library(nlme)

# The following data shows the results of tests carried over 6 rails. The response
# indicated the time needed for a an ultrasonic wave to travel the length of the rail
# It is a one way model, the factor being the rail used. However, the rails used represent just a sample of size 6 from a (much) larger population

<table>
<thead>
<tr>
<th>Rail</th>
<th>travel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>55</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
</tr>
<tr>
<td>8</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>85</td>
</tr>
<tr>
<td>10</td>
<td>92</td>
</tr>
<tr>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>96</td>
</tr>
<tr>
<td>13</td>
<td>49</td>
</tr>
<tr>
<td>14</td>
<td>51</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>16</td>
<td>80</td>
</tr>
<tr>
<td>17</td>
<td>85</td>
</tr>
<tr>
<td>18</td>
<td>83</td>
</tr>
</tbody>
</table>

fixed.fit<-lm(travel~1,data=Rail)

summary(fixed.fit)

Coefficients:

|             | Estimate | Std. Error | t value | Pr(>|t|)  |
|-------------|----------|------------|---------|-----------|
| (Intercept) | 66.500   | 5.573      | 11.93   | 1.10e-09 *** |

---

Signif. codes:  0 ** ** ** 0.001 ** 0.01 * 0.05 . 1  

Residual standard error: 23.65 on 17 degrees of freedom

#Alternatively we could fit a model where the rail values are levels of a factor with fixed effects

fRail<-as.factor(Rail[,1])

fixed.fit2<-lm(travel~fRail-1,data=Rail)

summary(fixed.fit2)

Coefficients:

|             | Estimate | Std. Error | t value | Pr(>|t|)  |
|-------------|----------|------------|---------|-----------|
| fRail2      | 31.667   | 2.321      | 13.64   | 1.15e-08 *** |
| fRail5      | 50.000   | 2.321      | 21.54   | 5.86e-11 *** |
| fRail11     | 54.000   | 2.321      | 23.26   | 2.37e-11 *** |
| fRail16     | 82.667   | 2.321      | 35.61   | 1.54e-13 *** |
| fRail13     | 84.667   | 2.321      | 36.47   | 1.16e-13 *** |
| fRail14     | 96.000   | 2.321      | 41.35   | 2.59e-14 *** |

---

Signif. codes:  0 ** ** ** 0.001 ** 0.01 * 0.05 . 1  

Residual standard error: 4.021 on 12 degrees of freedom

# This model takes into account only the specific rails used in the experiment
# but we may be interested in the attributes of the underlying population of rails
```r
mixed.fit.ML<-lme(travel~1,data=Rail, random=~1|Rail,method="ML")

summary(mixed.fit.ML)

Linear mixed-effects model fit by maximum likelihood
Data: Rail
   AIC      BIC    logLik
134.5600 137.2312 -64.28002

Random effects:
  Formula: ~1 | Rail
    (Intercept) Residual
    StdDev: 22.62435 4.020779

Fixed effects: travel ~ 1
  Value Std.Error DF  t-value p-value
(Intercept) 66.5  9.554026 12  6.960417       0

mixed.fit.REML<-lme(travel~1,data=Rail, random=~1|Rail,method="REML")

summary(mixed.fit.REML)

Linear mixed-effects model fit by REML
Data: Rail
   AIC      BIC   logLik
128.177 130.6766 -61.0885

Random effects:
  Formula: ~1 | Rail
    (Intercept) Residual
    StdDev: 24.80547 4.020779

Fixed effects: travel ~ 1
  Value Std.Error DF  t-value p-value
(Intercept) 66.5 10.17104 12  6.538173       0

# in this simple case the random effects variance is equal but this does not hold in general.

plot(mixed.fit.REML)

intervals(mixed.fit.REML)

############################################################ Ergonomic stools Examples ############################################################

# An experiment is conducted to assess the effort needed to get up from 4 different stools. Nine people are used to measure the effort.

ergoStool

plot.design(ergoStool)

# one can notice that there is a significant variability from subject to subject (here the subjects are blocks). The comparison seems to indicate that the type 1 stool requires the smallest effort.

fSubject<-as.factor(ergoStool$Subject)

options(contrasts=c("contr.treatment","contr.poly"))

fix.fit<-lm(effort~Type+fSubject,data=ergoStool)

anova(fix.fit)
```
summary(fix.fit)

Coefficients:

|            | Estimate | Std. Error | t value | Pr(>|t|) |
|------------|----------|------------|---------|----------|
| (Intercept)| 8.55556  | 0.36676    | 23.327  | < 2e-16  *** |
| TypeT2     | 3.88889  | 0.51868    | 7.498   | 9.75e-08 *** |
| TypeT3     | 2.22222  | 0.51868    | 4.284   | 0.000256 *** |
| TypeT4     | 0.66667  | 0.51868    | 1.285   | 0.210951  |
| fSubject.L | 4.00208  | 0.55015    | 7.275   | 1.63e-07 *** |
| fSubject.Q | 0.02849  | 0.55015    | 0.052   | 0.959128  |
| fSubject.C | 0.11124  | 0.55015    | 0.202   | 0.841468  |
| fSubject^4 | 0.05587  | 0.55015    | 0.102   | 0.919949  |
| fSubject^5 | -0.57781 | 0.55015    | -1.050  | 0.304047  |
| fSubject^6 | -0.33710 | 0.55015    | -0.613  | 0.545808  |
| fSubject^7 | -0.27312 | 0.55015    | -0.496  | 0.624100  |
| fSubject^8 | -0.26444 | 0.55015    | -0.481  | 0.635099  |

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 1

Residual standard error: 1.1 on 24 degrees of freedom

rand.fit.ML<-lme(effort~Type, data=ergoStool, random = ~1 | fSubject,method="ML")

summary(rand.fit.ML)

Random effects:

Formula: ~1 | fSubject

(Intercept)  Residual
StdDev:   1.25626 1.037368

Fixed effects: effort ~ Type

Value Std.Error DF  t-value p-value
(Intercept) 8.555556 0.5760123 24 14.853079 0.0000
TypeT2     3.888889 0.5186838 24  7.497610 0.0000
TypeT3     2.222222 0.5186838 24  4.284348 0.0003
TypeT4     0.666667 0.5186838 24  1.285304 0.2110

anova(rand.fit.ML)

numDF denDF  F-value p-value
(Intercept)   1   24  455.0075  <.0001
Type           3   24   22.3556  <.0001

rand.fit.REML<-lme(effort~Type, data=ergoStool, random = ~1 | fSubject,method="REML")

summary(rand.fit.REML)

Linear mixed-effects model fit by REML
Data: ergoStool
AIC  BIC  logLik
 133.1308 141.9252 -60.5654

Random effects:

Formula: ~1 | fSubject

(Intercept)  Residual
StdDev:   1.332465 1.100295

Fixed effects: effort ~ Type

Value Std.Error DF  t-value p-value
(Intercept) 8.555556 0.5760123 24 14.853079 0.0000
TypeT2     3.888889 0.5186838 24  7.497610 0.0000
TypeT3     2.222222 0.5186838 24  4.284348 0.0003
TypeT4     0.666667 0.5186838 24  1.285304 0.2110

anova(rand.fit.REML)
numDF denDF  F-value  p-value
(Intercept) 1    24  455.0075  <.0001
Type         3    24  22.3556  <.0001

plot(rand.fit.REML)
intervals(rand.fit.REML)

# The following data gives the productivity score for each of six randomly chosen
# workers tested on each of three different machine types. Each worker used each
# machine three times. There is little variability in the productivity score for the
# same worker using the same machine but there are differences between machines and between
# workers.
attach(Machines)
plot.design(Machines)
interaction.plot(Machine, Worker, score, mean)

# In the first model considered
# there is a fixed effect for each type of machine and random effect for each worker.
mixed.fit.1.REML<-lme(score ~ Machine, random = ~1 | Worker)
summary(mixed.fit.1.REML)

Linear mixed-effects model fit by REML
Data: NULL
AIC      BIC    logLik
296.8782 306.5373 -143.4391
Random effects:
Formula: ˜1 | Worker
          (Intercept) Residual
StdDev:   5.146552 3.161647
Fixed effects: score ~ Machine
             Value   Std.Error   DF t-value  p-value
(Intercept) 52.35556  2.229312 46 23.48507       0
MachineB     7.96667  1.053883 46  7.55935       0
MachineC    13.91667  1.053883 46 13.20514       0

# If we would like to model the interactions between workers and machines
# we need to consider the interactions as nested random effects
# We need to consider effects of each machine "within" each worker.
mixed.fit.2.REML<-lme(score ~ Machine, random = ~1 | Worker/Machine)
summary(mixed.fit.2.REML)

Linear mixed-effects model fit by REML
Data: NULL
AIC      BIC    logLik
227.6876 239.2785 -107.8438
Random effects:
Formula: ˜1 | Worker
          (Intercept)
StdDev:   4.78105
Formula: `~1 | Machine %in% Worker
(Intercept) Residual
StdDev: 3.729532 0.9615771

Fixed effects: score ` Machine

Value Std.Error DF t-value p-value
(Intercept) 52.35556 2.485828 36 21.061613 0.0000
MachineB 7.96667 2.176972 10 3.659518 0.0044
MachineC 13.91667 2.176972 10 6.392672 0.0001

# we can compare the two models with ANOVA

anova(mixed.fit.1.REML,mixed.fit.2.REML)

Model df      AIC      BIC    logLik   Test  L.Ratio p-value
mixed.fit.1.REML     1 5 296.8782 306.5373 -143.4391
mixed.fit.2.REML     2 6 227.6876 239.2785 -107.8438 1 vs 2 71.19063 <.0001

##### Repeated measurements over time #######

# Consider the orthodontic data (Orthodont) which contains measurements of the
distance between the pituitary gland and the pterygomaxillary fissure taken
every two years from 8 years of age until 14 years of age on a sample of 27
children - 16 males and 11 females.

attach(Orthodont)

# we can look at the female data first
OrthoFem<-Orthodont[Orthodont$Sex=="Female",]

# using the lmList command we fit separate linear regression model for each girl
fitOrthF.lis<-lmList(distance~age,data=OrthoFem)

# find the coefficients of each regression
coef(fitOrthF.lis)

# plot the confidence intervals
plot(intervals(fitOrthF.lis))

# the intervals have the same length due to the balanced design
# we may want to use a model with common slope
# we also notice that there is some (negative) dependence between the slope and the intercep

var(coef(fitOrthF.lis))

(Intercept)     age
(Intercept)  6.3636818 -0.35311364
age        -0.3531136  0.04822727

# the correlation coefficient is estimated to be
-0.35/sqrt(6.36*0.048)
[1] -0.6334595

# we can get rid of this correlation by recentering the covariate values.
# we subtract the mean of the x’s which is 11

fitOrthF.lis<-lmList(distance~I(age-11),data=OrthoFem)

plot(intervals(fitOrthF.lis))
# one can see that the negative correlation has vanished
# instead of representing the fitted response at age 0
# now the intercept represents the fitted response at age 11

# without considering a random effect we are essentially drawing conclusions
# only about the girls included in the study and not about the underlying
# population of girls. In a random effects model we make inferences about the
# fixed effects which represent characteristics of the population.

fitmixOrthF<-lme(distance~I(age-11),data = OrthoFem, random=~1|Subject)
summary(fitmixOrthF)

Linear mixed-effects model fit by REML
Data: OrthoFem
   AIC     BIC    logLik
 149.22  156.17 -70.60916
Random effects:
Formula: ~1 | Subject
   (Intercept)  Residual
    StdDev:     2.06847 0.7800331
Fixed effects: distance ~ I(age - 11)
                Value Std.Error   DF t-value p-value
(Intercept) 22.647727 0.6346568 32 35.6850       0
I(age - 11)  0.479545 0.0525898 32  9.1186       0
Correlation:
 (Intr)
I(age - 11) 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

# instead of a common growth rate we can assume random slopes as well

fitmixOrthF2<-lme(distance~I(age-11),data = OrthoFem, random=~age|Subject)
summary(fitmixOrthF2)

# we can compare the two models using the likelihood ratio test

anova(fitmixOrthF,fitmixOrthF2)

             Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fitmixOrthF 1  4  149.2183 156.1690 -70.60916
fitmixOrthF2 2  6 149.4287 159.8547 -68.71435 1 vs 2 3.789622  0.1503

# we conclude the first model is adequate

# using random.effects we can predict the random effects

random.effects(fitmixOrthF)

F10 -4.00532866
F09 -1.47044943
F06 -1.47044943
F01 -1.22903236
F05 -0.02194701
F04  0.34017860
F02  0.34017860
F08  0.70230420
F03  1.06442981
F07  2.15080662
F11  3.59930904
# coef finds the coefficients for each fitted line
# (the intercepts include the random effects)

coef(fitmixOrthF)

(Intercept) I(age - 11)
F10     18.64240   0.4795455
F09     21.17728   0.4795455
F06     21.17728   0.4795455
F01     21.41869   0.4795455
F05     22.62578   0.4795455
F07     22.98791   0.4795455
F02     22.98791   0.4795455
F08     23.35003   0.4795455
F03     23.71216   0.4795455
F04     24.79853   0.4795455
F11     26.24704   0.4795455

plot(coef(fitmixOrthF))
plot(coef(fitmixOrthF2))

# the augPred produces predictions for each group over the range of the covariate
# the predictions are superimposed with the actual observations

plot(augPred(fitmixOrthF),grid=T)
plot(augPred(fitmixOrthF2),grid=T)

# suppose now that we are interested in the comparison between
# boys and girls.

fitmix.RML<-lme(distance˜Sex*I(age-11),data=Orthodont, random=˜I(age-11)|Subject)
summary(fitmix.RML)

Linear mixed-effects model fit by REML
Data: Orthodont
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) 1.8303268 (Intr)
I(age - 11) 0.1803454 0.206
Residual    1.3100397

Fixed effects: distance ˜ Sex * I(age - 11)
  Value Std.Error DF  t-value p-value
(Intercept)           24.968750 0.4860007 79 51.37595  0.0000
SexFemale             -2.321023 0.7614168 25 -3.04829  0.0054
I(age - 11)            0.784375 0.0859995 79  9.12069  0.0000
SexFemale:I(age - 11) -0.304830 0.1347353 79 -2.26243  0.0264

Correlation:
     (Intr) SexFml I(-11)
SexFemale   -0.638
I(age - 11)  0.102 -0.065
SexFemale:I(age - 11) -0.065  0.102 -0.638

fitmix.ML<-lme(distance˜Sex*I(age-11),data=Orthodont, method="ML",random=˜I(age-11)|Subject)
summary(fitmix.ML)

Linear mixed-effects model fit by maximum likelihood
Data: Orthodont
# suppose we are interested in predicting new values
# the command used is predict

# first we create the new set of covariates
newOrth<-data.frame(Subject=rep(c("M11","F03"),c(3,3)),Sex=rep(c("Male","Female"),c(3,3)),
age=rep(16:18,2))
predict(fitmix.RML,newdata=newOrth) # this gives the within group predictions

M11      M11      M11      F03      F03      F03
26.96809 27.61195 28.25580 26.61357 27.20668 27.79979

predict(fitmix.RML,newdata=newOrth,level=0:1)

Subject predict.fixed predict.Subject
1     M11      28.89063        26.96809
2     M11      29.67500        27.61195
3     M11      30.45938        28.25580
4     F03      25.04545        26.61357
5     F03      25.52500        27.20668
6     F03      26.00455        27.79979

# the predict.fixed gives the population prediction
# the predict.Subject gives the within the group prediction

# All the methods discussed today work well with groupedData objects

Suppose we revisit the tire data

tire.data<-read.table("/Users/BaseCamp/Teaching/2101/tire.dat",he=T)
tire.data

Tire Compound y
1 1 A 238
2 1 B 238
3 1 C 279
4 2 A 196
5 2 B 213
6 2 D 308
7 3 A 254
8 3 C 334
9 3 D 367
10 4 B 312
11 4 C 421
12 4 D 412
> class(tire.data)
[1] "data.frame"

# Suppose we want to treat the tire as a random effect model

tire<groupedData(y~Compound|Tire, data=read.table("/Users/BaseCamp/Teaching/2101/tire.dat",he=T))

tire

ftire<-as.factor(tire$Tire)
fCompound<-as.factor(tire$Compound)
fitmix<-lme(y~fCompound, data=tire, random=~1|ftire)
summary(fitmix)
Linear mixed-effects model fit by REML
Data: tire
AIC      BIC    logLik
94.95613 95.43278 -41.47807

Random effects:
Formula: ~1 | ftire
 (Intercept) Residual
StdDev: 49.98397 18.71319

Fixed effects: y ~ fCompound
Value Std.Error DF  t-value p-value
(Intercept) 251.14521  27.41758  5 9.160005  0.0003
fCompoundB    5.40494  16.16108  5 0.334442  0.7516
fCompoundC   78.20169  16.16108  5 4.838891  0.0047
fCompoundD  102.47921  16.16108  5 6.341113  0.0014

fit.intra<-lm(y~ftire+fCompound, data=tire)
summary(fit.intra)

Call:
  lm(formula = y ~ ftire + fCompound, data = tire)

Residuals:
  1       2       3       4       5       6       7       8       9      10      11
12
042

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
(Intercept) 252.292     11.299  22.328 3.35e-06 ***
ftire.L 77.033     11.459  6.722  0.00110 **
ftire.Q   41.375     11.459  3.611  0.01537 *
ftire.C -15.597     11.459 -1.361  0.23164
fCompoundB  4.375     16.206  0.270  0.79798
fCompoundC  76.250     16.206  4.705  0.00531 **
fCompoundD 100.875     16.206  6.225  0.00157 **

Residual standard error: 18.71 on 5 degrees of freedom