17.

THE K-FACTOR DESIGN

17.1 INTRODUCTION

Suppose we have k factors, factor i takes a_i levels and we take n observations from treatment $(i_1, i_2, \ldots, i_k)$. Then for a single observation we have

$$E[y] = \sum_{i_1=1}^{a_1} \sum_{i_k=1}^{a_k} \beta_{i_1} \cdots i_k \quad X_{i_1} \cdots i_k$$

where $X_{i_1} \cdots i_k = 1$ if y is a response from treatment $(i_1, i_2, \ldots, i_k)$ and is 0 otherwise. Listing the observations in natural order we have

$$E[\mathbf{y}] = \mathbf{X}\beta = (X_{11\ldots1} \quad \cdots \quad X_{a_1 \cdots a_k})\ \beta$$

where $X_{1\ldots11}$ has 1 in the first n+1 through 2n entries and 0 elsewhere, etc.

Note that $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = \left(\frac{T_{11\ldots1}}{n}, \ \cdots \ , \frac{T_{a_1 \cdots a_k}}{n} right)'$

and $T_{i_1 \cdots i_k}$ is the total of all observations from treatment $(i_1, \ldots, i_k)$.

(a) The model with k-factors is

A_1: $A_{11}, \ldots, A_{a_1 1}$

A_2: $A_{12}, \ldots, A_{a_2 2}$

.
.
.

A_k: $A_{1_k}, \ldots, A_{a_k k}$

which gives the linear model

$$\boxed{E[y] = \beta_{1\ldots 1}\ x_{1\ldots 1} + \beta_{21\ldots 1}\ x_{21\ldots 1} + \cdots + \beta_{abc}\ X_{abc}}$$

Where $x_{i_1 \cdots i_k} = 1$ if treatment $A_{i_1} \cdots A_{i_k k}$ is applied and is 0 otherwise.
If we apply $A_{i_1 i} \cdots A_{i_k k}$, n times we obtain

$E[\mathbf{y}] = \mathbf{X}\beta$ where

$$\mathbf{X} = \begin{vmatrix} 1 & 0 & . & . & . & 0 \\ 1 & 0 & . & . & . & 0 \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ n & 0 & & & & . \\ 0 & 1 & & & & . \\ 0 & 1 & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ & n & & & & 0 \\ & . & & & & 1 \\ & . & & & & 1 \\ & . & & & & . \\ & . & & & & . \\ & . & & & & . \\ 0 & 0 & & & & n \end{vmatrix}$$

and

$$y = \begin{vmatrix} \text{n} & \text{obs} & \text{from} & A_{11} \\ \text{n} & \text{obs} & \text{from} & A_{12} \\ & & . & \\ & & . & \\ & & . & \\ \text{n} & \text{obs} & \text{from} & A_{a_k k} \end{vmatrix}$$

the least squares estimator is given by

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y} = \begin{bmatrix} \frac{T_{1\ldots1}}{n} \\ . \\ . \\ . \\ \frac{T_{21\ldots1}}{n} \end{bmatrix} = \begin{bmatrix} \bar{y}_{1\ldots1} \\ . \\ . \\ . \\ \bar{y}_{21\ldots1} \end{bmatrix}$$

and $\alpha = \mathbf{C}\beta$ where $a_{i_1 \cdots i_k} = (\mathbf{C}_{i_1 i} \times \cdots \times \mathbf{C}_{i_k k})\, \beta$

First test for interactions of order k-1, then k-2, etc. Then obtain the following ANOVA for single degrees of freedom.

ANOVA (coarse table)

Source	DF	SS
Mean	1	$na^2_{1\ldots1}$
Main Effects		
.		
.		
.		
First order Interactions		
.		
.		
.		
(k-1) order Interactions		
Error	(n-1) $a_1 \cdots a_k$	
	$na_1 \cdots a_k$	$\mathbf{y'y}$

If we collect appropriate degrees of freedom together we obtain the

ANOVA (fine table)

Source	DF	SS
Mean	1	$\frac{G^2}{na_1 \cdots a_k} = SL(M)$
A_1	$a_1 - 1$	$\sum_{i=1}^{a_1} \frac{T_i^2}{na_2 \cdots a_k} - SL(M) = SL(A)$
.		
.		
.		
A_k	$a_k - 1$	$\sum_{i=1}^{a_k} \frac{T_i^2}{na_1 \cdots a_{k-1}} - SL(M) = SL(A_k)$
$A_1 \ x \ A_2$	$(a_1-1)(a_2-1)$	$\sum_{i=1}^{a_1}\sum_{j=1}^{a_2} \frac{T_{ij}^2}{na_3 \cdots a_k} - SL(A_1) - SL(A_2) - SL(M)$
.		
.		
.		
$A_1 \ x \ A_2 \ x \ A_3$	$(a_1-1)(a_2-1)(a_3-1)$	1 (see below)
.		
.		
.		
$A_1 \ x \ \cdots \ x \ A_k$	$(a_1-1) \ \cdots \ (a_k-1)$	2 (see below)
Error	$(n-1)a_1 \ \cdots \ a_k$	subtraction
	$na_1 \ \cdots \ a_k$	$\mathbf{y'y}$

where

$$1 = \sum_{i_1=1}^{a_1}\sum_{i_2=1}^{a_2}\sum_{i_3=1}^{a_3} \frac{T^2{}_{i_1 i_2 i_3}}{na_4 \cdots a_k} - SL(A_1 \ x \ A_2) - SL(A_1 \ x \ A_3) - SL(A_2 \ x \ A_3) - SL(A_2) - SL(A_3) - SL(M)$$

$$2 = \sum_{i_1=1}^{a_1} \cdots \sum_{i_k=1}^{a_k} \frac{T_{i_1}{}^2}{n} - SL\,(A_1 \ x \ \cdots \ x \ A_{k-1}) - \ \cdots \ - SL\,(M)$$

Note: If n=1 then we have no degrees of freedom for estimating error, however for k ≥ 4 higher order interactions are sometimes assumed not to exist and these degrees of freedom become available for error.

Example of a 4 x 3 x 2 Factorial Experiment

A has 4 levels

B has 3 levels

C has 2 levels

Thus there are 4 x 3 x 2 = 24 treatment combinations. Say there are 5 observations on each treatment. This gives the following table.

ANOVA

Source		DF
Main Effects	A	3
	B	2
	C	1
Interactions	A x B	6
	A x C	3
	B x C	2
	A x B x C	6
Error	(Within Treat)	96
Total (corrected for mean)		

where $\frac{G^2}{N} = 4296.03 = \frac{(718)^2}{120}$

with Error equal to 364.80

The steps necessary for a hand calculation are briefly summarized below.

Step 1: Form the following table.
Each cell is the total of 5 observations

		A_1	A_2	A_3	A_4	Σ
C_1	B_1	21	33	35	44	133
	B_2	24	32	36	41	133
	B_3	18	30	37	49	134
	Σ	63	95	108	134	400
C_2	B_1	16	28	35	36	115
	B_2	14	26	33	34	107
	B_3	13	23	27	33	96
	Σ	43	77	95	103	318
	Σ	106	172	203	237	718

Step 2: Perform the following calculations

$$SS_T = \frac{(21)^2}{5} + \frac{(33)^2}{5} + \cdots + \frac{(33)^2}{5} - \frac{(718)^2}{120}$$

$$= 4691.20 - 4296.03 = 395.17$$

$$SS_C = \frac{(400)^2}{60} + \frac{(318)^2}{60} - \frac{(718)^2}{120}$$

$$= 4352.07 - 4296.03 = 56.04$$

$SS_A = 4606.60 - 4296.03 = 310.57$

To calculate the A x C interaction - look at cells of the A x C table.

$$SS_{cell} = \frac{(63)^2}{15} + \frac{(95)^2}{15} + \cdots + \frac{(103)^2}{15} - \frac{(718)^2}{120}$$

$$= 4668.40 - 4296.03 = 372.37$$

	A_1	A_2	A_3	A_4	$\sum$
C_1	63	95	108	134	400
C_2	43	77	95	103	318
$\sum$	106	172	203	237	718

$$SS_{AC} = SS_{cell} - SS_C - SS_A$$
$$= 372.37 - 56.04 - 310.57 = 5.76$$

For the A x B interaction follow the same technique.

$$SS_B = \frac{(248)^2}{40} + \frac{(240)^2}{40} + \frac{(230)^2}{40} - \frac{(718)^2}{120}$$

$$= 4300.10 - 4296.03 = 4.07$$

$$SS_{cell} = \frac{(37)^2}{10} + \frac{(61)^2}{10} + \cdots + \frac{(82)^2}{10} - \frac{(718)^2}{120}$$

$$= 4617.40 - 4296.03 = 321.37$$

$SS_{AB} = 321.37 - 4.07 - 310.57 = 6.73$

for the B x C interaction

$$SS_{cell} = \frac{(133)^2}{20} + \frac{(133)^2}{20} + \cdots + \frac{(96)^2}{20} - \frac{(718)^2}{120}$$

$$= 4361.21 - 4296.03 = 65.17$$

Then by subtraction we have

$$SS_{BxC} = 65.17 - 56.04 - 4.04 = 5.06$$

$$SS_{ABC} = SS_T - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$= 395.17 - 310.57 - 4.07 - 56.04 - 6.73 - 5.76 - 5.06$$

$$= 395.17 - 388.23 = 6.94$$

The complete ANOVA table is given below.

ANOVA

Source	DF	SS	MS	F
A	3	310.52	103.52	27.24·
B	2	4.07	2.04	
C	1	56.04	56.04	14.75·
AxB	6	6.73	1.12	
AxC	3	5.76	1.92	
BxC	2	5.06	2.57	
AxBxC	6	6.94	1.16	
Error	96	364.80	3.80	
Total		759.97		

where · indicates the P-value if < .05. Note the above was an example using a hand calculation method. A computer implementation for the multifactor design will be given at the end of the chapter.

17.2 2^k FACTORIAL DESIGN

Consider a design where we have k factors each taking two (2) levels. For each factor there is essentially one contrast $\left(\frac{1}{\sqrt{2}} - \frac{-1}{\sqrt{2}}\right)$ and accordingly the analysis of such designs is relatively straightforward. Since such designs are of such importance a special notation has been developed. As before the factors are denoted by capitals $A:A_1,A_2$; $B:B_1,B_2$; $C:C_1,C_2$; etc. And the total of all observations receiving treatment $A_i\,B_j\,C_k$ is denoted by $a^{i-1}\,b^{j-1}\,c^{k-1}$ where $a^0=1$ and $a^1=a$. Level 1 of a factor is called the low level and level 2 called the high level.

Example 1 2^2 **-** factorial

There are a total of 4 treatments.

$$a^0\ b^0 = (1)$$
$$a^1\ b^0 = a$$
$$a^0\ b^1 = b$$
$$a^1\ b^1 = ab$$

Example 2 2^3 **-** factorial

There are a total of 8 treatments.

$$a^0\ b^0\ c^0 = (1)$$
$$a^1\ b^0\ c^0 = a$$
$$a^0\ b^1\ c^0 = b$$
$$a^1\ b^1\ c^0 = ab$$
$$a^0\ b^0\ c^1 = c$$
$$a^1\ b^0\ c^1 = ac$$
$$a^0\ b^1\ c^1 = bc$$
$$a^1\ b^1\ c^1 = abc$$

The treatment in the above examples are written in what is known as *natural order*.

As mentioned above the contrast matrix for the full design is of the form:

$$C = \left|\begin{array}{ccccccccccc|cc}
 & & & & & & & & & & & \text{Divisor} & \text{Source} \\
\hline
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & \ldots & 1 & 1 & \sqrt{2}^k & \text{Mean} \\
1 & -1 & 1 & -1 & 1 & -1 & & & \ldots & 1 & -1 & \sqrt{2}^k & \text{A} \\
1 & 1 & -1 & -1 & 1 & -1 & & & & & & \sqrt{2}^k & \text{B} \\
1 & -1 & 1 & -1 & 1 & -1 & & & & & & \sqrt{2}^k & \text{A x B} \\
. & & & & & . & & & & & . & . & . \\
. & & & & & . & & & & & . & . & . \\
. & & & & & . & & & & & . & . & .
\end{array}\right.$$

Note that the rows of C are $\pm$ 1's divided by $\sqrt{2}^{-k}$.
The quantities we wish to make inferences about are given by $\alpha = C\beta$ and the elements of α are often denoted (1, A, B, AB, ...)

Example 1 (cont'd)

Source						Divisor
Mean		1	1	1	1	2
A	C =	1	-1	1	-1	2
B		1	1	1	-1	2
AxB		1	-1	-1	1	2

$$\alpha \;= (\beta_{11} + \beta_{21} + \beta_{12} + \beta_{22}) / 2$$
$$\alpha_{21} = (\beta_{11} - \beta_{21} + \beta_{12} - \beta_{22}) / 2$$
$$\alpha_{12} = (\beta_{11} + \beta_{21} - \beta_{12} - \beta_{22}) / 2$$
$$\alpha_{22} = (\beta_{11} + \beta_{21} - \beta_{12} + \beta_{22}) / 2$$
$$\alpha_{11} = ((1) + a + b + ab) / 2n$$
$$\alpha_{21} = ((1) - a + b - ab) / 2n$$
$$\alpha_{12} = ((1) + a - b - ab) / 2n$$
$$\alpha_{22} = ((1) - a - b + ab) / 2n$$

Then the ANOVA table takes the form

Source	DF	SS
Mean	1	$na_{11}{}^2$
A	1	$na_{21}{}^2$
B	1	$na_{12}{}^2$
AB	1	$na_{22}{}^2$
Error	4(n-1)	subtraction
Total	abn	$\mathbf{y'y}$

17.3 YATES ALGORITHM

Yates Algorithm is a fast way of performing the calculations for a 2^k factorial. We illustrate the method with a 2^3 - factorial.

Step 1	Step 2
write down totals in natural order	form first 4 entries by adding successive elements and last 4 by subtracting 2nd element from first
(1)	(1) + a
a	b + ab
b	c + ac
ab	bc + abc
c	(1) - a
ac	b - ab
bc	c - ac
abc	bc - abc

Step 3 (repeat)	Step 4 (repeat)	Source
(1) + a + b + ab	(1) + a + b + ab + c + ac + bc + abc	(mean)
c + ac + bc + abc	(1) - a + b - ab + c - ac + bc - abc	(A)
(1) - a + b - ab	(1) + a - b - ab + c - ac - bc - abc	(B)
c - ac + bc - abc	(1) - a - b + ab + c - ac - bc - abc	(C)
(1) + a - b - ab	(1) + a + b + ab - c - ac - bc - abc	(AB)
c + ac - bc - abc	(1) - a + b - ab - c + ac - bc - abc	(AC)
(1) - a - b + ab	(1) + a - b - ab - c - ac + bc + abc	(BC)
c - ac - bc + abc	(1) - a - b + ab - c + ac + bc - abc	(ABC)

In general we are required to repeat the step k times.

17.4 Example : An experimental psychologist is interested in determining whether verbal retention is influenced by the following factors

A: number of presentations: once, twice
B: mode of presentation: visual, auditory
C: time of presentation: immediate, delayed

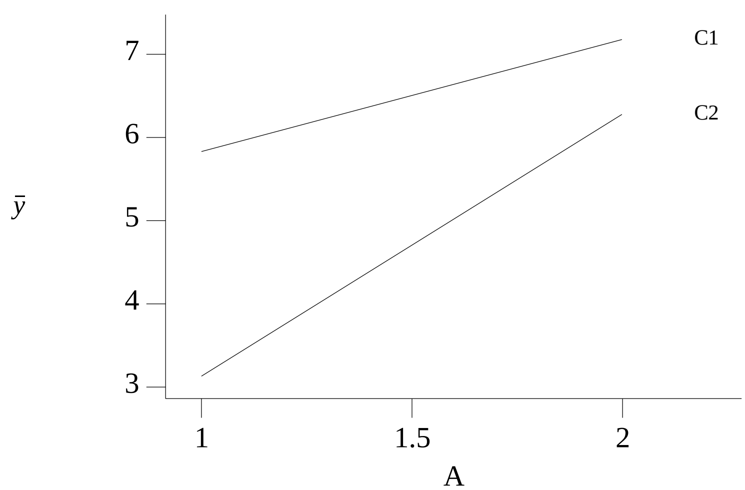
Ten subjects are run under each treatment condition. The results are listed below:

A_1				A_2			
B_1		B_2		B_1		B_2	
C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
8	4	4	4	9	7	7	7
7	4	8	2	9	7	6	6
4	5	7	2	8	6	7	5
6	2	5	2	8	9	6	5
6	4	6	1	8	7	6	5
4	4	6	3	8	7	6	4
4	5	5	2	7	6	6	6
6	4	6	2	8	6	7	6
8	4	5	2	6	8	8	7
7	4	5	3	6	6	8	6
60	40	57	23	77	69	67	57

ANOVA	DF	SS	MS	F
A	1	101.25		
B	1	22.05		20.05*
C	1	64.80		
AxB	1	.05		
AxC	1	16.20		14.73*
BxC	1	3.20		2.91
AxBxC	1	1.80		1.64
Error	72	79.40	1.10	
Total	79	288.75		

where * indicates p < .05.

A graph of the A x C interaction is given below.



17.5 SPSSX COMPUTER PROGRAM

```
data list/ A 1 B 3 C 5 Y 7
begin data
1 1 1 8
1 1 1 7
. . . .
. . . .
. . . .
2 2 2 6
end data
manova y by A B C (1,2)/
   design = A, B, C, A by B, A by C,
       B by C, A by B by C/
   print=cellinfo(means)
   omeans=TABLES(A,B,C,A by C)/
   residuals = casewise plot/
finish
```

*note that since there are only 2 levels to each factor we may omit the partition and contrast statements of SPSSX.

```
Cell Means and Standard Deviations
Variable .. Y
     FACTOR           CODE                 Mean  Std. Dev.          N   95 percent Conf. Interval

 A                     1
  B                     1
   C                     1               6.000      1.563        10      4.882      7.118
   C                     2               4.000       .816        10      3.416      4.584
  B                     2
   C                     1               5.700      1.160        10      4.871      6.529
   C                     2               2.300       .823        10      1.711      2.889
 A                     2
  B                     1
   C                     1               7.700      1.059        10      6.942      8.458
   C                     2               6.900       .994        10      6.189      7.611
  B                     2
   C                     1               6.700       .823        10      6.111      7.289
   C                     2               5.700       .949        10      5.021      6.379
For entire sample                        5.625      1.912        80      5.200      6.050

             Combined Observed Means for A
             Variable .. Y
                         A
                         1        WGT.     4.50000
                                UNWGT.     4.50000
                         2        WGT.     6.75000
                                UNWGT.     6.75000

             Combined Observed Means for B
             Variable .. Y
                         B
                         1        WGT.     6.15000
                                UNWGT.     6.15000
                         2        WGT.     5.10000
                                UNWGT.     5.10000

             Combined Observed Means for C
             Variable .. Y
                         C
                         1        WGT.     6.52500
                                UNWGT.     6.52500
                         2        WGT.     4.72500
                                UNWGT.     4.72500

             Combined Observed Means for A BY C
             Variable .. Y
                                     A           1           2
                         C
                         1        WGT.     5.85000     7.20000
                                UNWGT.     5.85000     7.20000
                         2        WGT.     3.15000     6.30000
                                UNWGT.     3.15000     6.30000
```

The ANOVA table for the model is given below. The A by C interaction is significant, $F(1,72) = 14.69$, $p<.001$. A graph of the means would indicate the type of interaction. The main effects of A, B, and C are also significant.

Tests of Significance for Y using UNIQUE sums of squares

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	79.40	72	1.10		
A	101.25	1	101.25	91.81	.000
B	22.05	1	22.05	19.99	.000
C	64.80	1	64.80	58.76	.000
A BY B	.05	1	.05	.05	.832
A BY C	16.20	1	16.20	14.69	.000
B BY C	3.20	1	3.20	2.90	.093
A BY B BY C	1.80	1	1.80	1.63	.205

The residuals look reasonable.

```
Plots of Observed, Predicted, and Residual Case Values (CONT.)
             Predicted Values VS Std Resid. for Y

         +----+----+----+----+----+----+----+----+-+
         :                                         :
         :                                         :
         :                         1               :
       2 +                           2    1        +
         :                                         :
         :   1                                     :
         :                                         :
         :                         3     2     2   :
R      1 +              2            2    1        + R
E        :                                         : E
S        :   2                                     : S
I        :                                         : I
D        :                         7     3     5   : D
U      0 +              7            3    4        + U
A        :   6                                     : A
L        :                                         : L
S        :                         7     5     1   : S
         :                                4        :
      -1 +                                         +
         :   1                                     :
         :                                         :
         :                         2           2   :
         :                                         :
      -2 +              1            3             +
         :                                         :
         :                                         :
         +----+----+----+----+----+----+----+----+-+
            2.4         4       5.6       7.2
                 3.2       4.8       6.4         8

                          PREDICTED
```

The normal probability plot looks reasonable.

```
Plots of Observed, Predicted, and Residual Case Values (CONT.)
                         Normal Plot

         +----+----+----+----+----+----+----+----+-+
         :                                    1   :
         :                                        :
E      2 +                                   1    + D
X        :                                 1      : E
P        :                              1  1      : V
E        :                           3            : I
C        :                           3            : A
T      1 +                         211            + T
E        :                       2 2              : I
D        :                     6                  : O
         :                     6                  : N
N        :                    33                  :
O      0 +                   51                   + F
R        :                 1 5                    : R
M        :             1   5                      : O
A        :             6                          : M
L        :             4                          :
      -1 +            22                          + N
V        :         1  2                           : O
A        :      3                                 : R
L        :   1  1                                 : M
U        :   1                                    : A
E     -2 +   1                                    + L
         :                                        :
         :   1                                    :
         :                                        :
         +----+----+----+----+----+----+----+----+-+
           -1.8      -.6        .6       1.8
                -1.2        0       1.2       2.4

                            Y
```