Random and Mixed Effects Models - Part II

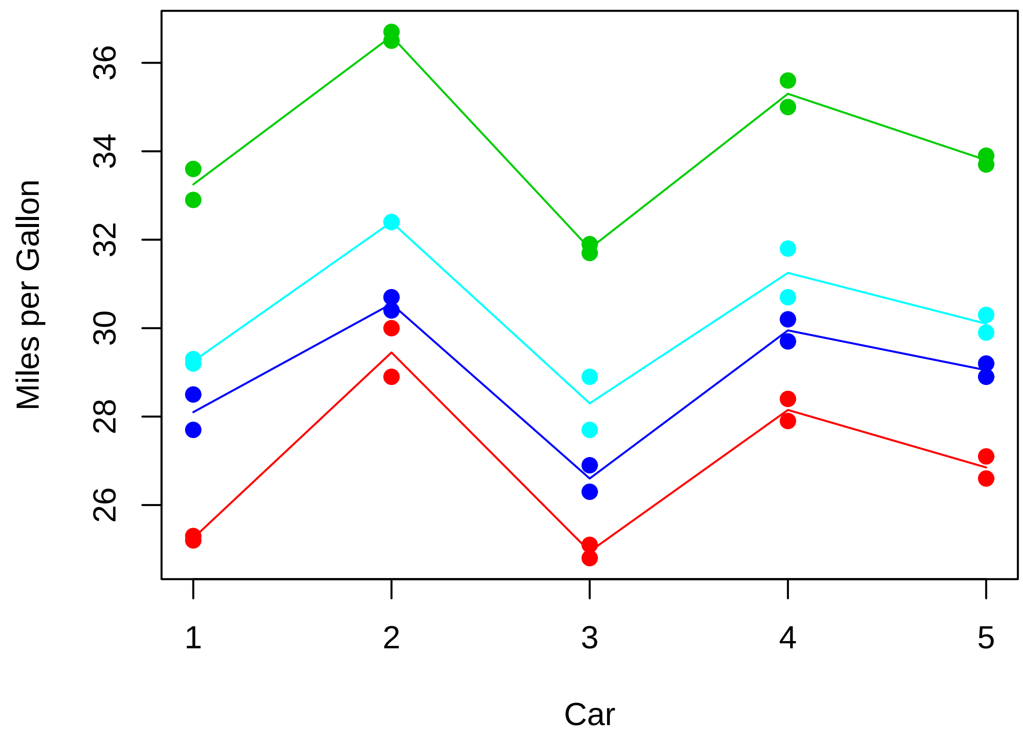
Statistics 149

Spring 2006

Two-Factor Random Effects Model

Example: Miles per Gallon (Neter, Kutner, Nachtsheim, & Wasserman, problem 24.15)

An automobile manufacturer studied the effects of differences between drivers (`Driver`) and cars (`Car`) on gasoline consumption. Four drivers and 5 cars of the same model were both selected at random. Each driver drove each car twice over a 40 mile test course and the miles per gallon (`MPG`) were recorded.



In this example, it seems reasonable to treat driver and car as random effects.

One possible model for this data is

$$
\begin{aligned}
y_{ijk} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \\
\alpha_i &\overset{iid}{\sim} N(0, \sigma^2_\alpha) \\
\beta_j &\overset{iid}{\sim} N(0, \sigma^2_\beta) \\
(\alpha\beta)_{ij} &\overset{iid}{\sim} N(0, \sigma^2_{\alpha\beta}) \\
\epsilon_{ijk} &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
$$

For what follows, lets assume a balance design, with a levels for factor A (`Driver`), b levels for factor B (`Car`), and m observations for each (i, j) combination.

(Of course you could have an unbalanced designed, however the following formulas get really ugly.)

For this design, the expected mean squares satisfy

Mean Square	df	$E[MS]$
MSA	$a-1$	$\sigma^2 + bm\sigma_\alpha^2 + m\sigma_{\alpha\beta}^2$
MSB	$b-1$	$\sigma^2 + am\sigma_\beta^2 + m\sigma_{\alpha\beta}^2$
MSAB	$(a-1)(b-1)$	$\sigma^2 + m\sigma_{\alpha\beta}^2$
MSE	$(m-1)ab$	σ^2

Based on these expected mean squares, we estimate the various variance components as follows

- σ^2:

$$\hat{\sigma}^2 = MSE$$

- $\sigma_{\alpha\beta}^2$:

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{MSAB - MSE}{m}$$

- σ_α^2:

$$\hat{\sigma}_\alpha^2 = \frac{MSA - MSAB}{bm}$$

- σ_β^2:

$$\hat{\sigma}_\beta^2 = \frac{MSB - MSAB}{am}$$

For the example, these happen to be

```
> mpg.rr <- lmer(MPG ~ 1 + (1|Driver) + (1|Car) + (1|Driver:Car) ,

> mpg.rr
Linear mixed-effects model fit by REML
Formula: MPG ~ 1 + (1 | Driver) + (1 | Car) + (1 | Driver:Car)
         Data: mpg
      AIC      BIC    logLik MLdeviance REMLdeviance
 94.77908 101.5346 -43.38954   89.67671     86.77908
Random effects:
 Groups     Name        Variance Std.Dev.
 Driver:Car (Intercept) 0.014063 0.11859
 Car        (Intercept) 2.934312 1.71298
 Driver     (Intercept) 9.322437 3.05327
 Residual               0.175750 0.41923
number of obs: 40, groups: Driver:Car, 20; Car, 5; Driver, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  30.0475     1.7096  17.576
```

So similarly to the single random effect case, we can use the ANOVA table to help us to estimate the variance components. However we need to know the structure of the expected mean squares. Different assumptions about which terms are to be included in the model and which factors are random and which are fixed, will lead to different mean square errors, and thus different estimates for the variance components.

The resulting structure also affects tests on the variance components.

If all of the factors are fixed, the three standard F tests are

$$F_A = \frac{MSA}{MSE} \qquad F_B = \frac{MSB}{MSE} \qquad F_{AB} = \frac{MSAB}{MSE}$$

For the test of the interaction $(H_0 : \sigma^2_{\alpha\beta})$,

$$F_{AB} = \frac{MSAB}{MSE}$$

since both mean squares have the same expectation under H_0.

For the example, this can be examined in **R** by

```
> mpg.rr <- lmer(MPG ~ 1 + (1|Driver) + (1|Car) + (1|Driver:Car),
      data=mpg)

> mpg.a.rr <- lmer(MPG ~ 1 + (1|Driver) + (1|Car), data=mpg)

> anova(mpg.rr, mpg.a.rr)
Data: mpg
Models:
mpg.a.rr: MPG ~ 1 + (1 | Driver) + (1 | Car)
mpg.rr: MPG ~ 1 + (1 | Driver) + (1 | Car) + (1 | Driver:Car)
         Df    AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
mpg.a.rr  3 92.863  97.929 -43.431
mpg.rr    4 94.779 101.535 -43.390 0.0836      1     0.7725
```

Note that the **R** approach using the `lme4` package is to use χ^2 based likelihood ratio tests instead of F tests. Asymptotically these will give the similar results.

However for the main effects, the tests aren't valid. For the test on factor A, under $H_0 : \sigma_\alpha^2$,

$$E[MSA] = \sigma^2 + m\sigma_{\alpha\beta}^2 \qquad E[MS] = \sigma^2$$

so the F used in the fixed effects case will not work here.

Instead

$$F_A = \frac{MSA}{MSAB}$$

will work as both mean squares have the same expectation under the null.

Similarly

$$F_B = \frac{MSB}{MSAB}$$

can be used to examine $H_0 : \sigma_\beta^2$

While these are valid test statistics, whether testing the corresponding hypotheses is usually questionable. There is usually no reason to test a main effect when an interaction containing that main effect is included in the model.

Two-Factor Mixed Model

As suggested before, it is possible to combine fixed and random factors in a model. For example, suppose that driver is a random effect and car is a fixed effect. This might occur if we are considering a single Zipcar location that only has 5 cars. In that case, the company might be interested in the individual cars and not some larger population of cars.

In this case we can model the data as

$$
\begin{aligned}
y_{ijk} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \\
\alpha_i &\overset{iid}{\sim} N(0, \sigma^2_\alpha) \\
\beta_j & \text{ fixed but unknown constants, subject to} \sum \beta_j = 0 \\
(\alpha\beta)_{ij} &\overset{iid}{\sim} N(0, \sigma^2_{\alpha\beta}) \\
\epsilon_{ijk} &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
$$

(Note the constraint on the β_j was chosen to make the expect mean square formula nicer.)

Usually when you include interactions involving fixed and random effects, the interactions are considered as random effects.

The expected mean squares for this model satisfy

Mean Square	df	$E[MS]$
MSA	$a-1$	$\sigma^2 + bm\sigma^2_\alpha + m\sigma^2_{\alpha\beta}$
MSB	$b-1$	$\sigma^2 + m\sigma^2_{\alpha\beta} + ma\frac{\sum \beta_j^2}{b-1}$
MSAB	$(a-1)(b-1)$	$\sigma^2 + m\sigma^2_{\alpha\beta}$
MSE	$(m-1)ab$	σ^2

Note that this table of mean squares doesn't match what you will see in other and software packages. Other books discuss a slightly different form of the mixed model, sometimes referred to as the restricted model.

In this case, they assume that

$$\sum_i (\alpha\beta)_{ij} = 0 \quad \text{for each } j$$

The unrestricted model was chosen as it matches with the function `lmer`. Also Dean and Vos argue that usually the unrestricted model makes for sense (what really is the correct way to set up the restrictions).

For a further discussion of the restricted form of the model, see Neter, Kutner, Nachtsheim, and Wasserman or Montgomery.

Under the unrestricted model, the estimated variance components are

- σ^2:

$$\hat{\sigma}^2 = MSE$$

- $\sigma^2_{\alpha\beta}$:

$$\hat{\sigma}^2_{\alpha\beta} = \frac{MSAB - MSE}{m}$$

- σ^2_α:

$$\hat{\sigma}^2_\alpha = \frac{MSA - MSAB}{bm}$$

```
> options(contrasts=c("contr.sum","contr.poly"))
> mpg.rf <- lmer(MPG ~ Car + (1|Driver) + (1|Car:Driver) ,
    data=mpg)
> mpg.rf
Linear mixed-effects model fit by REML
Formula: MPG ~ Car + (1 | Driver) + (1 | Car:Driver)
          Data: mpg
      AIC      BIC    logLik MLdeviance REMLdeviance
 86.69653 98.51868 -36.34826   66.10478     72.69653
Random effects:
 Groups     Name        Variance Std.Dev.
 Car:Driver (Intercept) 0.014063 0.11859
 Driver     (Intercept) 9.322437 3.05327
 Residual               0.175750 0.41923
number of obs: 40, groups: Car:Driver, 20; Driver, 4
```

```
Fixed effects:
            Estimate Std. Error  t value
(Intercept) 30.04750    1.52830  19.6607
Car1        -1.08500    0.14278  -7.5988
Car2         2.20250    0.14278  15.4253
Car3        -2.13500    0.14278 -14.9526
Car4         1.11500    0.14278   7.8090

Correlation of Fixed Effects:
     (Intr) Car1   Car2   Car3
Car1  0.000
Car2  0.000 -0.250
Car3  0.000 -0.250 -0.250
Car4  0.000 -0.250 -0.250 -0.250
```

Note that the constraint chosen on the fixed effects affects their estimates but not those of the random effects (at least under the default estimation scheme `method="REML"`). Switching to `contr.treatment` gives

```
> options(contrasts=c("contr.treatment","contr.poly"))
> mpg.rf2 <- lmer(MPG ~ Car + (1|Driver) + (1|Car:Driver),
     data=mpg)
> mpg.rf2
Linear mixed-effects model fit by REML
Formula: MPG ~ Car + (1 | Driver) + (1 | Car:Driver)
         Data: mpg
      AIC     BIC    logLik MLdeviance REMLdeviance
 83.47765 95.2998 -34.73883   66.10478     69.47765
Random effects:
 Groups     Name        Variance Std.Dev.
 Car:Driver (Intercept) 0.014062 0.11859
 Driver     (Intercept) 9.322436 3.05327
 Residual               0.175750 0.41923
number of obs: 40, groups: Car:Driver, 20; Driver, 4
```

```
Fixed effects:
            Estimate Std. Error t value
(Intercept) 28.96250    1.53496 18.8686
Car2         3.28750    0.22576 14.5618
Car3        -1.05000    0.22576 -4.6509
Car4         2.20000    0.22576  9.7447
Car5         0.98750    0.22576  4.3741

Correlation of Fixed Effects:
     (Intr) Car2   Car3   Car4
Car2 -0.074
Car3 -0.074  0.500
Car4 -0.074  0.500  0.500
Car5 -0.074  0.500  0.500  0.500
```

Again we can see whether there is an interaction effect as follows

```
> anova(mpg.a.rf, mpg.rf)
Data: mpg
Models:
mpg.a.rf: MPG ~ Car + (1 | Driver)
mpg.rf: MPG ~ Car + (1 | Driver) + (1 | Car:Driver)
         Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mpg.a.rf  6 84.780 94.913 -36.390
mpg.rf    7 86.697 98.519 -36.348 0.0836      1     0.7725
```

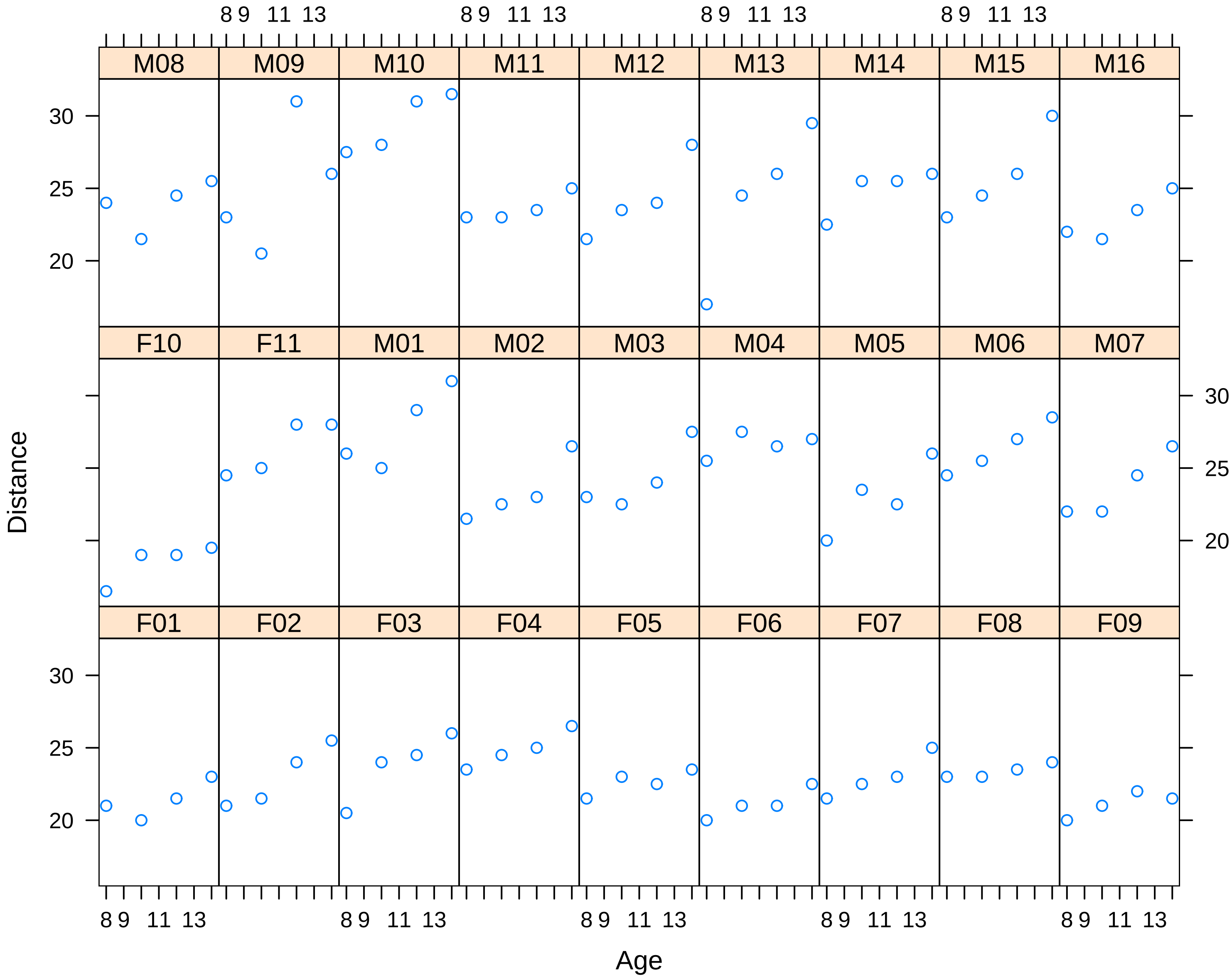
Actually this is the same test as with the random effects case, even though the base models are different. Be careful as this won't always happen in more complex cases.

Combining Continuous Predictors and Random Effects

In the next example, we want to combine a continuous predictor with a categorical factors, one fixed and one random. Of course we can have a wide range of different ways we can combine predictors in this sort of situation.

Example: This dataset, from the S-plus manual and collected by the Dental School of North Carolina, investigated the distance from the pituitary to the ptergomaxillary fissure (`Distance`). There were 27 subjects (`Subject`) (16 boys, 11 girls - `Sex`) each measured at 4 ages (8, 10, 12, 14 - `age`).

Of interest is the difference between the boys and girls, after accounting for the effects of age and subject to subject variability.


M08
M09
M10
M11
M12
M13
M14
M15
M16
F10
F11
M01
M02
M03
M04
M05
M06
M07
F01
F02
F03
F04
F05
F06
F07
F08
F09
Distance
Age
8 9 11 13
20
25
30

So we want to develop a model of the basic form

$$y = \mu + \underbrace{\alpha}_{\text{Sex effect}} + \underbrace{\beta}_{\text{Subject effect}} + \underbrace{\gamma t}_{\text{Age effect}} + \epsilon$$

Two possible models of this basic form are

- Common age effect:

$$y_{it} = \mu + \alpha(\mathsf{sex})_i + \beta_i + \gamma t + \epsilon_{it}$$

$$\beta_i \overset{iid}{\sim} N(0, \sigma_\beta^2)$$

$$\epsilon_{it} \overset{iid}{\sim} N(0, \sigma^2)$$

In this case, the effect of age is the same for each subject.

- Subject specific age effect:

$$
\begin{aligned}
y_{it} &= \mu + \alpha(\textsf{sex})_i + \beta_i + \gamma_i t + \epsilon_{it} \\
\beta_i &\overset{iid}{\sim} N(0, \sigma^2_\beta) \\
\gamma_i &\overset{iid}{\sim} N(\gamma, \sigma^2_\gamma) \\
\epsilon_{it} &\overset{iid}{\sim} N(0, \sigma^2)
\end{aligned}
$$

In this case, each subject in the trial has there own slope. As written, their is a fixed age effect γ and $\gamma_i - \gamma$ describe the the random slope deviations for each person.

For the first model, the fitted values are

```
> orthodont.mix <- lmer(distance ~ Sex + age + (1|Subject),
        data=Orthodont)
> orthodont.mix
Linear mixed-effects model fit by REML
Formula: distance ~ Sex + age + (1 | Subject)
          Data: Orthodont
      AIC     BIC    logLik MLdeviance REMLdeviance
 445.5125 456.241 -218.7563   434.8982     437.5125
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 3.2667   1.8074
 Residual             2.0495   1.4316
number of obs: 108, groups: Subject, 27

Fixed effects:
             Estimate Std. Error t value
(Intercept) 15.385690   0.895983 17.1718
SexMale      2.321023   0.761412  3.0483
age          0.660185   0.061606 10.7162
```

```
Correlation of Fixed Effects:
        (Intr) SexMal
SexMale -0.504
age     -0.756  0.000
```

The second model can be fit by

```
> orthodont.mix2 <- lmer(distance ~ Sex + age + (1|Subject)
                           + (age - 1| Subject), data=Orthodont)
```

To fit the model, we need to include the fixed effect for age to estimate γ. The term `(age - 1| Subject)` is written this way to guarantee that the random effects for `Subject` and the `age:Subject` are independent. If the `(age | Subject)` was used, there would be an interaction.

```
> orthodont.mix2
Linear mixed-effects model fit by REML
Formula: distance ~ Sex + age + (1 | Subject) + (age - 1 | Subject)
          Data: Orthodont
      AIC     BIC    logLik MLdeviance REMLdeviance
 446.6453 460.056 -218.3227   434.0781     436.6453
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 2.172948 1.47409
 Subject  age         0.009996 0.09998
 Residual             1.967260 1.40259
number of obs: 108, groups: Subject, 27; Subject, 27

Fixed effects:
             Estimate Std. Error t value
(Intercept) 15.568993   0.861517 18.0716
SexMale      2.011700   0.759759  2.6478
age          0.660185   0.063351 10.4211
```

```
Correlation of Fixed Effects:
        (Intr) SexMal
SexMale -0.523
age     -0.734  0.000
```

One potentially important question is whether the subject specific slopes are needed. Note that the estimate of σ_γ^2 is very small relative to the estimate σ^2, suggesting that a single slope is adequate.

This is supported by

```
> anova(orthodont.mix2,orthodont.mix)
Data: Orthodont
Models:
orthodont.mix: distance ~ Sex + age + (1 | Subject)
orthodont.mix2: distance ~ Sex + age + (1 | Subject) + (age - 1 | S
               Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
orthodont.mix   4 445.51 456.24 -218.76
orthodont.mix2  5 446.65 460.06 -218.32 0.8672      1     0.3517
```