Ch. 6 Two-way ANOVA: Fixed factor levels
Unequal sample sizes

1 Assumptions and models

- There are two factors, factors A and B, that are of interest.

- Factor A is studied at a levels, and factor B at b levels;

  All ab factor level combinations are included in the study.

  There may be empty cells.

Table 1: Format of (unbalanced) data set

		Factor B				
		level 1	$\ldots$	level j	$\ldots$	level b
	level 1	$Y_{111}, \ldots, Y_{11n_{11}}$		$Y_{1j1}, \ldots, Y_{1jn_{1j}}$		$Y_{1b1}, \ldots, Y_{1bn_{1b}}$
	$\vdots$					
Factor A	level i	$Y_{i11}, \ldots, Y_{i1n_{i1}}$		$Y_{ij1}, \ldots, Y_{ijn_{ij}}$		$Y_{ib1}, \ldots, Y_{ibn_{ib}}$
	$\vdots$					
	level a	$Y_{a11}, \ldots, Y_{a1n_{a1}}$		$Y_{aj1}, \ldots, Y_{ajn_{aj}}$		$Y_{ab1}, \ldots, Y_{abn_{ab}}$

1.1 Cell means model

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}.$$

$\mu_{ij}, \quad i = 1, \ldots, a, j = 1, \ldots, b$, are mean parameters;

Error terms $\quad \varepsilon_{ijk} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}.$

- Feature of model

$$Y_{ijk} \overset{\text{independent}}{\sim} \mathcal{N}(\mu_{ij}, \sigma^2), \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}.$$

Table 2: Cell means

		Factor B			
		level 1 ...	level j ...	level b	
	level 1	μ_{11}	μ_{1j}	μ_{1b}	$\mu_{1\cdot}$
Factor A	$\vdots$				
	level i	μ_{i1}	μ_{ij}	μ_{ib}	$\mu_{i\cdot}$
	$\vdots$				
	level a	μ_{a1}	μ_{aj}	μ_{ab}	$\mu_{a\cdot}$
		$\mu_{\cdot 1}$	$\mu_{\cdot j}$	$\mu_{\cdot b}$	

1.2   Factor effects model

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}.$$

$\mu_{..}$ is a constant or overall mean   ($\mu_{..} = \frac{1}{ab} \sum\limits_{i=1}^{a} \sum\limits_{i=1}^{b} \mu_{ij}$);

α_i are constants subject to the restriction $\sum\limits_{i=1}^{a} \alpha_i = 0$

   ($\alpha_i = \mu_{i\cdot} - \mu_{..}$,   the main effect for factor A at the ith level);

β_j are constants subject to the restriction $\sum\limits_{j=1}^{b} \beta_j = 0$

   ($\beta_j = \mu_{\cdot j} - \mu_{..}$,   the main effect for factor B at the jth level);

$(\alpha\beta)_{ij}$ are constants subject to the restrictions

$$\sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, \quad j = 1, \ldots, b, \qquad\qquad \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0, \quad i = 1, \ldots, a$$

($(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{..}$,

the interaction effect for factor A at the ith level and factor B at the jth level).

- Feature of model

$$Y_{ijk} \overset{\text{independent}}{\sim} \mathcal{N}(\mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \ \sigma^2), \quad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}.$$

1.3   Notations

1.3.1   Means

$$\bar{Y}_{...} = \frac{1}{n_T} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} Y_{ijk} \qquad \text{(The overall mean)}$$

$$\bar{Y}_{i..} = \frac{1}{n_{i.}} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} Y_{ijk} \qquad \text{(The mean for the } i\text{th factor level of A)}$$

$$\bar{Y}_{\cdot j \cdot} = \frac{1}{n_{\cdot j}} \sum_{i=1}^{a} \sum_{k=1}^{n_{ij}} Y_{ijk} \qquad \text{(The mean for the } j\text{th factor level of B)}$$

$$\bar{Y}_{ij\cdot} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} Y_{ijk} \qquad \text{(The mean for the } (i,j)\text{th treatment)}$$

where

$$n_T = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}, \qquad n_{i.} = \sum_{j=1}^{b} n_{ij}, \qquad n_{\cdot j} = \sum_{i=1}^{a} n_{ij}.$$

Relationships (partitioning)

$$\underbrace{Y_{ijk} - \bar{Y}_{...}}_{\text{Total deviation}} = \underbrace{\bar{Y}_{ij\cdot} - \bar{Y}_{...}}_{\substack{\text{Deviation of} \\ \text{estimated treatment mean} \\ \text{around overall mean}}} + \underbrace{Y_{ijk} - \bar{Y}_{ij\cdot}}_{\substack{\text{Deviation around} \\ \text{estimated treatment mean}}}$$

1.3.2   Sums of squares

We can define sums of squares as the balanced data case.

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} \left(Y_{ijk} - \bar{Y}_{ij\cdot}\right)^2$$

$$SSTO = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} \left(Y_{ijk} - \bar{Y}_{...}\right)^2$$

$$SSA = \sum_{i=1}^{a} n_{i.} (\bar{Y}_{i..} - \bar{Y}_{...})^2 \qquad \text{(Factor A sum of squares)}$$

$$SSB = \sum_{j=1}^{b} n_{\cdot j} (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{...})^2 \qquad \text{(Factor B sum of squares)}$$

$$SSAB = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} (\bar{Y}_{ij\cdot} - \bar{Y}_{i..} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{...})^2 \qquad \text{(AB interaction sum of squares)}$$

However, the following equation does NOT hold for the unbalanced data.

$$SSTO = SSA + SSB + SSAB + SSE$$

Growth hormone example, p.892.

Factor A (gender) has 2 factor levels, and factor B (bone development) has 3 factor levels.

```
> y <- read.table("CH23TA01.DAT")
      V1 V2 V3 V4
   1 1.4  1  1  1
    ......
   7  1.1  1  3  2
   8  2.4  2  1  1
  ......
   14 1.3  2  3  3
```

Let's compute means and sum of squares. An important question is: $SSA + SSB + SSAB + SSE = SSTO$?

```
> y... <- mean(y[,1])


> y1.. <- mean(y[,1][y[,2]==1])
> y2.. <- mean(y[,1][y[,2]==2])
> SSA <-  sum((c(rep(y1..,length(y[,1][y[,2]==1])),
                rep(y2..,length(y[,1][y[,2]==2])))-y...)^2)
    [1] 0.002857143


> y.1. <- mean(y[,1][y[,3]==1])
> y.2. <- mean(y[,1][y[,3]==2])
> y.3. <- mean(y[,1][y[,3]==3])
> SSB <- sum((c(rep(y.1.,length(y[,1][y[,3]==1])),
              rep(y.2.,length(y[,1][y[,3]==2])),
              rep(y.3.,length(y[,1][y[,3]==3])))-y...)^2)
    [1] 4.306286

 > y11. <- mean(y[,1][y[,2]==1&y[,3]==1])
 > y12. <- mean(y[,1][y[,2]==1&y[,3]==2])
 > y13. <- mean(y[,1][y[,2]==1&y[,3]==3])
 > y21. <- mean(y[,1][y[,2]==2&y[,3]==1])
 > y22. <- mean(y[,1][y[,2]==2&y[,3]==2])
 > y23. <- mean(y[,1][y[,2]==2&y[,3]==3])
 > SSAB <- (y11.-y1..-y.1.+y...)^2*length(y[,1][y[,2]==1&y[,3]==1])+
           (y12.-y1..-y.2.+y...)^2*length(y[,1][y[,2]==1&y[,3]==2])+
           (y13.-y1..-y.3.+y...)^2*length(y[,1][y[,2]==1&y[,3]==3])+
           (y21.-y2..-y.1.+y...)^2*length(y[,1][y[,2]==2&y[,3]==1])+
           (y22.-y2..-y.2.+y...)^2*length(y[,1][y[,2]==2&y[,3]==2])+
           (y23.-y2..-y.3.+y...)^2*length(y[,1][y[,2]==2&y[,3]==3])
    [1] 0.2017143
```

```
> fitted <-  c(rep(y11., length(y[,1][y[,2]==1&y[,3]==1])),
               rep(y12., length(y[,1][y[,2]==1&y[,3]==2])),
               rep(y13., length(y[,1][y[,2]==1&y[,3]==3])),
               rep(y21., length(y[,1][y[,2]==2&y[,3]==1])),
               rep(y22., length(y[,1][y[,2]==2&y[,3]==2])),
               rep(y23., length(y[,1][y[,2]==2&y[,3]==3]))))
> SSE <- sum((y[,1]-fitted)^2)
     [1] 1.3


> SSTO <- sum((y[,1]-rep(y...,length(y[,1])))^2)
     [1] 5.774286


> SSA+SSB+SSAB+SSE-SSTO
     [1] 0.03657143         # SSA+SSB+SSAB+SSE > SSTO
```

Next we compare these sums of squares with that obtained via the two-way ANOVA table.

```
> data <- y[,1]
> gender <- y[,2]
> bone <- y[,3]
> growth.df <- data.frame(data=data,gender=factor(gender),bone=factor(bone))

> summary(aov(data~gender+bone+gender*bone,growth.df)  )
             Df  Sum Sq  Mean Sq  F value   Pr(>F)
 gender       1  0.0029  0.0029    0.0176   0.897785
 bone         2  4.3960  2.1980   13.5262   0.002713
 gender:bone  2  0.0754  0.0377    0.2321   0.798034
 Residuals    8  1.3000  0.1625

> aov(data~gender+bone+gender*bone,growth.df)
  Terms:
                   gender       bone    gender:bone  Residuals
  Sum of Squares   0.002857   4.396000  0.075429     1.300000
  Deg. of Freedom         1          2         2            8

  Residual standard error: 0.4031129
  Estimated effects may be unbalanced
```

Something is wrong here. What's the problem?

2   Estimation for mean parameters and σ^2

The least squares and maximum likelihood estimators of μ_{ij}

$$\hat{\mu}_{ij} = \bar{Y}_{ij\cdot}$$

Fitted value

$$\bar{Y}_{ijk} = \bar{Y}_{ij\cdot}$$

Residual

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ijk} = Y_{ijk} - \bar{Y}_{ij\cdot}$$

Parameter	Estimator	Confidence interval
μ_{ij}	$\hat{\mu}_{ij} = \bar{Y}_{ij\cdot} \ \sim \ \mathcal{N}(\mu_{ij}, \frac{\sigma^2}{n_{ij}})$	$\bar{Y}_{ij\cdot} \pm qt(1 - \frac{\alpha}{2}; n_T - ab)\sqrt{\frac{MSE}{n_{ij}}}$
$\mu_{i\cdot}$	$\hat{\mu}_{i\cdot} = \bar{Y}_{i\cdot\cdot} \ \sim \ \mathcal{N}(\mu_{i\cdot}, \frac{\sigma^2}{b^2}\sum_{j}\frac{1}{n_{ij}})$	$\bar{Y}_{i\cdot\cdot} \pm qt(1 - \frac{\alpha}{2}; n_T - ab)\sqrt{\frac{MSE}{b^2}\sum_{j}\frac{1}{n_{ij}}}$
$\mu_{\cdot j}$	$\hat{\mu}_{\cdot j} = \bar{Y}_{\cdot j\cdot} \ \sim \ \mathcal{N}(\mu_{\cdot j}, \frac{\sigma^2}{a^2}\sum_{i}\frac{1}{n_{ij}})$	$\bar{Y}_{\cdot j\cdot} \pm qt(1 - \frac{\alpha}{2}; n_T - ab)\sqrt{\frac{MSE}{a^2}\sum_{i}\frac{1}{n_{ij}}}$
σ^2	$\widehat{\sigma^2} = MSE, \quad \frac{SSE}{\sigma^2} \sim \chi^2_{n_T - ab}$	

3   Use of regression approach for testing factor effects

Consider the factor effects model

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n_{ij}.$$

Let's count the number of (independent) parameters among $\alpha, \beta, (\alpha\beta)$.

For α, there are $a - 1$ (independent) parameters because of the constraint $\sum\limits_{i=1}^{a} \alpha_i = 0$.

Similarly, for β, there are $b - 1$ (independent) parameters.

For $(\alpha\beta)$, there are $(a-1)(b-1)$ (independent) parameters.

In all, the number of (independent) parameters in the pool is

$$(a - 1) + (b - 1) + (a - 1)(b - 1) = ab - 1.$$

Growth hormone example, p.892.   In this case, $a = 2, b = 3$. Thus, there are only 6 (independent) parameters among $\alpha, \beta, (\alpha\beta)$. Choose $\alpha_1, \beta_1, \beta_2, (\alpha\beta)_{11}, (\alpha\beta)_{12}$ as our variables. The rest are expressed as their linear combinations

Table 3: Cell means

		Factor B			
		level 1	level 2	level 3	
	level 1	$(\alpha\beta)_{11}$	$(\alpha\beta)_{12}$	$(\alpha\beta)_{13} = -(\alpha\beta)_{11} - (\alpha\beta)_{12}$	α_1
Factor A					
	level 2	$(\alpha\beta)_{21} = -(\alpha\beta)_{11}$	$(\alpha\beta)_{22} = -(\alpha\beta)_{12}$	$(\alpha\beta)_{23} = (\alpha\beta)_{11} + (\alpha\beta)_{12}$	$\alpha_2 = -\alpha_1$
		β_1	β_2	$\beta_3 = -\beta_1 - \beta_2$	

Regression model

$$
\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{221} \\ Y_{222} \\ Y_{223} \\ Y_{231} \\ Y_{232} \\ Y_{233} \end{pmatrix}
=
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \mu_{..} +
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \alpha_1 +
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ -1 \\ -1 \\ -1 \end{pmatrix} \beta_1 +
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} \beta_2 +
\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} (\alpha\beta)_{11} +
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (\alpha\beta)_{12} + \varepsilon
$$

The regression model is

$$
Y = \mu_{..} + X_1\alpha_1 + X_2\beta_1 + X_3\beta_2 + X_{11}(\alpha\beta)_{11} + X_{12}(\alpha\beta)_{12} + \varepsilon.
$$

In this case, $X_{11} = X_1 * X_2, X_{12} = X_1 * X_3$.

The question is then: how to set up X_1, X_2, X_3? Of course, we can type X_1, X_2, X_3 as

```
> x1 <- c(rep(1, length(y[,1][y[,2]==1])), rep(-1, length(y[,1][y[,2]==2])))
> x2 <- c(1,1,1,0,0,-1,-1,1,0,0,0,-1,-1,-1)
> x3 <- c(0,0,0,1,1,-1,-1,0,1,1,1,-1,-1,-1)
```

In practice, the length of such a vector may be too large to type. Thus, it's important to define X_1, X_2, X_3 more efficiently.

```
> y <- read.table("CH23TA01.DAT")
> y
        V1 V2 V3 V4
    1  1.4  1  1  1
    2  2.4  1  1  2
       ...  ...  ...
    14 1.3  2  3  3
```

Now add three more columns using *cbind*, which will be X_1, X_2, X_3.

```
> x1 <- 1:length(y[,1])
> x2 <- 1:length(y[,1])
> x3 <- 1:length(y[,1])

> y <- cbind(y, x1, x2, x3)
        V1 V2 V3 V4 x1 x2 x3
    1  1.4  1  1  1  1  1  1
    2  2.4  1  1  2  2  2  2
       ...  ...  ...
    14 1.3  2  3  3 14 14 14
```

For x1, indicators 1 and -1   are assigned as follows.

```
> x1[y[,2]==1]<-1
> x1[y[,2]==2]<--1

> y
        V1 V2 V3 V4 x1 x2 x3
 1  1.4  1  1  1  1  1  1
 2  2.4  1  1  2  1  2  2
 3  2.2  1  1  3  1  3  3
 4  2.1  1  2  1  1  4  4
 5  1.7  1  2  2  1  5  5
 6  0.7  1  3  1  1  6  6
 7  1.1  1  3  2  1  7  7
 8  2.4  2  1  1 -1  8  8
 9  2.5  2  2  1 -1  9  9
10 1.8  2  2  2 -1 10 10
11 2.0  2  2  3 -1 11 11
12 0.5  2  3  1 -1 12 12
13 0.9  2  3  2 -1 13 13
14 1.3  2  3  3 -1 14 14
```

For x2, indicators 1, 0 and -1 are assigned as follows.

```
> x2[y[,3]==1]<-1
> x2[y[,3]==2]<-0
> x2[y[,3]==3]<--1
       V1 V2 V3 V4 x1 x2 x3
  1   1.4  1  1  1  1  1  1
  2   2.4  1  1  2  1  1  2
  3   2.2  1  1  3  1  1  3
  4   2.1  1  2  1  1  0  4
  5   1.7  1  2  2  1  0  5
  6   0.7  1  3  1  1 -1  6
  7   1.1  1  3  2  1 -1  7
  8   2.4  2  1  1 -1  1  8
  9   2.5  2  2  1 -1  0  9
 10   1.8  2  2  2 -1  0 10
 11   2.0  2  2  3 -1  0 11
 12   0.5  2  3  1 -1 -1 12
 13   0.9  2  3  2 -1 -1 13
 14   1.3  2  3  3 -1 -1 14
```

Similarly, set up x3.

```
> x3[y[,3]==1]<-0
> x3[y[,3]==2]<-1
> x3[y[,3]==3]<--1
       V1 V2 V3 V4 x1 x2 x3
  1   1.4  1  1  1  1  1  0
  2   2.4  1  1  2  1  1  0
  3   2.2  1  1  3  1  1  0
  4   2.1  1  2  1  1  0  1
  5   1.7  1  2  2  1  0  1
  6   0.7  1  3  1  1 -1 -1
  7   1.1  1  3  2  1 -1 -1
  8   2.4  2  1  1 -1  1  0
  9   2.5  2  2  1 -1  0  1
 10   1.8  2  2  2 -1  0  1
 11   2.0  2  2  3 -1  0  1
 12   0.5  2  3  1 -1 -1 -1
 13   0.9  2  3  2 -1 -1 -1
 14   1.3  2  3  3 -1 -1 -1
```

Define x11 and x13 as

```
> x11 <- x1*x2
> x12 <- x1*x3
```

In general, for a two-factor study, factor A has a levels, and factor B has b levels.

Question 1. How many indicator variables we need for factor A and B, respectively?

Question 2. How to obtain indicators for interaction terms?

Fit of full regression model

$$Y = \mu_{..} + X_1\alpha_1 + X_2\beta_1 + X_3\beta_2 + X_{11}(\alpha\beta)_{11} + X_{12}(\alpha\beta)_{12} + \varepsilon.$$

The full regression model is fitted via

```
> full <- lm(y[,1]~x1+x2+x3+x11+x12)
> summary(full)
  Call:  lm(formula = y[, 1] ~ x1 + x2 + x3 + x12 + x13)
  Residuals:
        Min         1Q      Median         3Q         Max
  -6.000e-01 -2.000e-01 -8.816e-17  2.000e-01  4.000e-01
  Coefficients:
              Estimate Std. Error   t value Pr(>|t|)
  (Intercept)   1.7000     0.1164    14.609 4.73e-07
  x1           -0.1000     0.1164    -0.859   0.4152
  x2            0.5000     0.1778     2.813   0.0227
  x3            0.3000     0.1576     1.904   0.0934
  x1.x2        -0.1000     0.1778    -0.563   0.5891
  x1.x3      -1.77e-17     0.1576 -1.12e-16   1.0000


  Residual standard error: 0.4031 on 8 degrees of freedom
  Multiple R-Squared: 0.7749,     Adjusted R-squared: 0.6342
  F-statistic: 5.507 on 5 and 8 degrees of freedom,
  p-value: 0.01722
```

The residual sum of squares, $SSE(FULL)$, is of particular interest and will be used for the ANOVA below.

```
> sse.full <- (0.4031*sqrt(8))^2
  [1] 1.299917
```

Alternatively, $SSE(FULL)$ can be found from

```
> anova(full)
  Analysis of Variance Table
  Response: y[, 1]
            Df    Sum Sq   Mean Sq   F value    Pr(>F)
  x1         1    0.0029    0.0029    0.0176  0.897785
  x2         1    3.6509    3.6509   22.4668  0.001464
  x3         1    0.7451    0.7451    4.5855  0.064638
  x11        1    0.0754    0.0754    0.4642  0.514913
  x12        1 2.045e-33 2.045e-33 1.259e-32  1.000000

  Residuals  8    1.3000    0.1625
```

Test for AB interaction effects

$$\mathcal{H}_0: \quad \text{all } (\alpha\beta)_{ij} = 0, \qquad \mathcal{H}_1: \quad \text{not all } (\alpha\beta)_{ij} \text{ equal } 0.$$

In this case, $\mathcal{H}_0$ is equivalent to the following reduced regression model

$$Y = \mu_{..} + X_1\alpha_1 + X_2\beta_1 + X_3\beta_2 + \varepsilon,$$

which is fitted by

```
> ab <- lm(y[,1]~x1+x2+x3)
```

The corresponding SSE is obtained from either `summary(ab)` or `anova(ab)`.

```
> anova(ab)
  Analysis of Variance Table

  Response: y[, 1]
            Df  Sum Sq   Mean Sq  F value     Pr(>F)
  x1         1  0.0029   0.0029   0.0208     0.8882630
  x2         1  3.6509   3.6509   26.5435    0.0004302
  x3         1  0.7451   0.7451   5.4175     0.0422260
  Residuals 10  1.3754   0.1375
> sse.ab <-  anova(ab)[4,2]
  [1]  1.375429
```

The F test statistic is

$$F^* = \frac{SSE(AB) - SSE(FULL)}{df_{SSE(AB)} - df_{SSE(FULL)}} \div \frac{SSE(FULL)}{df_{SSE(FULL)}}.$$

```
> f.ab <- ((sse.ab-sse.full)/(10-8))/(sse.full/8)
  [1] 0.2330956
> 1-pf(f.ab, 10-8,8)          # P-value for F-test
  [1] 0.7972741
```

The p-value is 79.7%, which is large enough to accept $\mathcal{H}_0$. That is, factors A and B do not interact. Alternatively, `interaction.plot` tells the same story.

```
> interaction.plot(gender,bone,data)
```

The lines in the interaction plot are approximately parallel.

Test for factor A main effects

$$\mathcal{H}_0: \quad \text{all } \alpha_i = 0, \qquad \mathcal{H}_1: \quad \text{not all } \alpha_i \text{ equal } 0.$$

The reduced regression model corresponding to $\mathcal{H}_0$ is

$$Y = \mu_{..} + X_2\beta_1 + X_3\beta_2 + X_{11}(\alpha\beta)_{11} + X_{12}(\alpha\beta)_{12} + \varepsilon,$$

which is fitted by

```
> a <- lm(y[,1]~x2+x3+x11+x12)   # Don't use a <- lm(y[,1]~x2+x3+x1*x2+x1*x3)
```

The F test statistic is

$$F^* = \frac{SSE(A) - SSE(FULL)}{df_{SSE(A)} - df_{SSE(FULL)}} \div \frac{SSE(FULL)}{df_{SSE(FULL)}}.$$

```
> sse.a <- anova(a)[5,2]
  [1] 1.42
> f.a <- ((sse.a-sse.full)/(9-8))/(sse.full/8)
  [1] 0.7384699
> 1-pf(f.a, 9-8,8)            # P-value for F-test
  [1] 0.4151579              # What's your conclusion?
```

Test for factor B main effects

$$\mathcal{H}_0: \quad \text{all } \beta_j = 0, \qquad \mathcal{H}_1: \quad \text{not all } \beta_j \text{ equal } 0.$$

The reduced regression model corresponding to $\mathcal{H}_0$ is

$$Y = \mu_{..} + X_1\alpha_1 + X_{11}(\alpha\beta)_{11} + X_{12}(\alpha\beta)_{12} + \varepsilon,$$

which is fitted by

```
> b <- lm(y[,1]~x1+x11+x12)        # Not  b <- lm(y[,1]~x1+x1*x2+x1*x3
```

The F test statistic is

$$F^* = \frac{SSE(B) - SSE(FULL)}{df_{SSE(B)} - df_{SSE(FULL)}} \div \frac{SSE(FULL)}{df_{SSE(FULL)}}.$$

```
> sse.b <- anova(b)[4,2]
> f.b <- ((sse.b-sse.full)/(10-8))/(sse.full/8)
   [1] 12.89132
> 1-pf(f.b, 10-8, 8)       # P-value for F-test
   0.003144746             # What's your conclusion?
```

4   Empty cell

Question: What happens if $n_{ij} = 0$?

df for Error(Residuals) $= 0$;   SSE$=0$; $SSTO = SSTR = SSA + SSB + SSAB.$

No-interaction model:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b,$$

where

$$\alpha_i = \mu_{i.} - \mu_{..}, \qquad\qquad \beta_j = \mu_{.j} - \mu_{...}$$