Ch. 5 Two-way ANOVA: Fixed effect model
Equal 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.
- Format of (balanced) data set

		Factor B				
		level 1	$\dots$	level j	$\dots$	level b
Factor A	level 1	$Y_{111}, \dots, Y_{11n}$		$Y_{1j1}, \dots, Y_{1jn}$		$Y_{1b1}, \dots, Y_{1bn}$
	$\vdots$					
	level i	$Y_{i11}, \dots, Y_{i1n}$		$Y_{ij1}, \dots, Y_{ijn}$		$Y_{ib1}, \dots, Y_{ibn}$
	$\vdots$					
	level a	$Y_{a11}, \dots, Y_{a1n}$		$Y_{aj1}, \dots, Y_{ajn}$		$Y_{ab1}, \dots, Y_{abn}$

Mean parameters

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

1.1 Cell means model (for balanced data)

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

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

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

- Feature of model

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

1.2 Factor Effects Model

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

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

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

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

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

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

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

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

($(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot}$,
the interaction effect for factor A at the ith level and factor B at the jth level).

- Feature of model

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

1.3 Notations

Means

$$
\begin{aligned}
\bar{Y}_{\cdots} &= \frac{1}{nab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} && \text{(The overall mean)}\\
\bar{Y}_{i\cdot\cdot} &= \frac{1}{nb}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} && \text{(The mean for the } i\text{th factor level of A)}\\
\bar{Y}_{\cdot j\cdot} &= \frac{1}{na}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{ijk} && \text{(The mean for the } j\text{th factor level of B)}\\
\bar{Y}_{ij\cdot} &= \frac{1}{n}\sum_{k=1}^{n} Y_{ijk} && \text{(The mean for the } (i,j)\text{th treatment)}
\end{aligned}
$$

Sums of Squares

$$
\begin{aligned}
SS_{TR} &= n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij\cdot} - \bar{Y}_{\cdots})^2\\
SS_{E} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{ij\cdot})^2\\
SS_{T} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{\cdots})^2
\end{aligned}
$$

$$
\begin{aligned}
SS_{A} &= nb\sum_{i=1}^{a}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2 && \text{(Factor A sum of squares)}\\
SS_{B} &= na\sum_{j=1}^{b}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2 && \text{(Factor B sum of squares)}\\
SS_{AB} &= n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 && \text{(AB interaction sum of squares)}
\end{aligned}
$$

Relationships (partitioning)

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

$$
SS_T = SS_{TR} + SS_E, \qquad SS_{TR} \perp SS_E
$$

$$
\underbrace{\bar{Y}_{ij\cdot} - \bar{Y}_{\cdots}}_{\substack{\text{Deviation of}\\ \text{estimated treatment mean}\\ \text{around overall mean}}} = \underbrace{(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})}_{\text{A main effect}} + \underbrace{(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})}_{\text{B main effect}} + \underbrace{(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})}_{\text{AB interaction effect}}
$$

$$
SS_{TR} = SS_A + SS_B + SS_{AB}
$$

2 Estimation for mean parameters and σ^2

2.1 Cell means model

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

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

Fitted value of Y_{ijk}:

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

Residual related to Y_{ijk}:

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

Parameter	Estimator	Confidence interval
μ_{ij}	$\hat{\mu}_{ij} = \bar{Y}_{ij.} \sim \mathcal{N}(\mu_{ij}, \frac{\sigma^2}{n})$	$\bar{Y}_{ij.} \pm qt(1-\alpha/2; (n-1)ab)\sqrt{\frac{MSE}{n}}$
$\mu_{i.}$	$\hat{\mu}_{i.} = \bar{Y}_{i..} \sim \mathcal{N}(\mu_{i.}, \frac{\sigma^2}{bn})$	$\bar{Y}_{i..} \pm qt(1-\alpha/2; (n-1)ab)\sqrt{\frac{MS_E}{bn}}$
$\mu_{.j}$	$\hat{\mu}_{.j} = \bar{Y}_{.j.} \sim \mathcal{N}(\mu_{.j}, \frac{\sigma^2}{an})$	$\bar{Y}_{.j.} \pm qt(1-\alpha/2; (n-1)ab)\sqrt{\frac{MS_E}{an}}$
σ^2	$\widehat{\sigma^2} = MS_E, \quad \frac{SS_E}{\sigma^2} \sim \chi^2_{(n-1)ab}$	

2.2 Factor effects model

Parameter	(LS or MLE) Estimator
$\mu_{..}$	$\hat{\mu}_{..} = \bar{Y}_{...}$
$\alpha_i = \mu_{i.} - \mu_{..}$	$\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$
$\beta_j = \mu_{.j} - \mu_{..}$	$\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$
$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$	$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$

3 F tests

Two-Way ANOVA table

Source of Variation	df	Sum of Squares (SS)	Mean Square (MS)	F^*-value	Test for
Factor A	$a-1$	SS_A	$MS_A = \frac{SS_A}{a-1}$	$F^* = \frac{MS_A}{MS_E}$	$\mathcal{H}_0:\ \mu_{1\cdot} = \cdots = \mu_{a\cdot}$ (factor A main effects)
Factor B	$b-1$	SS_B	$MS_B = \frac{SS_B}{b-1}$	$F^* = \frac{MS_B}{MS_E}$	$\mathcal{H}_0:\ \mu_{\cdot 1} = \cdots = \mu_{\cdot b}$ (factor B main effects)
AB interactions	$(a-1)(b-1)$	SSAB	MSAB=$\frac{SS_{AB}}{(a-1)(b-1)}$	$F^* = \frac{MS_{AB}}{MS_E}$	$\mathcal{H}_0$: all $(\alpha\beta)_{ij} = 0$ (interactions)
Error(Residuals)	$ab(n-1)$	SS_E	$MS_E = \frac{SS_E}{ab(n-1)}$		
Total	$nab-1$	SS_T			

Castle Bakery Example, p.817.
Factor A (display height) has 3 factor levels, and factor B (display width) has 2 factor levels.

```
> y <- read.table("CH19TA07.DAT")
     V1 V2 V3 V4
1    47  1  1  1
  ......
4    40  1  2  2
5    62  2  1  1
  ......
8    71  2  2  2
9    41  3  1  1
  ......
12 46  3  2  2
```

Set the data frame, as the one-way layout case.

```
> data <- y[,1]
> height <- y[,2]
> width <- y[,3]

> sales.df <- data.frame(data=data,height=factor(height),width=factor(width))
```

The first step is often to look at the data graphically.

```
> plot.factor(height,data)      # Factor A  (boxplot)
> plot(height,data)                         (dot plot)

> plot.factor(width,data)       # Factor B
> plot(width,data)

> plot(sales.df)                # data frame
```

To obtain the ANOVA table, type `aov`, and then `summary`.

```
> anova <- aov(data~height+width+height*width,sales.df)
          # or        anova <- aov(data~height*width,sales.df)
          # compare   aov(data~height,sales.df)   aov(data~width,sales.df)
> summary(anova)
          # or        anova(lm(data~height*width,sales.df))
               Df  Sum Sq   Mean Sq  F value    Pr(>F)
  height        2   1544.00  772.00   74.7097    5.754e-05
  width         1   12.00    12.00    1.1613     0.3226
  height:width  2   24.00    12.00    1.1613     0.3747
  Residuals     6   62.00    10.33


 > model.tables(anova,type="means")        # treatment means

   Tables of means
     Grand mean  51

     height       1   2   3
                  44  67  42

     width        1   2
                  50  52

     height:width    Dim  1 = height, 2 = width
                       1   2
                 1  45  43
                 2  65  69
                 3  40  44
```

4 Analysis of factor effects when factors do not interact

- Where are we?

  Two possibilities:

  (i) Based on a F test for testing $\mathcal{H}_0$: all $(\alpha\beta)_{ij} = 0$, we accept $\mathcal{H}_0$. This means that factors do not interact.
  (ii) Factors A and B interact only in an unimportant fashion.

- What's next?

  Examine whether the main effects for factors A and B are important.
  For important A or B main effects, describe the nature of these effects in terms of the factor level means $\mu_{i\cdot}$ or $\mu_{\cdot j}$, respectively.

Castle Bakery Example, p.817.

```
> y <- read.table("CH19TA07.DAT")
> anova(lm(data~height*width,sales.df))
              Df  Sum Sq   Mean Sq  F value    Pr(>F)
 height       2   1544.00  772.00   74.7097    5.754e-05         # factor A
 width        1   12.00    12.00    1.1613     0.3226            # factor B
 height:width 2   24.00    12.00    1.1613     0.3747            # AB interaction
 Residuals    6   62.00    10.33
```

The p-value for testing $\mathcal{H}_0$: all $(\alpha\beta)_{ij} = 0$ is 37.5%, which leads us to accept $\mathcal{H}_0$. This means that factors do not interact.

The p-value for testing $\mathcal{H}_0$: all $\beta_j = 0$ is 32.3%, which leads us to accept $\mathcal{H}_0$. This means that factor B is not important.

How about factor A? Look at the p-value, which is `5.754e-05`. This means that factor A is important. We need analyze the factor effects in terms of $\mu_{1\cdot}, \mu_{2\cdot}$, and $\mu_{3\cdot}$.

$\rightarrowtail$ Multiple comparison procedure

$\circlearrowright$ Ch. 3 for analysis of one-way (fixed) factor level effects

Multiple comparison procedures

(cf. Ch. 3 for analysis of one-way (fixed) factor level effects)

⋇ (1-α)100 % Tukey simultaneous confidence intervals for all pairwise comparisons $\mu_i - \mu_{i'}$:

$$\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \pm T * \sqrt{MS_E\Big(\frac{1}{n_i} + \frac{1}{n_{i'}}\Big)}, \qquad i \neq i', i, i' \in \{1, \ldots, r\}$$

where

$$T \triangleq \frac{1}{\sqrt{2}} q(1-\alpha; r, n_T - r), \qquad \text{(Table B.9)}$$

⋇ (1-α)100 % Scheffé simultaneous confidence intervals for the family of contrasts L:

$$\sum_{i=1}^{r} c_i \bar{Y}_{i\cdot} \pm S * \sqrt{MS_E \sum_{i=1}^{r} \frac{c_i^2}{n_i}},$$

where

$$S^2 \triangleq (r-1) * qF(1-\alpha; r-1, n_T - r).$$

⋇ (1-α)100 % Bonferroni simultaneous confidence intervals for the g linear combinations L:

$$\sum_{i=1}^{r} c_i \bar{Y}_{i\cdot} \pm B * \sqrt{MS_E \sum_{i=1}^{r} \frac{c_i^2}{n_i}},$$

where

$$B \triangleq qt(1 - \frac{\alpha}{2g}; n_T - r).$$

5 Analysis of factor effects when interactions are important

- Where are we?

  Based on a F test for testing $\mathcal{H}_0$: all $(\alpha\beta)_{ij} = 0$, we accept $\mathcal{H}_a$. This means that factors A and B interact in an important fashion.

- What's next?

  Analyze the two factor effects jointly in terms of the treatment means μ_{ij}

Hay fever relief, CH19PR14.DAT.

```
> y <- read.table("CH19PR14.DAT")
> data <- y[,1]
> ingredient1 <- y[,2]
> ingredient2 <- y[,3]

> fever.df <- data.frame(data=data, ingredient1=factor(ingredient1),
                                    ingredient2=factor(ingredient2))


> anova(lm(data~ingredient1+ingredient2+ingredient1*ingredient2, fever.df))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ingredient1	2	220.020	110.010	1827.86	< 2.2e-16
ingredient2	2	123.660	61.830	1027.33	< 2.2e-16
ingredient1:ingredient2	4	29.425	7.356	122.23	< 2.2e-16
Residuals	27	1.625	0.060		

The p-value for testing $\mathcal{H}_0$: all $(\alpha\beta)_{ij} = 0$ is `< 2.2e-16`, which leads us to accept $\mathcal{H}_a$. This means that factors do interact. We need analyze the factor effects in terms of μ_{ij}

$\rightarrowtail$ Multiple comparison procedure

Multiple pairwise comparisons of treatment means

1° **Tukey procedure**

The Tukey $1-\alpha$ multiple comparison confidence limits for all pairwise comparisons:

$$D = \mu_{ij} - \mu_{i'j'}$$

are

$$\bar{Y}_{ij\cdot} - \bar{Y}_{i'j'\cdot} \pm T\sqrt{\frac{2MS_E}{n}},$$

where

$$T = \frac{1}{\sqrt{2}}q[1-\alpha; ab, (n-1)ab] \qquad \text{(Table B.9)}$$

2° **Bonferroni procedure**

The Bonferroni $1-\alpha$ multiple comparison confidence limits for a group of g pairwise comparisons:

$$\bar{Y}_{ij\cdot} - \bar{Y}_{i'j'\cdot} \pm B\sqrt{\frac{2MS_E}{n}},$$

where

$$B = t[1-\alpha/(2g); (n-1)ab].$$

Multiple Contrasts of Treatment Means

3° **Scheffe procedure**

The joint (1-α) confidence limits for contrasts of the form:

$$L = \sum\sum c_{ij}\mu_{ij}, \qquad \text{where} \quad \sum\sum c_{ij} = 0,$$

are

$$\sum\sum c_{ij}\bar{Y}_{ij\cdot} \pm S * \sqrt{\frac{MS_E}{n}\sum\sum c_{ij}^2},$$

where

$$S^2 = (ab-1)F[1-\alpha; ab-1, (n-1)ab].$$

4° **Bonferroni procedure** (*for small number of contrasts*)

$$\sum\sum c_{ij}\bar{Y}_{ij\cdot} \pm B * \sqrt{\frac{MS_E}{n}\sum\sum c_{ij}^2},$$

where

$$B = t[1-\alpha/(2g); (n-1)ab].$$

Problem 20.7: *Hay fever relief*, CH19PR14.DAT.

Part c. The Scheffe multiple comparison.

```
> scheffe <- function(coef){
            s.value <- sqrt((3*3-1)*qf(1-.10,3*3-1,(4-1)*3*3));
            lowerbd <- sum(coef*cell.mean)-s.value*sqrt(MSE/n*sum(coef^2));
            upperbd <- sum(coef*cell.mean)+s.value*sqrt(MSE/n*sum(coef^2));
            scheffe <- c(lowerbd, upperbd);
            scheffe
            }
```

where `cell.mean` is obtained from

```
> model.tables(aov(data~ingredient1+ingredient2+ingredient1*ingredient2, fever.df),
  type="means")
  Grand mean   7.183333

  ingredient1    1     2     3
              3.883 7.833 9.833

  ingredient2    1     2     3
              4.633 7.933 8.983

  ingredient1:ingredient2
                          ingredient2
            ingredient1  1      2      3
                      1  2.475  4.600  4.575
                      2  5.450  8.925  9.125
                      3  5.975 10.275 13.250

 > cell.mean <- matrix(c(2.475,5.450,5.975,4.600,8.925,10.275,4.575,9.125,13.250),
                      3, 3)        #  Enter a matrix by column.
                                   #  The matrix can also be filled row-wise as
                                   #  matrix(c(2.475, 4.600, ...), 3, 3, byrow=T))
```

Now CI for $L_1 = (\mu_{12} + \mu_{13})/2 - \mu_{11}$ is

```
 > scheffe(matrix(c(-1,.5,.5,0,0,0,0,0,0),3,3, byrow=T))
   [1] 1.526296 2.698704
```

Part d. The Tukey multiple comparison procedure

```
>  q.value <- 4.32   #  In this case, a=3, b=3, n=4, alpha=.1
                     # From Table B.9, q.value=q(1-alpha; a*b, (n-1)*a*b)
                     #                        =q(.9, 3*3, 3*3*3)=4.32
>  T <- q.value/sqrt(2)

>  tukey <- function(x,y){
           tukey <- c(mean(x)-mean(y)-sqrt(2*MS_E/n)*T,
                      mean(x)-mean(y)+sqrt(2*MS_E/n)*T);
                      tukey
                      }
```

Now CI for $\mu_{11} - \mu_{12}$ is

```
> tukey(x[,1][x[,2]==1&x[,3]==1],  x[,1][x[,2]==1&x[,3]==2])
  [1] -2.654090 -1.595910
```

Similarly, for $\mu_{11} - \mu_{13}$

```
> tukey(x[,1][x[,2]==1&x[,3]==1], x[,1][x[,2]==1&x[,3]==3])
  [1] -2.629090 -1.570910
```

There are $\binom{3\times 4}{2} = 66$ pairwise CIs. From the Tukey multiple comparison procedure, the longest mean relief is that with largest cell mean, i.e., μ_{33}.

6 One case per treatment

Question: What happens if $n = 1$?

df for Error(Residuals) $= 0$; $SS_E = 0$; $SS_T = SS_B = SS_A + SS_B + SS_{AB}$.

No-interaction model:

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

where

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

Two-Way ANOVA Table (n=1)

Source of Variation	df	Sum of Squares	Mean Square (MS)	F^*-value	Null hypothesis
Factor A	$a-1$	SS_A	$MS_A = \frac{SSA}{a-1}$	$F^* = \frac{MS_A}{MS_{AB}}$	$\mathcal{H}_0: \ \alpha_1 = \cdots = \alpha_a = 0$
Factor B	$b-1$	SS_B	$MS_B = \frac{SS_B}{b-1}$	$F^* = \frac{MS_B}{MS_{AB}}$	$\mathcal{H}_0: \ \beta_1 = \cdots = \beta_b = 0$
Error	$(a-1)(b-1)$	SS_{AB}	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	σ^2	
Total	$ab-1$	SS_T			

Sums of Squares

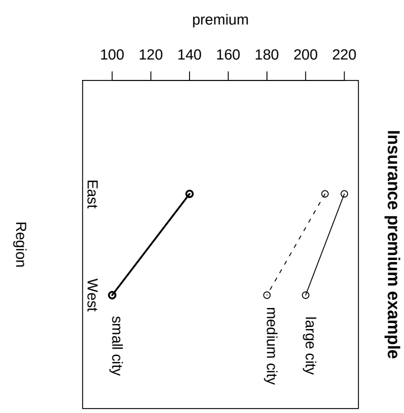
$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$$

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

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

$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 \qquad \text{(Error)}$$

Insurance premium example, p.878



The **R** code for the figure is

```
> y <- read.table("CH21TA02.DAT")
> small <- y[,1][y[,2]==1]
> medium <- y[,1][y[,2]==2]
> large <- y[,1][y[,2]==3]
> region <- c(1,2)

> plot(region, small, type="o", lwd=2,xlim=c(0,3), ylim=c(90,222),
       xlab="Region", ylab="premium", xaxt="n", main="Insurance premium example")

> lines(region, medium,type="o", lty=2 )
> lines(region, large, type="o")

> text(2.5, 102, "small city")
```

```
> text(2.5, 182, "medium city")
> text(2.5, 202, "large city")

> text(1, 90, "East")
> text(2, 90, "West")
```

The ANOVA table is obtained as before.

```
> data <- y[,1]
> size <- y[,2]
> region <- y[,3]
> premium.df <- data.frame(data=data, size=factor(size), region=factor(region))

> anova <- aov(data~size+region, premium.df)
           # compare:  aov(data~size+region+size*region, premium.df)
> summary(anova)
            Df Sum Sq Mean Sq  F value  Pr(>F)
size         2   9300    4650      93   0.01064
region       1   1350    1350      27   0.03510
Residuals    2    100      50
```