In class, I described the ideas of *random effects* and *fixed effects*. It is appropriate to model a factor as a *random effect* if you wish to think of the levels as having been randomly selected from a larger population whereas fixed effects are factors where the levels in the problem are the only ones of interest. We also talked about the difference between factors that are *crossed* or *nested*. Two factors are crossed if there is at least one observation for each pair of levels. Two factors are nested if the levels of one factor are completely contained in the levels of another factor. For the chimp example in Chapter 14, the factors *sign* and *chimp* are crossed because each chimp is measured at each sign. If we created a factor *sex*, then *chimp* would be nested within *sex* as the levels *Bruno* and *Booee* are nested within *male* and *Cindy* and *Thelma* are female.

For this problem, it makes sense to consider treating *chimp* as a random effect. However, R is quite limited in the types of mixed effects models that can be fit. Fitting nested factors works quite well, but fitting crossed factors when one or more are random is far from straight-forward. So, I'm not going to show you how to do it.

But here is an example of fitting a model with nested factors adding the factor *sex*. In the formula, we indicate the nesting of *chimp* within *sex* by `sex/CHIMP`. We will have to create the *sex* factor within R. (Alternatively, we could edit the file in which we read in the data). The strategy will be to create a variable that is female everywhere and then change the proper observations to male.

```
> case1401 <- read.table("sleuth/case1401.csv", header = T, sep = ",")
> attach(case1401)
> sex = rep("female", nrow(case1401))
> sex[CHIMP == "BOOEE"] <- "male"
> sex[CHIMP == "BRUNO"] <- "male"
> sex <- factor(sex)
> new1401 <- data.frame(case1401, sex)
> detach()
> attach(new1401)
> fit1 <- lm(log(MINUTES) ~ SIGN + CHIMP)
> fit2 <- lm(log(MINUTES) ~ SIGN + sex/CHIMP)
> anova(fit1)

Analysis of Variance Table

Response: log(MINUTES)
          Df Sum Sq Mean Sq F value    Pr(>F)
SIGN       9 45.690   5.077  7.7649 1.498e-05 ***
CHIMP      3  5.333   1.778  2.7189   0.06421 .
Residuals 27 17.653   0.654
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fit2)

Analysis of Variance Table

Response: log(MINUTES)
          Df Sum Sq Mean Sq F value    Pr(>F)
SIGN       9 45.690   5.077  7.7649 1.498e-05 ***
sex        1  0.167   0.167  0.2558   0.61710
sex:CHIMP  2  5.166   2.583  3.9505   0.03127 *
Residuals 27 17.653   0.654
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Notice that the second model just breaks up the sum of squares from *chimp* into two pieces. There is little evidence of a sex effect to explain the differences between the chimps. The estimated regression coefficients for *sign* are identical, and the estimates for *chimp* are simply parameterized in a different way.

```
> summary(fit1)

Call:
lm(formula = log(MINUTES) ~ SIGN + CHIMP)

Residuals:
     Min       1Q   Median       3Q      Max
-1.70580 -0.38195  0.03512  0.41762  1.73890

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.69487    0.46096   5.846 3.18e-06 ***
SIGNfood     0.96866    0.57175   1.694 0.101736
SIGNfruit    1.35652    0.57175   2.373 0.025047 *
SIGNhat      1.85663    0.57175   3.247 0.003108 **
SIGNkey      0.12460    0.57175   0.218 0.829124
SIGNlisten  -1.05303    0.57175  -1.842 0.076520 .
SIGNlook     2.05422    0.57175   3.593 0.001286 **
SIGNmore     0.30099    0.57175   0.526 0.602881
SIGNshoe     0.05467    0.57175   0.096 0.924527
SIGNstring   2.55070    0.57175   4.461 0.000129 ***
CHIMPBRUNO   0.68133    0.36161   1.884 0.070351 .
CHIMPCINDY   0.09287    0.36161   0.257 0.799270
CHIMPTHELMA  0.84713    0.36161   2.343 0.026766 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8086 on 27 degrees of freedom
Multiple R-Squared: 0.743,        Adjusted R-squared: 0.6287
F-statistic: 6.503 on 12 and 27 DF,  p-value: 2.776e-05

> summary(fit2)

Call:
lm(formula = log(MINUTES) ~ SIGN + sex/CHIMP)

Residuals:
     Min       1Q   Median       3Q      Max
-1.70580 -0.38195  0.03512  0.41762  1.73890

Coefficients: (4 not defined because of singularities)
                     Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)           3.54200    0.46096   7.684  2.9e-08 ***
SIGNfood              0.96866    0.57175   1.694 0.101736
SIGNfruit             1.35652    0.57175   2.373 0.025047 *
SIGNhat               1.85663    0.57175   3.247 0.003108 **
SIGNkey               0.12460    0.57175   0.218 0.829124
SIGNlisten           -1.05303    0.57175  -1.842 0.076520 .
SIGNlook              2.05422    0.57175   3.593 0.001286 **
SIGNmore              0.30099    0.57175   0.526 0.602881
SIGNshoe              0.05467    0.57175   0.096 0.924527
SIGNstring            2.55070    0.57175   4.461 0.000129 ***
sexmale              -0.84713    0.36161  -2.343 0.026766 *
sexmale:CHIMPBRUNO    0.68133    0.36161   1.884 0.070351 .
sexfemale:CHIMPCINDY -0.75426    0.36161  -2.086 0.046571 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8086 on 27 degrees of freedom
Multiple R-Squared: 0.743,        Adjusted R-squared: 0.6287
F-statistic: 6.503 on 12 and 27 DF,  p-value: 2.776e-05
```