Ch. 6.3 Multiple linear regression
Data in Table 6.3 (Carbohydrate data)

Table 6.5 *Analysis of Variance table comparing models (6.6) and (6.7).*

Source variation	Degrees of freedom	Sum of squares	Mean square
Model (6.7)	3	28761.978	
Improvement due to model (6.6)	1	38.359	38.36
Residual	16	567.663	35.48
Total	20	29368.000	

Try to see the connection between this table and the MINITAB table on the next page! Note that it is given in the example that N y-bar^2 = 28275.2

Then do the analysis of Table 6.5 by using R.

MINITAB output, Carbohydrate data

```
Regression Analysis: y versus x2; x3; x1

The regression equation is
y = 37,0 - 0,228 x2 + 1,96 x3 - 0,114 x1

Predictor      Coef  SE Coef      T      P
Constant      36,96    13,07   2,83  0,012
x2         -0,22802  0,08329  -2,74  0,015
x3           1,9577   0,6349   3,08  0,007
x1          -0,1137   0,1093  -1,04  0,314

S = 5,95642   R-Sq = 48,1%   R-Sq(adj) = 38,3%

Analysis of Variance

Source          DF        SS      MS     F      P
Regression       3    525,14  175,05  4,93  0,013
Residual Error  16    567,66   35,48
Total           19   1092,80

Source  DF  Seq SS
x2       1  181,38
x3       1  305,40
x1       1   38,36
```

Two-factor ANOVA

Table 6.9 *Fictitious data for two-factor ANOVA with equal numbers of observations in each subgroup.*

Levels of factor A	Levels of factor B: B_1	B_2	Total
A_1	6.8, 6.6	5.3, 6.1	24.8
A_2	7.5, 7.4	7.2, 6.5	28.6
A_3	7.8, 9.1	8.8, 9.1	34.8
Total	45.2	43.0	88.2

The main hypotheses are:
H_I: there are no interaction effects, i.e., the effects of A and B are additive;
H_A: there are no differences in response associated with different levels of factor A;
H_B: there are no differences in response associated with different levels of factor B.

Model with Sum-to-zero constraint:

$$E(Y_{jkl}) = \mu_{jk} = \mu + \alpha_j + \beta_k + \gamma_{jk}, \quad j = 1,2,3, \ k = 1,2$$

6 free parameters: $\mu, \alpha_1, \alpha_2, \beta_1, \gamma_{11}, \gamma_{21}$.

$$\begin{aligned} \alpha_3 &= -\alpha_1 - \alpha_2 \\ \beta_2 &= -\beta_1 \\ \gamma_{12} &= -\gamma_{11} \\ \gamma_{22} &= -\gamma_{21} \\ \gamma_{31} &= -\gamma_{11} - \gamma_{21} \\ \gamma_{32} &= \gamma_{11} + \gamma_{21} \end{aligned}$$

Design-matrix with one observation per cell:

	μ	α_1	α_2	β_1	γ_{11}	γ_{21}
A_1B_1	1	1	0	1	1	0
A_1B_2	1	1	0	−1	−1	0
A_2B_1	1	0	1	1	0	1
A_2B_2	1	0	1	−1	0	−1
A_3B_1	1	−1	−1	1	−1	−1
A_3B_2	1	−1	−1	−1	1	1

```
> X
      [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]    1    1    0    1    1    0
 [2,]    1    1    0    1    1    0
 [3,]    1    1    0   -1   -1    0
 [4,]    1    1    0   -1   -1    0
 [5,]    1    0    1    1    0    1
 [6,]    1    0    1    1    0    1
 [7,]    1    0    1   -1    0   -1
 [8,]    1    0    1   -1    0   -1
 [9,]    1   -1   -1    1   -1   -1
[10,]    1   -1   -1    1   -1   -1
[11,]    1   -1   -1   -1    1    1
[12,]    1   -1   -1   -1    1    1
```

```
> t(X)%*%X
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]   12    0    0    0    0    0
[2,]    0    8    4    0    0    0
[3,]    0    4    8    0    0    0
[4,]    0    0    0   12    0    0
[5,]    0    0    0    0    8    4
[6,]    0    0    0    0    4    8
```



```
> y
      [,1]
 [1,]  6.8
 [2,]  6.6
 [3,]  5.3
 [4,]  6.1
 [5,]  7.5
 [6,]  7.4
 [7,]  7.2
 [8,]  6.5
 [9,]  7.8
[10,]  9.1
[11,]  8.8
[12,]  9.1
> b <- solve(t(X)%*%X)%*%t(X)%*%y
> b
           [,1]
[1,]  7.3500000
[2,] -1.1500000
[3,] -0.2000000
[4,]  0.1833333
[5,]  0.3166667
[6,]  0.1166667
```

```
> X1 <- X[,1]
> X1
 [1] 1 1 1 1 1 1 1 1 1 1 1 1
> X2 <- cbind(X[,2],X[,3])
> X2
      [,1] [,2]
 [1,]    1    0
 [2,]    1    0
 [3,]    1    0
 [4,]    1    0
 [5,]    0    1
 [6,]    0    1
 [7,]    0    1
 [8,]    0    1
 [9,]   -1   -1
[10,]   -1   -1
[11,]   -1   -1
[12,]   -1   -1
> X3 <- X[,4]
> X3
 [1]  1  1 -1 -1  1  1 -1 -1  1  1 -1 -1
> X4 <- cbind(X[,5],X[,6])
> X4
      [,1] [,2]
 [1,]    1    0
 [2,]    1    0
 [3,]   -1    0
 [4,]   -1    0
 [5,]    0    1
 [6,]    0    1
 [7,]    0   -1
 [8,]    0   -1
 [9,]   -1   -1
[10,]   -1   -1
[11,]    1    1
[12,]    1    1
```



```
> b1 <- solve(t(X1)%*%X1)%*%t(X1)%*%y
> b1
     [,1]
[1,] 7.35
> S1 <- t(b1)%*%t(X1)%*%y
> S1
       [,1]
[1,] 648.27
> b2 <- solve(t(X2)%*%X2)%*%t(X2)%*%y
> b2
      [,1]
[1,] -1.15
[2,] -0.20
> S2 <- t(b2)%*%t(X2)%*%y
> S2
      [,1]
[1,] 12.74
> b3 <- solve(t(X3)%*%X3)%*%t(X3)%*%y
> b3
          [,1]
[1,] 0.1833333
> S3 <- t(b3)%*%t(X3)%*%y
> S3
          [,1]
[1,] 0.4033333
> b4 <- solve(t(X4)%*%X4)%*%t(X4)%*%y
> b4
          [,1]
[1,] 0.3166667
[2,] 0.1166667
> S4 <- t(b4)%*%t(X4)%*%y
> S4
         [,1]
[1,] 1.206667
> Res <- t(y)%*%y-S1-S2-S3-S4
> Res
     [,1]
[1,] 1.48
```

F-tests for HI, HA, HB

```
> FI <- (S4/2)/(Res/6)
> FA <- (S2/2)/(Res/6)
> FB <- (S3/1)/(Res/6)
> FI
         [,1]
[1,] 2.445946
> FA
         [,1]
[1,] 25.82432
> FB
         [,1]
[1,] 1.635135
```


Table 6.11 *ANOVA table for data in Table 6.8.*

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Mean	1	648.2700		
Levels of A	2	12.7400	6.3700	25.82
Levels of B	1	0.4033	0.4033	1.63
Interactions	2	1.2067	0.6033	2.45
Residual	6	1.4800	0.2467	
Total	12	664.1000		