t-value p-value
(Intercept)        22.647727 0.6346562 32 35.68503       0
I(age - mean(age))  0.479545 0.0662140 32  7.24235       0
 Correlation:
                   (Intr)
I(age - mean(age)) 0.384

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.85438178 -0.46784876  0.06779756  0.42976662  1.59215811

Number of Observations: 44
Number of Groups: 11
```

^ The function lme() in package {nlme} provides a fit to the data using the "Restricted Maximum Liklihood" (REML) method, as detailed in PB. This method is default, and need not be specified in the command line. REML produces larger variance components than the alternative "Maximum Liklihood" (ML) method, and is generally preferred when used alone in this manner. PB p. 18 caution that when *comparing* REML models, identical fixed-effects structure must be maintained including the same contrasts. The default in R is "treatments" contrasts. MJ Crowley in *The R Book* suggests, in addition, the use of ML instead of REML in making all comparisons.

In general, REML and ML estimates of fixed effects and test will be similar. Variance components with ML will be smaller than those with REML. >

```
> summary(FMe1)
Linear mixed-effects model fit by maximum likelihood
 Data: Of
       AIC      BIC    logLik
  146.5093 157.2144 -67.25463

Random effects:
 Formula: ~1 + I(age - mean(age)) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
                   StdDev    Corr
(Intercept)        1.9789501 (Intr)
I(age - mean(age)) 0.1466746 0.556
Residual           0.6682746

Fixed effects: distance ~ I(age - mean(age))
                       Value Std.Error DF  t-value p-value
(Intercept)        22.647727 0.6193615 32 36.56625       0
I(age - mean(age))  0.479545 0.0646183 32  7.42120       0
 Correlation:
                   (Intr)
I(age - mean(age)) 0.384

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.93507292 -0.47348302  0.08334365  0.45964949  1.60921714

Number of Observations: 44
Number of Groups: 11
```

Fitting a Linear Mixed Model with random intercept but common slope:

Model:

$Y_{ij} = (\beta_0 + b_{0j}) + \beta_1 X_{ij} + \varepsilon_{ij}$

$Y \sim X + (1 \mid G)$

$\text{distance} \sim \text{age}_c + (1 \mid \text{Subject})$

where: ij = index for objects i within each level j of the random factor

Y_{ij} = dependent variable

X_{ij} = covariate variable

β_0 = fixed intercept coefficient

b_{0j} = random intercept coefficient for each level of j

β_1 = common slope coefficient for all levels of j

$b_j \sim N(0,\sigma_b^2)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$

```
# FITTING LINEAR MIXED MODEL WITH RANDOM INTERCEPT BUT COMMON SLOPE:
RME1=lme(distance~I(age-mean(age)),data=Of,random= ~1|Subject,method="REML")
summary(RME1)
RMe1=update(RME1,method="ML")
summary(RMe1)
```

```
> summary(RME1)
Linear mixed-effects model fit by REML
 Data: Of
       AIC     BIC    logLik
  149.2183 156.169 -70.60916

Random effects:
 Formula: ~1 | Subject
        (Intercept)  Residual
StdDev:     2.06847 0.7800331

Fixed effects: distance ~ I(age - mean(age))
                       Value Std.Error DF t-value p-value
(Intercept)        22.647727 0.6346568 32 35.6850       0
I(age - mean(age))  0.479545 0.0525898 32  9.1186       0
 Correlation:
                   (Intr)
I(age - mean(age)) 0

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
```

```
> summary(RMe1)
Linear mixed-effects model fit by maximum likelihood
 Data: Of
       AIC      BIC   logLik
  146.0304 153.1672 -69.0152

Random effects:
 Formula: ~1 | Subject
        (Intercept)  Residual
StdDev:    1.969870 0.7681235

Fixed effects: distance ~ I(age - mean(age))
                       Value Std.Error DF  t-value p-value
(Intercept)        22.647727 0.6193616 32 36.56624       0
I(age - mean(age))  0.479545 0.0530056 32  9.04708       0
 Correlation:
                   (Intr)
I(age - mean(age)) 0

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-2.3056159 -0.7192392  0.1763611  0.4257994  1.6689361

Number of Observations: 44
Number of Groups: 11
```

Testing adequacy of random slope:

Assumptions:

Models are as specified above. The test is valid only if one model - the Reduced Model (RM) fits within specification of the other model - the Full Model (FM). Since the FM contains more constraint it is by definition less parsimonious than the RM.

Full Model:

$Y_{ij} = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij} + \varepsilon_{ij}$ < model includes random coefficients for both intercept and slope.

$Y \sim X + (1 + X \mid G)$

Reduced Model:

$Y_{ij} = (\beta_0 + b_j) + \beta_1 X_{ij} + \varepsilon_{ij}$ < model includes random coefficient for intercept alone.

$Y \sim X + (1 \mid G)$

Hypotheses:

H_0: The more parsimonious Reduced Model is an adequate description of the relationship between X_{ij} and Y_{ij}

H_1: Full Model is a significantly better fit.

Decision Rule:

$\alpha := 0.05$ < set as desired

IF $P < \alpha$ THEN REJECT H_0 OTHERWISE ACCEPT H_0

```
# TESTING RELATIVE ADEQUACY OF RANDOM SLOPE:
anova(RME1,FME1)
anova(RMe1,FMe1)
```

```
> anova(RME1,FME1)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
RME1     1  4 149.2183 156.1690 -70.60916
FME1     2  6 149.4287 159.8547 -68.71435 1 vs 2 3.789622  0.1503

> anova(RMe1,FMe1)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
RMe1     1  4 146.0304 153.1671 -69.01520
FMe1     2  6 146.5093 157.2144 -67.25463 1 vs 2 3.521127  0.1719
```

^ both REML & ML version of the test fail to reject H_0: the more parsimonious model with common slopes for all Subjects j is an adequate description of fit, and therefore preferred.

```
# PREDICTIONS OF RESPONSE, COEFFICIENTS AND RANDOM EFFECTS
# FOR MODEL WITH RANDOM INTERCEPT BUT COMMON SLOPE:

predict(RME1)
coef(RME1)
ranef(RME1)  # BLUPs

augPred(RME1,~age)
plot(augPred(RME1,~age),aspect="fill", grid="T")
```

< features of model object RME1 may now be extracted.

predicted values for each regression line j (Subject) >

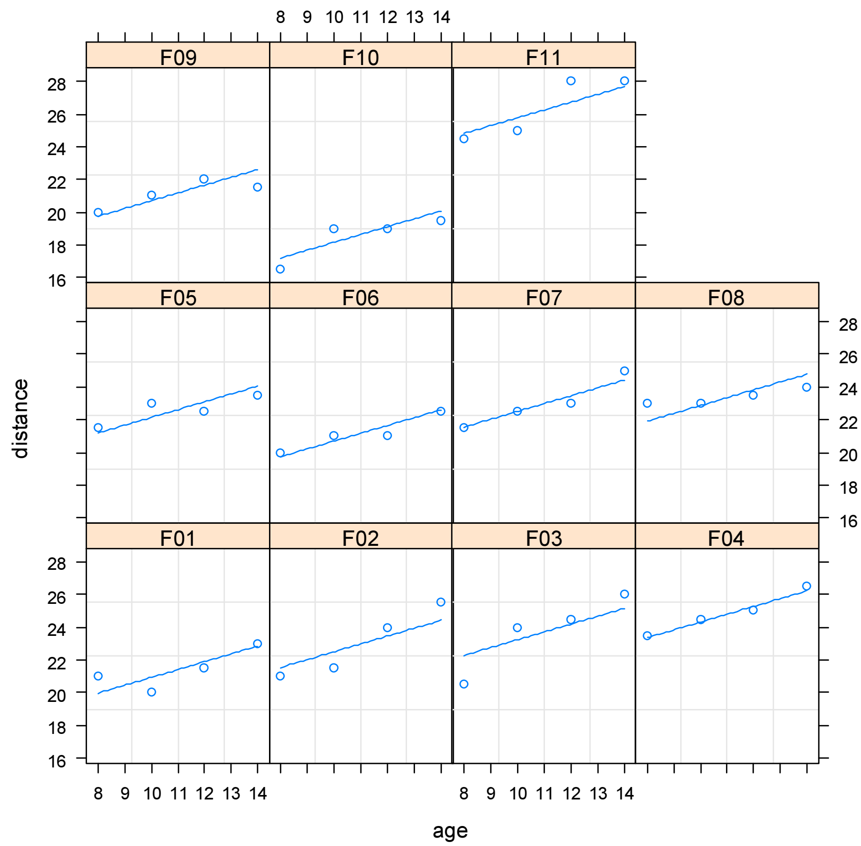
```
> predict(RME1)
     F01      F01      F01      F01      F02      F02      F02      F02 
19.98006 20.93915 21.89824 22.85733 21.54927 22.50836 23.46745 24.42654 
     F03      F03      F03      F03      F04      F04      F04      F04 
22.27352 23.23261 24.19170 25.15079 23.35990 24.31899 25.27808 26.23717 
     F05      F05      F05      F05      F06      F06      F06      F06 
21.18714 22.14623 23.10533 24.06442 19.73864 20.69773 21.65682 22.61591 
     F07      F07      F07      F07      F08      F08      F08      F08 
21.54927 22.50836 23.46745 24.42654 21.91140 22.87049 23.82958 24.78867 
     F09      F09      F09      F09      F10      F10      F10      F10 
19.73864 20.69773 21.65682 22.61591 17.20376 18.16285 19.12194 20.08103 
     F11      F11      F11      F11 
24.80840 25.76749 26.72658 27.68567 
```

```
> coef(RME1)
    (Intercept) I(age - mean(age))
F01    21.41869          0.4795455
F02    22.98791          0.4795455
F03    23.71216          0.4795455
F04    24.79853          0.4795455
F05    22.62578          0.4795455
F06    21.17728          0.4795455
F07    22.98791          0.4795455
F08    23.35003          0.4795455
F09    21.17728          0.4795455
F10    18.64240          0.4795455
F11    26.24704          0.4795455
```

< fixed estimates of coefficients for each regression line j

random predictions of unique effect of each individual j. These are > the "Best Unbiased Linear Predictors" (BLUPs)

```
> ranef(RME1) # BLUPs
    (Intercept)
F01 -1.22903236
F02  0.34017860
F03  1.06442981
F04  2.15080662
F05 -0.02194701
F06 -1.47044943
F07  0.34017860
F08  0.70230420
F09 -1.47044943
F10 -4.00532866
F11  3.59930904
```



Testing adequacy of random intercept:

Assumptions:

Models are as specified above. The test is valid only if one model - the Reduced Model (RM) fits within specification of the other model - the Full Model (FM). Since the FM contains more constraint it is by definition less parsimonious than the RM.

Full Model:

$Y_{ij} = (\beta_0 + b_j) + \beta_1 X_{ij} + \varepsilon_{ij}$ **< model with random intercept coefficient**

Y ~ X + (1 | G)

Reduced Model:

$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \varepsilon_{ij}$ **< model with only fixed coefficients**

Y ~ X

Hypotheses:

H_0: The more parsimonious Reduced Model is a adequate description of the relationship between X_{ij} and Y_{ij}

H_1: Full Model is a significantly better fit.

Decision Rule:

$\alpha := 0.05$ < set as desired

IF $P < \alpha$ THEN REJECT H_0 OTHERWISE ACCEPT H_0

```
# FITTING LINEAR MIXED MODEL WITH RANDOM INTERCEPT BUT COMMON SLOPE:
FME2=lme(distance~I(age-mean(age)),data=Of,random= ~1|Subject,method="REML")
summary(FME2)
FMe2=update(FME2,method="ML")
summary(FMe2)
# FITTING LINEAR FIXED MODE WITH COMMON INTERCEPT AND COMMON SLOPE:
RM2=lm(distance~I(age-mean(age)),data=Of)
summary(RM2)
#FITTING LINEAR MODEL USING gls()
RM2a=gls(distance~I(age-mean(age)),method="REML",data=Of)
RM2b=gls(distance~I(age-mean(age)),method="ML",data=Of)
summary(RM2)
# TESTING RELATIVE ADEQUACY OF RANDOM INTERCEPT:
anova(RM2a,FME2)
anova(RM2b,FMe2)
```

Model comparisons using lm() vs lme() often do not work, presumably because anova.lm or anova.lme expect to see corresponding data structures for both models.

As a general solution , it's best to fit linear models using gls() in {nlme}, but here you should also specify a method of estimation - either REML or ML.

```
> anova(RM2a,FME2)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
RM2a     1  3 199.2166 204.4296 -96.60830
FME2     2  4 149.2183 156.1690 -70.60916 1 vs 2 51.99827  <.0001
> anova(RM2b,FMe2)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
RM2b     1  3 196.7557 202.1082 -95.37782
FMe2     2  4 146.0304 153.1671 -69.01520 1 vs 2 52.72525  <.0001
```

^ Both indicate a preference for FM

More Comprehensive Tests involving both age (covariate) and Sex (additional factor):

The following tests, following discussion in PB, looks at similar Full versus Reduced Model tests as above. However, now the factor Sex with two levels enters into the Fixed Component of each model allowing one to expand analysis to an ANCOVA analysis of the Fixed Component, while at the same time using Subject (Group factor) to enter also within the Random Component of the model.

Testing adequacy of random intercept and random slope versus fixed linear model:

Assumptions:

Same as above. GLM Reduced model vs Full model approach.

Full Model:

$Y_{ij} = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij} + \beta_2\tau_j + \beta_3 X_{ij}\tau_j + \varepsilon_{ij}$
Y ~ X*T +(1 + X | G)
distance ~ age_c*Sex + (1 + age_c | Subject)

< model with fixed factor T (Sex) crossed with covariate X (age_c) plus random coefficients for intercept and slope.

where: ij = index for objects i within each level j of the random factor
Y_{ij} = dependent variable
X_{ij} = covariate variable
β_0 = fixed intercept coefficient
b_{0j} = random intercept coefficient for each level of j
β_1 = common slope coefficient for all levels of j
b_{1j} = random slope coefficient for each level of j
β_2 = fixed regression coefficient for fixed factor τ_j
β_3 = fixed regression coefficient for fixed interaction $X_{ij}\tau_j$
$b_{0j} \sim N(0,\sigma_{b0}^2)$ and $b_{1j} \sim N(0,\sigma_{b1}^2)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$
b_{0j} and b_{1j} may be correlated but both are independent of ε_{ij}

Reduced Model:

$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2\tau_j + \beta_3 X_{ij}\tau_j + \varepsilon_{ij}$
Y ~ X*T
distance ~ age_c*Sex

^ fixed factor model

Hypotheses:

H_0: The more parsimonious Reduced Model is a adequate description of the relationship between X_{ij} and Y_{ij}

H_1: Full Model is a significantly better fit.

Decision Rule:

$\alpha := 0.05$ < set as desired

IF $P < \alpha$ THEN REJECT H_0 OTHERWISE ACCEPT H_0

```
#MORE COMPREHENSIVE TESTS INVOLVING AGE SEX AND SUBJECT:
# SWITCHING TO HELMERT CONTRASTS TO COMPARE RESULTS WITH PB p.148-149:
options(contrasts = c("contr.helmert", "contr.poly"))
#TESTING MIXED VERSUS FIXED MODEL:
#FULL MODEL: Sex*age + (1+age|Subject) USING REML
FME3=lme(distance~Sex*I(age-mean(age)),data=O,random= ~1+I(age-mean(age))|Subject,method="REML")
#REDUCED MODEL: Sex*age USING REML
RM3=gls(distance~Sex*I(age-mean(age)),method="REML",data=O)
anova(FME3,RM3)
```

values confirmed for REML PB p. 155 >

```
> anova(FME3,RM3)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
FME3     1  8 451.3543 472.5094 -217.6771
RM3      2  5 496.3317 509.5537 -243.1659 1 vs 2 50.97746  <.0001
```

summary(FME3)

values confirmed PB p. 151 >
with Helmert contrasts

```
> summary(FME3)
Linear mixed-effects model fit by REML
 Data: O
       AIC      BIC    logLik
  451.3543 472.5094 -217.6771

Random effects:
 Formula: ~1 + I(age - mean(age)) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
                   StdDev    Corr
(Intercept)        1.8303267 (Intr)
I(age - mean(age)) 0.1803454 0.206
Residual           1.3100397

Fixed effects: distance ~ Sex * I(age - mean(age))
                           Value Std.Error DF  t-value p-value
(Intercept)            23.808239 0.3807084 79 62.53668  0.0000
Sex1                    1.160511 0.3807084 25  3.04829  0.0054
I(age - mean(age))      0.631960 0.0673677 79  9.38076  0.0000
Sex1:I(age - mean(age))  0.152415 0.0673677 79  2.26243  0.0264
 Correlation:
                        (Intr) Sex1   I(-m()
Sex1                    -0.185
I(age - mean(age))       0.102 -0.019
Sex1:I(age - mean(age)) -0.019  0.102 -0.185

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-3.168078495 -0.385939137  0.007103929  0.445154689  3.849463224

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

Testing fixed sex*age interaction with random intercept and random slope:

Assumptions:

Same as above. GLM Reduced model vs Full model approach.

Full Model:

$Y_{ij} = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij} + \beta_2\tau_j + \beta_3 X_{ij}\tau_j + \varepsilon_{ij}$ < mixed cross factor model with random intercept and slope

Y ~ X*T + (1 + X | G)

distance ~ age_c*Sex + (1 + age_c | Subject)

where: ij = index for objects i within each level j of the random factor

Y_{ij} = dependent variable

X_{ij} = covariate variable

β_0 = fixed intercept coefficient

b_{0j} = random intercept coefficient for each level of j

β_1 = common slope coefficient for all levels of j

b_{1j} = random slope coefficient for each level of j

β_2 = fixed regression coefficient for fixed factor τ_j

β_3 = fixed regression coefficient for fixed interaction $X_{ij}\tau_j$

$b_{0j} \sim N(0,\sigma_{b0}^2)$ and $b_{1j} \sim N(0,\sigma_{b1}^2)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$

b_{0j} and b_{1j} may be correlated but both are independent of ε_{ij}

Reduced Model:

$Y_{ij} = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij} + \beta_2\tau_j + \varepsilon_{ij}$

Y ~ X+T + (1 + X | G)

distance ~ age_c+Sex + (1 + age_c | Subject)

^ fixed model without interaction
but random intercept and slope

Hypotheses:

H_0: The more parsimonious Reduced Model is a adequate description of the relationship between X_{ij} and Y_{ij}

H_1: Full Model is a significantly better fit.

Decision Rule:

$\alpha := 0.05$ < set as desired

IF P < α THEN REJECT H_0 OTHERWISE ACCEPT H_0

```
#TESTING FIXED INTERACTION:
#FULL MODEL: Sex*age + (1+age|Subject) USING ML
FME4=update(FME3,method="ML")
#REDUCED MODEL: Sex+age + (1+age|Subject) USING ML
RME4=lme(distance~Sex+I(age-mean(age)),data=O,random= ~1+I(age-mean(age))|Subject,method="ML")
anova(FME4,RME4)
```

```
> anova(FME4,RME4)
     Model df      AIC      BIC    logLik   Test L.Ratio p-value
FME4     1  8 443.8060 465.2630 -213.9030
RME4     2  7 446.8352 465.6101 -216.4176 1 vs 2 5.02921  0.0249
```

^ Full Model containing interaction is slightly favored

Testing fixed sex*age interaction with random intercept and random slope correcting for Heteroscedasticity:

```
#TESTING HETEROSCEDASTICITY:
#FULL MODEL: Sex*age+(1+age|Subject) varIdent(~1|Sex) USING ML
FME5=lme(distance~Sex*I(age-mean(age)),data=O,random= ~1+I(age-mean(age))|Subject,method="REML",
        weights=varIdent(form=~1|Sex))
#REDUCED MODEL: Sex*age+(1+age|Subject) USING ML
RME5=lme(distance~Sex*I(age-mean(age)),data=O,random=
~1+I(age-mean(age))|Subject,method="REML")
anova(FME5,RME5)
```

```
> anova(FME5,RME5)
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
FME5     1  9 432.2951 456.0946 -207.1475
RME5     2  8 451.3543 472.5094 -217.6771 1 vs 2 21.05918  <.0001
```

^ very close, but not exactly the same as values in PB p. 178.

Full model including heteroscedasticity is strongly favored.

^ PB provide diagnostic plots for residuals from lme() model fits indicating differences between Males and Females in scatter. In {nlme} they have implemented functions to specify particular covariance structures within linear mixed models designed, in part, to handle heteroscesdasticity. Their use of weights and varIdent() is repeated here, but requires additional discussion. Chapter 5 in PB provide their introduction to the problem. Chapter 4 in Zuur et al. 2009, *Mixed Effects Models and Extensions in Ecology with R* provides a more readable introduction. See also *Statistical Models* Worksheet LMM 050.