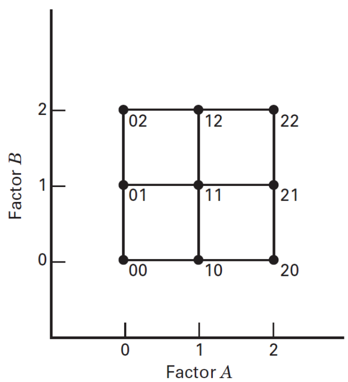
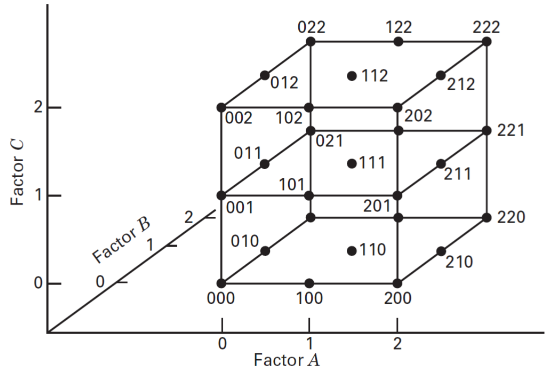
Lesson 2.3 3^k Factorial ExperimentS

Introduction

The 3^k factorial design is a factorial arrangement with k factors, each at three levels. Factors and interactions will be denoted by capital letters. We will refer to the three levels of the factors as *low*, *intermediate*, and *high*. Several different notations may be used to represent these factor levels; one possibility is to represent the factor levels by the digits *0 (low)*, *1 (intermediate)*, and *2 (high)*. Each treatment combination in the 3k design will be denoted by k digits, where the first digit indicates the level of factor A, the second digit indicates the level of factor B, . . . , and the k^th digit indicates the level of factor K. For example, in a 3^2 design, 00 denotes the treatment combination corresponding to A and B both at the low level, and 01 denotes the treatment combination corresponding to A at the low level and B at the intermediate level. The figures below show the geometry of the 3^2 (left) and the 3^3 (right) design, respectively.





In the 3^k system of designs, when the factors are quantitative, we often denote the low, intermediate, and high levels by -1, 0, and +1, respectively. This facilitates fitting a regression model relating the response to the factor levels. For example, consider the 3^2 design, and let x_1 represent factor A and x_2 represent factor B. A regression model relating the response y to x_1 and x_2 that is supported by this design is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \boldsymbol{\epsilon} \tag{1}$$

Notice that the addition of a third factor level allows the relationship between the response and design factors to be modeled as a quadratic.

The 3^2 design

The simplest design in the 3^k system is the 3^2 design, which has two factors, each at three levels. The treatment combinations for this design are shown in the previous figure (left). Because there are $3^2 = 9$ treatment combinations, there are eight degrees of freedom between these treatment combinations. The main effects of A and B each have two degrees of freedom, and the AB interaction has four degrees of freedom. If there are n replicates, there will be $3^2 n - 1$ total degrees of freedom and $3^2(n-1)$ degrees of freedom for error.

The sums of squares for A, B, and AB may be computed by the usual methods for factorial designs discussed in Stat 141. Each main effect can be represented by a linear and a quadratic component, each with a single degree of freedom, as demonstrated in Equation 1. Of course, this is meaningful only if the factor is quantitative. The two-factor interaction AB may be partitioned in two ways. Suppose that both factors A and B are quantitative. The first method consists of subdividing AB into the four single-degree-of-freedom components corresponding to $AB_{L\times L}$, $AB_{L\times Q}$, $AB_{Q\times L}$, and $AB_{Q\times Q}$. This can be done by fitting the terms $\beta_{12}x_1x_2$, $\beta_{122}x_1x_2^2$, $\beta_{112}x_1^2x_2$, and $\beta_{1122}x_1^2x_2^2$, respectively.

Consider the following example.

An example

The effective life of a cutting tool installed in a numerically controlled machine is thought to be affected by the cutting speed and the tool angle. Three speeds and three angles are selected, and a 3^2 factorial experiment with two replicates is performed. The data is contained in the Excel file *three-to-the-k factorial.xlsx*.

```r
library(readxl)
library(tidyverse)
df <- read_excel("three-to-the-k factorial.xlsx")

df <- df %>%
  mutate(Angle = factor(Angle),
         Speed = factor(Speed),
         x1x2=x1*x2,
         x1sq = x1^2,
         x2sq = x2^2,
         x1x2sq = x1*x2^2,
         x1sqx2 = x1^2*x2,
         x1sqx2sq = x1^2*x2^2)
fit <- lm(Life ~ x1 + x2 + x1sq + x2sq + x1x2 + x1x2sq + x1sqx2 + x1sqx2sq, data =df)
anova(fit)
```

```
## Analysis of Variance Table
##
```

```
## Response: Life
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1         1  8.333   8.333  5.7692 0.0397723 *
## x2         1 21.333  21.333 14.7692 0.0039479 **
## x1sq       1 16.000  16.000 11.0769 0.0088243 **
## x2sq       1  4.000   4.000  2.7692 0.1304507
## x1x2       1  8.000   8.000  5.5385 0.0430650 *
## x1x2sq     1 42.667  42.667 29.5385 0.0004137 ***
## x1sqx2     1  2.667   2.667  1.8462 0.2073056
## x1sqx2sq   1  8.000   8.000  5.5385 0.0430650 *
## Residuals  9 13.000   1.444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

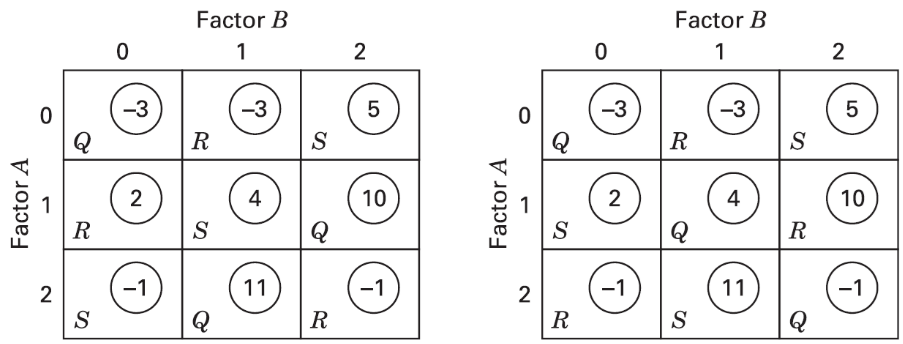
```r
mod1 <- aov(Life ~ Angle + Speed + Angle:Speed, data = df)
anova(mod1)
```

```
## Analysis of Variance Table
##
## Response: Life
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Angle        2 24.333 12.1667  8.4231 0.008676 **
## Speed        2 25.333 12.6667  8.7692 0.007703 **
## Angle:Speed  4 61.333 15.3333 10.6154 0.001844 **
## Residuals    9 13.000  1.4444
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have $SS_{AB_{L\times L}} = 8.00$, $SS_{AB_{L\times Q}} = 42.67$, $SS_{AB_{Q\times L}} = 2.67$, and $SS_{AB_{Q\times Q}} = 8.00$. Because this is an orthogonal partitioning of AB, note that $SS_{AB} = SS_{AB_{L\times L}} + SS_{AB_{L\times Q}} + SS_{AB_{Q\times L}} + SS_{AB_{Q\times Q}} = 61.34$.

The second method is based on orthogonal Latin squares. Two Latin squares are orthogonal if one square is superimposed on the other, each letter (number) in the first square will appear exactly once with each letter in the second square **Click HERE**.

This method does not require that the factors be quantitative, and it is usually associated with the case where all factors are qualitative. Consider the totals of the treatment combinations for the data in the previous example on effective life of a cutting tool. These totals are shown in the figure below as the circled numbers in the squares. The two factors A and B correspond to the rows and columns, respectively, of a 3×3 Latin square. verify that the Latin squares in the figure are indeed orthogonal.



(a) (b)

The totals for the letters in the (a) square are $Q = 18$, $R = -2$, and $S = 8$, and the sum of squares between these totals is

$$\frac{18^2 + (-2)^2 + 8^2}{(3)(2)} - \frac{24^2}{9(2)} = 33.34$$

,

with two degrees of freedom.

Similarly, the letter totals in the (b) square are $Q = 0$, $R = 6$, and $S = 18$, and the sum of squares between these totals is

$$\frac{0^2 + 6^2 + 18^2}{(3)(2)} - \frac{24^2 2}{(9)(2)} = 28.00$$

,

with two degrees of freedom.

Note that the sum of these two components is

$$33.34 + 28.00 = 61.34 = SS_{AB}$$

with $2 + 2 = 4$ degrees of freedom.

In general, the sum of squares computed from square (a) is called the AB **component** of interaction, and the sum of squares computed from square (b) is called the AB^2 **component** of interaction. The components AB and AB^2 each have two degrees of freedom. This terminology is used because if we denote the levels (0, 1, 2) for A and B by x_1 and x_2, respectively, then we find that the letters occupy cells according to the following pattern:

Square (*a*)	**Square (*b*)**
Q: $x_1 + x_2 = 0 \pmod 3$	Q: $x_1 + 2x_2 = 0 \pmod 3$
R: $x_1 + x_2 = 1 \pmod 3$	S: $x_1 + 2x_2 = 1 \pmod 3$
S: $x_1 + x_2 = 2 \pmod 3$	R: $x_1 + 2x_2 = 2 \pmod 3$

For example, in square (b), note that the middle cell corresponds to $x_1 = 1$ and $x_2 = 1$; thus,

$$x_1 + 2x_2 = 1 + (2)(1) = 3(mod3) = 0,$$

and Q would occupy the middle cell.

The AB and AB^2 components of the AB interaction have no actual meaning and are usually not displayed in the analysis of variance table. However, this rather arbitrary partitioning of the AB interaction into two orthogonal two-degree-of-freedom components is very useful in constructing more complex designs. Also, there is no connection between the AB and AB^2 components of interaction and the sums of squares for $AB_{L\times L}$, $AB_{L\times Q}$, $AB_{Q\times L}$, and $AB_{Q\times Q}$.

The AB and AB^2 components of interaction may be computed another way. Consider the treatment combination totals in either square in the example. If we add the data by diagonals downward from left to right, we obtain the totals $-3 + 4 + (-1) = 0$, $-3 + 10 + (-1) = 6$, and $5 + 11 + 2 = 18$. The sum of squares between these totals is 28.00 (AB^2). Similarly, the diagonal totals downward from right to left are $5 + 4 + (-1) = 8$, $-3 + 2 + (-1) = -2$, and $-3 + 11 + 10 = 18$. The sum of squares between these totals is 33.34 (AB). Yates called these components of interaction as the I and J components of interaction, respectively. We use both notations interchangeably; that is,

$$I(AB) = AB^2$$
$$J(AB) = AB$$

The 3^3 Design

Now suppose there are three factors (A, B, and C) under study and that each factor is at three levels arranged in a factorial experiment. This is a 3^3 factorial design. The 27 treatment combinations have 26 degrees of freedom. Each main effect has two degrees of freedom, each two factor interaction has four degrees of freedom, and the three-factor interaction has eight degrees of freedom. If there are n replicates, there are $n3^3 - 1$ total degrees of freedom and $3^3(n - 1)$ degrees of freedom for error.

The sums of squares may be calculated using the standard methods for factorial designs. In addition, if the factors are quantitative, the main effects may be partitioned into linear and quadratic components, each with a single degree of freedom. The two-factor interactions may be decomposed into $linear \times linear$, $linear \times quadratic$, $quadratic \times linear$, and $quadratic \times quadratic$ effects. Finally, the three-factor interaction ABC can be partitioned into eight single-degree-of-freedom components corresponding to $linear \times linear \times linear$, $linear \times linear \times quadratic$, and so on. Such a breakdown for the three-factor interaction is generally not very useful.

It is also possible to partition the two-factor interactions into their I and J components. These would be designated AB, AB^2, AC, AC^2, BC, and BC^2, and each component would have two degrees of freedom. As in the 32 design, these components have no physical significance.

The three-factor interaction ABC may be partitioned into four orthogonal two degrees-of-freedom components, which are usually called the W, X, Y, and Z components of the interaction. They are also referred to as the AB^2C^2, AB^2C, ABC^2, and ABC components of the ABC interaction, respectively. The two notations are used interchangeably; that is,

$$
\begin{aligned}
W(ABC) &= AB^2C^2 \\
X(ABC) &= AB^2C \\
Y(ABC) &= ABC^2 \\
Z(ABC) &= ABC
\end{aligned}
$$

Note that no first letter can have an exponent other than 1. Like the I and J components, the W, X, Y, and Z components have no practical interpretation. They are, however, useful in constructing more complex designs.

The General 3^k Design

The concepts utilized in the 3^2 and 3^3 designs can be readily extended to the case of k factors, each at three levels, that is, to a ^{k} factorial design. The usual digital notation is employed for the treatment combinations, so 0120 represents a treatment combination in a 3^4 design with A and D at the low levels, B at the intermediate level, and C at the high level. There are 3^k treatment combinations, with $3^k - 1$ degrees of freedom between them. These treatment combinations allow sums of squares to be determined for k main effects, each with two degrees of freedom; $\binom{n}{2}$ two-factor interactions, each with four degrees of freedom; . . . ; and one k-factor interaction with 2^k degrees of freedom. If there are n replicates, there are $n3^k - 1$ total degrees of freedom and $3^k(n-1)$ degrees of freedom for error.

The size of the design increases rapidly with k. For example, a 3^3 design has 27 treatment combinations per replication, a 3^4 design has 81, a 3^5 design has 243, and so on. Therefore, only a single replicate of the 3^4 design is frequently considered, and higher order interactions are combined to provide an estimate of error. As an illustration, if three-factor and higher interactions are negligible, then a single replicate of the 3^3 design provides 8 degrees of freedom for error, and a single replicate of the 3^4 design provides 48 degrees of freedom for error. These are still large designs for $k \geq 3$ factors and, consequently, not too useful.