Balanced Complete Factorial Designs

Consider a **factorial design** (FD) with two factors A and B, with levels $1, \ldots, a$ and $1, \ldots, b$ respectively, yielding a total of $k = ab$ factor combinations (treatments), and suppose that there are r **replications** in each treatment, giving $n = rab$ observations in total. Let x_{ijt} be the tth replicated observation in the (i, j)th factor-level combination.

- sample mean for Factor A level i

$$\overline{x}_{i.} = \frac{1}{br} \sum_{j=1}^{b} \sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, a$$

- sample mean for Factor B level j

$$\overline{x}_{.j} = \frac{1}{ar} \sum_{i=1}^{a} \sum_{t=1}^{r} x_{ijt} \qquad j = 1, \ldots, b$$

- sample mean for replicates in (i, j)th factor combination

$$\overline{x}_{ij} = \frac{1}{r} \sum_{t=1}^{r} x_{ijt} \qquad i = 1, \ldots, a, \ j = 1, \ldots, b$$

- overall sample mean

$$\overline{x}_{..} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r} x_{ijt}$$

- Sum of Squares for Treatments due to factor A ($\mathrm{SST_A}$)

$$SST_A = \sum_{i=1}^{a} br(\overline{x}_{i.} - \overline{x}_{..})^2$$

- Sum of Squares for Treatments due to factor B ($\mathrm{SST_B}$)

$$SST_B = \sum_{j=1}^{b} ar(\overline{x}_{.j} - \overline{x}_{..})^2$$

- Sum of Squares for Interaction ($\mathrm{SSI_{AB}}$)

$$SSI_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} r(\overline{x}_{ij} - \overline{x}_{i.} - \overline{x}_{.j} + \overline{x}_{..})^2$$

- Overall Sum of Squares (SS)

$$SS = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r} (x_{ijt} - \overline{x}_{..})^2$$

The following decomposition holds

$$SS = SST_A + SST_B + SSI_{AB} + SSE \qquad \therefore \qquad SSE = SS - SST_A - SST_B - SSI_{AB}$$

Define

$$MST_A = \frac{SST_A}{a-1} \qquad MST_B = \frac{SST_B}{b-1} \qquad MSI_{AB} = \frac{SSI_{AB}}{(a-1)(b-1)}$$

and

$$MSE = \frac{SSE}{n - ab}$$

HYPOTHESIS TESTING

- For testing for a **FACTOR A** effect, use

$$F = \frac{MST_A}{MSE}$$

Under the assumption of **NO FACTOR A EFFECT**, then

$$F \sim \text{Fisher-F}(a-1, n-ab)$$

which defines the rejection region and p-value in the usual way.

- For testing for a **FACTOR B** effect, use

$$F = \frac{MST_B}{MSE}$$

Under the assumption of **NO FACTOR B EFFECT**, then

$$F \sim \text{Fisher-F}(b-1, n-ab)$$

- For testing for an **INTERACTION**, use

$$F = \frac{MSI_{AB}}{MSE}$$

Under the assumption of **NO INTERACTION**, then

$$F \sim \text{Fisher-F}((a-1)(b-1), n-ab)$$

Note: The only difference between a randomized block design and a factorial design is that in the block design, one of the factors is known or strongly believed to have a significant effect on the response. The method of analysis for interaction and no interaction models are identical.

BALANCED COMPLETE FACTORIAL DESIGNS: EXAMPLES

EXAMPLE 1: Butterfat data (Sokal, R. R. and Rohlf F. J. (1981). *Biometry*, 2nd edition)
The data give the average butterfat content (percentages) for random samples of twenty cows (ten two-year old and ten mature (greater than four years old)) from each of five breeds. The data are from Canadian records of pure-bred dairy cattle. There are 100 observations on two age groups (two years and mature) and five breeds.

The **response variable** is butterfat level. **Factor A** is the *age* and there are $a = 2$ **factor levels**:

1. Mature
2. Two years

Factor B is the *breed* and there are $b = 5$ **factor levels**:

1. Ayrshire
2. Canadian
3. Guernsey
4. Holstein-Fresian
5. Jersey

$r = 2$ replicate measurements were made, so that $n = 2 \times 5 \times 2 = 20$ data were obtained in total. The data are available from the course website as **Butterfat.sav**

Results:

1. **Interaction model**: First note that the Levene test **REJECTS** the null hypothesis of equal group variances ($p = 0.008$), so the following ANOVA results are questionable. However, the p-value is not too small, so we proceed but with caution.

   There is a **significant difference** due to Factor B (breed, $F = 49.565$, p-value < 0.001), but there is no effect of Factor A (age, F=1.580, $p = 0.212$), and no significant interaction (age*breed, $F = 0.742$, $p = 0.566$.

2. **Factor B only**: If we omit the Factor A and interaction term, and refit the model, we confirm the strong effect of Factor B ($F = 49.802, p < 0.000$), and then can estimate the Factor B treatment means. Note how the error degrees of freedom changes when terms in the model are omitted.

EXAMPLE 2: Lyrics data (McClave and Sincich, Statistics, 10th Edition (p577, Ex 10.88))
The effect of violent song lyrics on the aggression level of listeners is to be investigated. Two songs (classified as *Violent* and *Non-Violent*) were played to two groups (or "pools") of students, one volunteer group and one group drawn from a psychology class. The students then rated the songs lyrical content, and from this (by means of a word-association test), the aggression level of the students was computed.

The **response variable** is aggression level. **Factor A** is the *song* and there are $a = 2$ **factor levels**:

1. Violent
2. Non-violent

Factor B is the *pool* and there are $b = 2$ **factor levels**:

1. Volunteer
2. Psychology class

$r = 15$ replicate measurements were made, so that $n = 2 \times 2 \times 15 = 60$ data were obtained in total. The data are available from the course website as **Lyrics.sav**

Results:

1. **Interaction model**: First note that the Levene test **DOES NOT REJECT** the null hypothesis of equal group variances ($p = 0.804$)

   There is a **significant difference** due to Factor A (song, $F = 26.114$, p-value < 0.001), but there is no effect of Factor B (pool, F=0.579, $p = 0.450$), and no significant interaction (song*pool, $F = 1.563, p = 0.216$.

2. Fits of the main-effects model (Factor A and Factor B but no interaction), and the Factor A only model confirm the results.

EXAMPLE 3: Gravel data
A company produces gravel from a number of quarries and in each quarry there are morning and afternoon shifts of workers. The company wishes to know whether there are differences in the quantity of gravel produced from these quarries and gathers the following data on the amount of gravel produced by each shift in one week (in tonnes). It can be assumed that the week being studied was a typical week, and that there was no systematic differences due to different workers etc.

The **response variable** is amount of gravel produced. **Factor A** is the *shift* and there are $a = 2$ **factor levels**:

1. AM
2. PM

Factor B is the *quarry* and there are $b = 4$ **factor levels**:

1. A
2. B
3. C
4. D

$r = 5$ replicate measurements were made, so that $n = 2 \times 4 \times 5 = 40$ data were obtained in total.

The data are available from the course website as **Gravel.sav**

Results:

1. **Interaction model**: First note that the Levene test **DOES NOT REJECT** the null hypothesis of equal group variances ($p = 0.969$).

   There is a **significant difference** due to Factor A (shift, $F = 13.667, p = 0.001$), and due to Factor B (quarry, F=19.996, $p < 0.001$), but **no significant interaction** (shift*quarry, $F = 1.099, p = 0.364$.

2. **Factor A and B only**: If we omit the interaction term, and refit the model, we confirm the strong effect of both factors (*shift* $F = 13.552, p = 0.001$, *quarry* $F = 19.829, p < 0.001$. The conclusion is that there is a difference between the two levels of factor *shift* and the four levels of factor *quarry*, but that there is no interaction, that is, the difference between morning and afternoon shift is the same in each block; this is depicted in the Marginal Means plot.

   Note again how the error degrees of freedom changes when terms in the model are omitted.