# Contents


1 Principles of Experimental Design
1.1 What this Chapter is About . . . 1
1.2 Experimental Design . . . 1
1.2.1 Definitions . . . 1
1.2.2 Randomization . . . 1
1.2.3 Replication . . . 3
1.2.4 Blocking . . . 4
1.3 Exercises . . . 6
. . . 6

2 Completely Randomized Designs
2.1 What this Chapter is About . . . 9
2.2 Completely Randomized Designs (CRD's) . . . 9
2.3 Analysis of the CRD . . . 9
2.3.1 Models . . . 10
2.3.2 ANOVA . . . 11
2.3.3 Estimation . . . 12
2.4 Additional Topics . . . 15
2.4.1 Unbalanced CRD / Missing Observations . . . 16
2.4.2 Checking Model Assumptions . . . 16
2.5 Exercises . . . 17
. . . 19

3 Randomized Block Designs
3.1 What this Chapter is About . . . 20
3.2 Randomized Block Designs (RBD's) . . . 20
3.3 Analysis of the RBD . . . 20
3.3.1 Models . . . 22
3.3.2 ANOVA . . . 23
3.3.3 Estimation . . . 25
3.3.4 Checking Model Assumptions . . . 27
3.4 Exercises . . . 28
. . . 28

4 Latin Square and Crossover Designs
4.1 What this Chapter is About . . . 29
4.2 Latin Square Designs . . . 29
4.3 Analysis of a Latin Square Design . . . 29
4.3.1 Models . . . 31
4.3.2 ANOVA . . . 32
4.3.3 Estimation . . . 34
4.4 Replication of Latin Squares . . . 37
4.5 Crossover Designs . . . 37
4.6 Exercises . . . 40
. . . 44

5 Factorial Designs
5.1 What this Chapter is About . . . 46
5.2 Factorial Designs . . . 46
5.3 Analysis of Factorial Designs . . . 46
5.3.1 Completely Randomized Design (CRD) . . . 47
5.3.2 Randomized Block Design (RBD) . . . 47
5.4 Exercises . . . 57
. . . 63

6 Nested Designs 65
6.1 What this Chapter is About . . . 65
6.2 Nested Designs . . . 65
6.3 2-stage Nested Design . . . 66
6.4 Analysis of 2-stage Nested Design . . . 66
6.4.1 Models . . . 66
6.4.2 ANOVA . . . 68
6.4.3 Estimation . . . 71
6.5 Exercises . . . 72

7 Split-Plot Designs 73
7.1 What this Chapter is About . . . 73
7.2 Split-Plot Designs . . . 73
7.3 Analysis of Split-Plot Design . . . 74
7.3.1 Model . . . 74
7.3.2 ANOVA . . . 75
7.4 Exercise . . . 77

# Chapter 1
# Principles of Experimental Design

## 1.1 What this Chapter is About

One of the topics in BMET 210 was the analysis of variance (ANOVA) of experimental data to compare "population" (hereafter called treatment or factor level) means. The experimental data that were analyzed were obtained from experiments that were planned following certain rules and procedures. These rules and procedures constitute what are called the **principles of experimental design**. In this chapter we present and discuss the three basic principles of experimental design which are: **randomization**, **replication** and **blocking**.

By the end of this chapter you should be able to

* identify the treatments or factor levels, the experimental units and the response of interest in a given experiment;
* explain by means of examples the objectives of randomization, replication and blocking;
* randomize a simple experiment;
* determine the optimum number of replications for a simple experiment;
* identify the blocking factor, block and randomize a given experiment.

## 1.2 Experimental Design

The process of designing an experiment for comparing treatment or factor level means begins by stating the objective(s) of the experiment clearly. The statement of the objectives indicates to us what measurements are to be made (**how, when, where**) and on **what**. Consider the following two examples of the objectives of an experiment.

**Example 1.1** *To compare the mean weight gains of steers that are fed diets $A$, $B$ and $C$.* ◇

**Example 1.2** *To compare the mean weight gains of two-year-old Holstein steers that are fed diets $A$, $B$ and $C$ for a period of six months.*

The objective of the experiment in example 1.1 is vague. **Why?** On the other hand the objective of the experiment in example 1.2 is clear/specifc. The statement of the objective indicates how the experiment should be conducted. **Suggest how?** Note that the statement of the objectives in example 1.2, in particular, specifies

(a) the set of treatments (diets $A$, $B$ and $C$) whose effects are to be investigated;

(b) the set of experimental units (two year old Holstein steers) to be used;

(c) the values of the response variable(s) (initial weight, final weight or weight gain) of interest.

### 1.2.1 Definitions

We define, giving examples, some of the terms (treatment, experimental unit, response variable, etc) that are used in experimental design. Consider the following experiment.

**Example 1.3** *To investigate the effects of the* **Rate of Application** *of a certain type of fertilizer (at levels (rates):* 1=0 kg/ha, 2=100 kg/acre, 3=200 kg/acre*), and the* **Watering Frequency** *(at frequencies:* 1= once/week, 2= twice/week*) on the* **yield** *of two (2) maize varieties $X$ and $Y$ grown under the same management and climatic conditions.*◇

### Factor

A **factor** is a quantitative or qualitative variable which affects (determines, influences) the values of one or more **response** or **dependent** variables. A factor **level** is a particular value or set of value(s) of a factor.

**Example 1.4** *Refer to examples 1.2 and 1.3.*

| Factor | Levels | Response Variable |
|---|---|---|
| *Diet (qualitative)* | *A, B, C* | *Weigh Gain* |
| *Rate of Application (quantitative)* | *0, 1, 2* | *Yield* |
| *Watering Frequency (quantitative)* | *1,2* | *Yield* |
| *Variety* | *X, Y* | *Yield* |

◇

### Treatment

A **treatment** is the level of a factor or a combination of the levels of 2 or more factors whose effects on some specified variable are to be investigated.

**Example 1.5**

*Refer to example 1.2. The* **Diet** *treatments are $A, B$ and $C$.* ◇

*Refer to example 1.3. The 12* **Rate of Application x Watering Frequency x Variety** *factor level combinations are the treatments, e.g. Treatment 1=(0,1,X), Treatment 2=(0,1,Y), ..., Treatment 12=(2,2,Y).* ◇

### Experimental Unit

An **experimental unit** is the smallest **experimental material** upon which a treatment is applied or imposed. ◇

**Example 1.6**

*Refer to example 1.2. The* **experimental material** *are the 2-year-old Holstein steers. If the two year old Holstein steers are individually fed their assigned diets, then the steers are the* **experimental units**, *otherwise, if the steers are group fed their assigned diets, then the* **groups** *are the experimental units.* ◇

*Refer to example 1.3. The* **experimental material** *is the experimental are and the* **experimental units** *are the plots.* ◇

### Response Variable

The **response variable** is the characteristic of the experimental unit that is measured after applying the treatment on the unit. A response variable is also called a **dependent variable**.

**Example 1.7**

*Refer to example 1.2. The response variable is the weight gain of a steer (if steer is the experimental unit) or the weight gain of a group of steers under the same diet (if the steers are group fed) .* ◇

*Refer to example 1.3. The response variable is the yield per unit area of the plot.* ◇

**Exercise 1.1** *Suppose that we wish to conduct an experiment to compare the mean nicotine content of three brands of cigarettes - A, B and C. We will analyst ten cigarettes of each brand for nicotine content and record the amount of nicotine in each cigarette in milligrams.*

(a) *What is our response variable?*

(b) *Identify the factor we wish to study. What are the factor levels of the factor?*

(c) *Identify the experimental units.* ◇

### 1.2.2 Randomization

Refer to example 1.3. If the mean yields of the two varieties are actual the same, what would be the conclusion from the experiment in which fertile plots are assigned to variety $X$ and poor plots are assigned to variety $Y$? The answer to this question is simple. The experimental procedure favors variety $X$. An erroneous conclusion that variety $X$ has a higher yield than variety $Y$ would be made when in actual fact the varieties have the same mean yield. If the allocation of the plots to the varieties is done (as above) deliberately, then the **bias** in the conclusions is called **subjective** bias or bias due to **deliberate selection**, otherwise, the bias is called **systematic** bias.

**Example 1.8** *Suppose that a new diet $A$ is believed to be better than an existing diet $B$ in terms of increasing the daily weight gain of steers (of the same age) fed the diets. Furthermore, suppose that four two-year-old Afrikander steers and four two year old Jersey steers are available for experimentation to verify the claim. Known or unknown to the experimenter is that naturally Afrikander steers grow faster ihan Jersey steers of the same age raised under the same management conditions. If in actual fact diet $A$ is as good as diet $B$, what will be the conclusion from the experiment whereby Afrikander steers are assigned to diet $A$ and Jersey steers are assigned to diet $B$?* $\diamond$

Again the experimental procedure described in example 1.8 favors diet $A$ since naturally Afrikander steers grow faster than Jersey steers of the same age raised under the same environmental conditions. An erroneous conclusion that diet $A$ is better than diet $B$ would be made when in actual fact the two diets are equally good/bad.

Subjective and systematic biases can be eliminated by **randomizing** the experiment. Furthermore, randomizing the experiment makes the **errors in the response measurements statistical independent** - an assumption that is required by many statistical methods of analyzing data including ANOVA. But, what is **randomization**?

**Randomization** of an experiment is the random allocation of the experimental units to the treatments. That is, it is the allocation of the experimental units to the treatments in a haphazard way. For example, we can randomly allocate $n = tr$ identical experimental units to $t$ treatments as follows:

(a) Number the $n$ experimental units with numbers from 1 to $n$.

(b) Number $n$ slips of paper with numbers from 1 to $n$.

(c) After mixing the slips of paper thoroughly, the first $r$ numbers drawn blindly from the hat or other container are units assigned to the first treatment, the second $r$ numbers are units assigned to the second treatment, etc.

For example, we can randomize the experiment in example 1.8 as follows:

(a) Number the available steers with numbers from 1 to 8.

(b) Number 8 slips of paper with numbers from 1 to 8.

(c) After mixing the slips of paper thoroughly, draw 8 numbers blindly from the hat or other container to obtain a random sequence of numbers which might be 5, 4, 7, 1, 6, 8, 3, 2. (The random sequence of the numbers 1 to 8 is called a **random permutation** of the numbers 1 to 8.) Assign steers 5, 4, 7, 1 to diet $A$ and the remainder (6, 8, 3, 2) to diet $B$ to obtain the experimental layout displayed in the table that follows.

| Steer | 5 | 4 | 7 | 1 | 6 | 8 | 3 | 2 |
|---|---|---|---|---|---|---|---|---|
| Diet | $A$ | $A$ | $A$ | $A$ | $B$ | $B$ | $B$ | $B$ |

The above guarantees that each steer is equally likely to be assigned to any of the two diets and hence avoids subjective or systematic bias in the conclusions of our experiment. Furthermore, randomization guarantees the independence of the measurements from the experiment as was mentioned before.

**Exercise 1.2** *Suppose that you wish to compare the effects of three types of fertilizer ($X$, $Y$, $Z$) on the yield of a certain variety of maize grown under the natural conditions of SA. Three one-acre plots are available for experimentation in each of the three research stations: Limpopo, Mpumalanga, KZN. In terms of the suitability of the environmental conditions for maize growth, KZN has the best followed by Mpumalanga.*

(a) *How would assigning plots in Limpopo to $X$, plots in Mpumalanga to $Y$ and plots in KZN to $Z$ affect your conclusions if in actual fact the effects of the fertilizers on yield are the same?*

(b) *Suppose that the 9 plots are identified by labels* **L1, L2, L3, M4, M5, M6, K7, K8, K9**, *where the alpha character is the first letter of the station name. Randomize your experiment (complete the table below) using the following random permutation of numbers from 1 to 9: 3, 7, 2, 6, 9, 8, 4, 1, 5.* ◇

| *Plot* | ... | ... | ... | ... | ... | ... | ... | ... | ... |
|---|---|---|---|---|---|---|---|---|---|
| *Fertilizer* | $X$ | $X$ | $X$ | $Y$ | $Y$ | $Y$ | $Z$ | $Z$ | $Z$ |

### 1.2.3 Replication

We define a **basic experiment** to be one in which only one experimental unit is assigned to each treatment. Thus, each treatment appears once in basic experiment.

**Replication** is the independent repetition of the **basic experiment**. In other words, replication is the assignment of at least two experimental units to each of the treatments whose effects are under investigation. For example, if we have four treatments, then the basic experiment is four experimental units randomly assigned to the four treatments. Two or more replications of this basic experiment constitute our whole experiment for comparing treatment means. Why replicate the basic experiment? Replication allows the accurate estimation of the **experimental error**, improves the reliability of the estimates of the treatment means, and also improves the sensitivity of statistical tests for comparing treatment means. We illustrate these purposes of replication by means of examples.

#### Examples

Suppose that we wish to compare the effects of two treatments ($T_1$, $T_2$) on some response. Furthermore, suppose that we have $2r$ ($r$ an integer greater than or equal to 1) identical experimental units available for experimentation. The plan of conduct of the experiment is to randomly allocate $n$ experimental units to $T_1$ and the remainder to $T_2$. The experiment for $r = 1$ is our basic experiment, for $r = 2$ we have two replications of our basic experiment, etc. The output from the whole experiment is displayed in table 1.1. In the table, $Y_{ij}$ is the $j^{th}$ response ($j = 1, 2, .., r$) to the $i^{th}$ treatment ($i = 1, 2$).

*Table 1.1 The data and the statistics for comparing two treatment means*

| | Replication | | | | | Mean | Variance |
|---|---|---|---|---|---|---|---|
| Treatment | 1 | 2 | . | . | r | | |
| $T_1$ | $Y_{11}$ | $Y_{12}$ | . | . | $Y_{1r}$ | $\bar{Y}_{1.} = \frac{1}{r}\sum_{j=1}^{r} Y_{1j}$ | $S_1^2 = \frac{1}{r-1}\sum_{j=1}^{r}(Y_{1j} - \bar{Y}_{1.})^2$ |
| $T_2$ | $Y_{21}$ | $Y_{22}$ | . | . | $Y_{2r}$ | $\bar{Y}_{2.} = \frac{1}{r}\sum_{j=1}^{r} Y_{2j}$ | $S_2^2 = \frac{1}{r-1}\sum_{j=1}^{r}(Y_{2j} - \bar{Y}_{2.})^2$ |

Recall that the **pooled t-test** assumptions are that the errors in the $Y_{ij}'s$ are independent and normally distributed with mean 0 and variance $\sigma^2$ (unknown). The $\sigma^2$ is a measure of the experimental error and it is estimated by the **pooled sample variance** which, in this case, is given by:

$$S_p^2 = \frac{1}{2}(S_1^2 + S_2^2).$$

What happens to $S_p^2$ if we do not replicate the basic experiment, i.e., if $r = 1$? Clearly, $S_1^2$, $S_2^2$ and hence $S_p^2$ are undefined, i.e., $\sigma^2$ can not be estimated. However, when $r > 1$, then the estimate of $\sigma^2$ ($S_p^2$) is well defined. This confirms that replication of the basic experiment allows the estimation the experimental error $\sigma^2$. If $r$ is large then $S_p^2$ will be very close to $\sigma^2$.

The estimate of the difference between the $T_1$ mean and the $T_2$ mean is:

$$\bar{Y}_{1.} - \bar{Y}_{2.} \text{ with variance } \sigma_{12}^2 = \frac{2}{r}\sigma^2.$$

The variance $\sigma_{12}^2$ is estimated by $\hat{\sigma}_{12}^2 = \frac{2}{r}S_p^2$. The $\sigma_{12}^2$ is a measure of the precision or the reliability of $\bar{Y}_{1.} - \bar{Y}_{2.}$ in estimating the difference between the treatment means. The estimate is precise or reliable if $\sigma_{12}^2$ is small. What happens to $\sigma_{12}^2$ if we increase $r$? The precision or the reliability of the estimate improves as we increase the number of replications of the basic experiment since $\sigma_{12}^2 \to 0$ as $r \to \infty$. You will learn from exercise 1.3 that the width of the 95% confidence interval for the difference between treatment means decreases as $r$ increases. That is, the confidence interval becomes more and more informative (accurate and precise) as $r$ is increased.

**Exercise 1.3** *Refer to table 1.1. Assume that the errors in the $Y_{ij}'s$ are normally with mean 0 and variance $\sigma^2$ (known).*

(a) *Derive a formula for the width ($w$) of a 95% confidence interval for the difference between the two treatment means.*

(b) *How does $w$ vary with $r$?* ◇

Suppose that we wish to test the hypotheses:

$$H_o: \ T_1 \text{ mean} = T_2 \text{ mean versus } H_a: \ T_1 \text{ mean} \neq T_2 \text{ mean}.$$

If the variance of the errors $\sigma^2$ is known, then the appropriate test is the **z-test**. The test statistic for the $z - test$ is:

$$z = \frac{\bar{Y}_{1.} - \bar{Y}_{2.}}{\sigma_{12}},$$

which has a **standard normal** distribution. The $z - test$ rejects $H_o$ in favor of $H_a$ if $|z|$ is large. How does $|z|$ vary with $r$? If the treatment means are not different, then we expect $|\bar{Y}_{1.} - \bar{Y}_{2.}| \to 0$ as $r \to \infty$, i.e., we expect $|z| \to 0$ as $r \to \infty$. However, if the treatment means are different, then $|\bar{Y}_{1.} - \bar{Y}_{2.}|$ is expected to approach the true absolute difference between the treatment means while $\sigma_{12} \to 0$ as $r \to \infty$. Hence, we expect $|z| \to \infty$ as $r \to \infty$. This means the $z$ test can reject $H_o$ in favor of $H_a$ even for small differences between the treatment means if $r$ is large enough. Thus, the sensitivity of our test to small differences between treatment means increases as the number of replications increases. Similar arguments can be used to show that replication increases the sensitivity of the **pooled** $t - tests$ and the $F - tests$ for comparing treatment means.

A measure of the sensitivity of a test for comparing treatment means is the **power** of the test to detect differences between treatment means. The power of a test is the probability of rejecting $H_o$ (treatment means are not different) when $H_a$ is true (treatment means are different). In other words, the power of a test is the probability that it will detect differences between treatment means. For example, suppose that we wish to use the $z - test$ to test the hypotheses:

$$H_o: \ T_1 \text{ mean} = T_2 \text{ mean versus } H_a: \ T_1 \text{ mean} \neq T_2 \text{ mean}$$

at the $\alpha$ **level of significance**. Let $\delta$ be the true difference between the treatment means. Then it can be shown that the power of the $z - test$ is given by:

$$\text{Power} = P[z > z_{\alpha/2} - \sqrt{r}\frac{|\delta|}{\sigma}],$$

where $z_{\alpha/2}$ is the $(1 - \alpha/2)100^{th}$ percentile of a standard normal distribution.

Yield = Treatment + Error
Yield = Treatment + Block + Error

**Exercise 1.4** *Consider the above $z-$test. How does the power of the test vary with $r$?* ◇

Similarly, it can be shown that the power of the **pooled** $t-$*tests* and the $F-$*tests* for comparing treatment means increases with $r$.

### 1.2.4 Blocking

**Blocking** an experiment refers to the arrangement of the experimental units into groups (called **blocks**) within each of which the experimental units are relatively homogeneous (identical) with respect to one or more characteristics of the units that may influence the response of interest. Randomization is then done independently within each block. Blocking can also be based on external variables (that may influence the response) associated with the experimental setting. For example, **time** if the experiment is to be repeated over time, **observer** if two or more people perform the experiment, etc.

Refer to example 1.8. In this example **breed** is a characteristic of the steers that affects the growth rate of the steers. Therefore, it should be used as a blocking factor. This is done as follows:

* Stratify the steers into groups by breed.

* Randomly assign the steers within each breed to the diets using different sets of random numbers.

**Exercise 1.5** *Refer to example 1.8. Suppose that the Afrikandér steers are $A_1, A_2, A_3, A_4$, and the Jersey steers are $J_1, J_2, J_3, J_4$. Furthermore, suppose that we have the following sets of random numbers: Set 1 = $\{4, 2, 1, 3\}$ and Set 2 = $\{2, 3, 4, 1\}$. Use these sets of random numbers to randomly allocate the steers to the diets. Show the layout of the experiment in the table below.* ◇

| *Diet* | *Afrikander* | *Jersey* |
|---|---|---|
| *A* | ..... | ..... |
| *B* | ..... | ..... |

**Exercise 1.6** *Refer to exercise 1.2. Suggest the blocking factor(s) for the experiment.* ◇

**Why block an experiment?** Blocking an experiment allows us to account for the variation in the responses that is due to differences among the experimental units. If we block an experiment using external variables such as time or observer, then blocking allows us to account for the variation in the responses that is due to these external variables. The consequence of not blocking when one is supposed to block is that the variation due to differences among the experimental units or due to the external variables can not be separated from that due to the random errors. This results in an estimate of the experimental error ($\sigma^2$) that is biased upwards. On the other hand, effective blocking reduces the experimental error, and hence obtains precise estimates of the treatment means and makes the $t-$*tests* and the $F-$*tests* sensitive to treatment differences. You will learn from exercise 4 in section 1.3 (below) how effective blocking reduces the experimental error.

There are many methods of blocking experiments in order to improve the precision of the experimental results. In all cases, blocking is a restriction of the random assignment of the experimental units to the treatments. We will learn more about the most important methods of blocking experiments in the this module.

## 1.3 Exercises

An experiment was conducted to compare the effect three different insecticides (A, B, C) on the number of seedlings, of a particular variety of string bean, that emerged per subplot. Four different plots (in terms of moisture content, fertility, etc.) of the same size were prepared, with each plot divided into three subplots of the same size. A suitable distance was maintained between the subplots within a plot. Each subplot was planted with 100 seeds and then maintained under the insecticide randomly assigned to it.

1. Identify the experimental units, the treatments and the response of variable of interest in the experiment.

2. What was the blocking factor in the experiment and why?

3. Let $P_{ij}$ ($i = 1, 2, 3, 4$, $j = 1, 2, 3$) be the $j^{th}$ subplot in the $i^{th}$ plot, e.g., $P_{12}$ is subplot 2 in plot 1. Show the layout of the design of the experiment. Use the following sets of random numbers to randomize your experiment: $\{2, 1, 3\}$, $\{2, 3, 1\}$, $\{1, 2, 3\}$ and $\{3, 2, 1\}$. How many times was the basic experiment replicated?

4. The data collected from the experiment are given in the table below.

| | Plot | | | |
|---|---|---|---|---|
| Insecticide | 1 | 2 | 3 | 4 |
| $A$ | 56 | 49 | 65 | 60 |
| $B$ | 84 | 78 | 94 | 93 |
| $C$ | 80 | 72 | 83 | 85 |

(a) One-way analysis of variance ignoring **Plots**. What is the estimate of the experimental error ($\sigma^2$)?

```
***** Analysis of variance *****
Variate: number

Source of variation      d.f.      s.s.       m.s.      v.r.   F pr.
insect                     2    1925.17     962.58     21.14   <.001
Residual                   9     409.75      45.53
Total                     11    2334.92
```

Yield = Treatment + Error — residual

one-way anova

2. What was the blocking factor in the experiment and why?

3. Let $P_{ij}$ $(i = 1, 2, 3, 4,\ j = 1, 2, 3)$ be the $j^{th}$ subplot in the $i^{th}$ plot, e.g., $P_{12}$ is subplot 2 in plot 1. Show the layout of the design of the experiment. Use the following sets of random numbers to randomize your experiment: $\{2, 1, 3\}$, $\{2, 3, 1\}$, $\{1, 2, 3\}$ and $\{3, 2, 1\}$. How many times was the basic experiment replicated?

4. The data collected from the experiment are given in the table below.

| | Plot | | | |
|---|---|---|---|---|
| Insecticide | 1 | 2 | 3 | 4 |
| $A$ | 56 | 49 | 65 | 60 |
| $B$ | 84 | 78 | 94 | 93 |
| $C$ | 80 | 72 | 83 | 85 |

(a) One-way analysis of variance ignoring **Plots**. What is the estimate of the experimental error $(\sigma^2)$?

```
***** Analysis of variance *****
Variate: number

Source of variation      d.f.      s.s.       m.s.      v.r.   F pr.
insect                      2   1925.17     962.58     21.14   <.001
Residual                    9    409.75      45.53
Total                      11   2334.92
```

field = Treatment + Error

one-way anova

(b) Two-way ANOVA with **Plot** as a second factor. What is the estimate of the experimental error ($\sigma^2$)?

```
***** Analysis of variance *****
Variate: number

Source of variation      d.f.       s.s.       m.s.     v.r.   F pr.

plot stratum               3    386.250    128.750    32.87

plot.*Units* stratum
insect                     2   1925.167    962.583   245.77   <.001
Residual                   6     23.500      3.917

Total                     11   2334.917
```

(c) Which of the two estimates of the experimental error in ($a$) and ($b$) is larger and why?

(d) Which of the two $F-tests$, in ($a$) and ($b$), for comparing the insecticides is more powerful and why?

$Y = \alpha + bx$

Yield = Treatment + plot + error.

3. The statistical analysis of the experiment (ANOVA and estimation) is easy even when

   (a) there are observations that are missing by accident or by design;

   (b) a treatment is dropped.

The major disadvantage of a CRD is that it can only be used with homogeneous (identical or uniform) experimental units. These may be difficult to obtain if number of treatments is large. heterogeneous experimental units in a CRD leads to inflated estimates of the experimental error and hence making it more difficult to detect differences among the treatment means (see exercise 4 in section 1.3 of chapter 1).

**Exercise 2.1** *Suppose that you want to compare the effects of three types of fertilizer (X, Y, Z) on the yield of a certain variety of maize. Furthermore, suppose that 9 plots of the same size are available for the experiment.*

(a) *List at least two characteristics of the plots that can affect the response of interest.*

(b) *List at least two environmental factors that can affect the response of interest.*

(c) *List at least two management practices that can affect the response of interest besides the method and the level of application of the fertilizers.*

(d) *In view of your answers to (a) – (c), under what conditions would a CRD design be appropriate for the experiment?* ◇

## 2.3 Analysis of the CRD

In this section we review the analysis of a CRD. The analysis of the design was previously discussed BMET 210 (ANOVA). Table 2.1 displays the layout of the data in a CRD with $t$ treatments and $r$ experimental units per treatment. The symbols in the table have the following meanings:

* $Y_{ij}$ is the $j^{th}$ response to the $i^{th}$ treatment;

* $Y_{i.} = \sum_{j=1}^{r} Y_{ij}$ and $\bar{Y}_{i.} = \frac{1}{r} Y_{i.}$ are the $i^{th}$ treatment total and treatment mean, respectively;

* $i = 1, 2, ..., t;\ j = 1, .., r;\ n = tr = \#(\text{observations})$.

*Table 2.1 Layout of the data in CRD*

| | Replication | | | | | |
|---|---|---|---|---|---|---|
| Treatment | 1 | 2 | . . | r | Total | Mean |
| $T_1$ | $Y_{11}$ | $Y_{12}$ | . . | $Y_{1r}$ | $Y_{1.}$ | $\bar{Y}_{1.}$ |
| $T_2$ | $Y_{21}$ | $Y_{22}$ | . . | $Y_{2r}$ | $Y_{2.}$ | $\bar{Y}_{2.}$ |
| . | . | . | . . | . | . | . |
| $T_t$ | $Y_{t1}$ | $Y_{t2}$ | . . | $Y_{tr}$ | $Y_{t.}$ | $\bar{Y}_{t.}$ |

We shall use the data in the example that follows to illustrate the analysis of a CRD.

**Example 2.1** *A CRD was used to compare the mileage (km) per litre obtained by competing brands of petrol. Fifteen identical cars were randomly assigned to brands I, II and III, with each brand assigned five cars. The cars were operated under the same conditions and the distance travelled by each car per litre of the assigned brand of petrol recorded. The results are displayed in table 2.2. Analyst the data and draw appropriate conclusions.* ◇

# Chapter 2
# Completely Randomized Designs

## 2.1 What this Chapter is About

In this chapter we discuss the design and the analysis of an experiment for comparing more than two treatment (or factor level) means under the following conditions:

* the experimental units are identical with respect to their characteristics that can affect the response of interest;

* there are no external variables that can affect the response of interest;

* the inferences about the treatments (or factor levels) are to pertain to the set of treatments included in the experiment as well as when the inferences are to be extended to the population from which the set of treatments was randomly selected.

Under these conditions the simplest design for the experiment is a **completely randomized design** (CRD). Furthermore, the conditions prescribe the following two types of one-way analysis of variance (ANOVA) models for the CRD:

**Fixed effects model (model I):** a model that is used when the inferences about the treatments are to pertain to the set of treatments included in the experiment.

**Random effects model (model II):** a model that is used when inferences about the treatments are to be extended to the population from which the set of treatments were randomly selected.

By the end of this chapter you should be able to

* describe a CRD and state the conditions under which it is used;

* state the advantages and the disadvantages of using a CRD;

* identify the appropriate model (I or II) for a given CRD;

* analyst a CRD and draw appropriate conclusions.

## 2.2 Completely Randomized Designs (CRD's)

In a CRD **uniform/identical** experimental units are assigned to the treatments completely at random. For example, if there are $t$ treatments and $n = tr$ uniform/identical experimental units, then the $n$ experimental units are assigned to the $t$ treatments completely at random and in such a way that each treatment is assigned $r$ experimental units. Complete randomization ensures that every unit has an equal chance to be assigned to any one of the $t$ treatments. The value of complete randomization is insurance against subjective and systematic biases as was mentioned in section 1.2.2 of chapter 1.

We only use a CRD when both the experimental units and experimental conditions are **uniform**. That is, when the experimental units are identical with respect to their characteristics that can affect the response and there are no external variables associated with the experimental setting that can also affect the response (see section 1.2.4 of chapter 1). The advantages of using a CRD are as follows:

1. The design can be used with any number of treatments ($t$) if the resources are permitting.

2. The number of experimental units ($r$) can be varied from treatment to treatment without complicating the analysis of the experiment. This might be desirable when comparisons between certain treatments are to be more precise than others.

3. The statistical analysis of the experiment (ANOVA and estimation) is easy even when

   (a) there are observations that are missing by accident or by design;

   (b) a treatment is dropped.

The major disadvantage of a CRD is that it can only be used with homogeneous (identical or uniform) experimental units. These may be difficult to obtain if number of treatments is large. heterogeneous experimental units in a CRD leads to inflated estimates of the experimental error and hence making it more difficult to detect differences among the treatment means (see exercise 4 in section 1.3 of chapter 1).

**Exercise 2.1** *Suppose that you want to compare the effects of three types of fertilizer (X, Y, Z) on the yield of a certain variety of maize. Furthermore, suppose that 9 plots of the same size are available for the experiment.*

(a) *List at least two characteristics of the plots that can affect the response of interest.*

(b) *List at least two environmental factors that can affect the response of interest.*

(c) *List at least two management practices that can affect the response of interest besides the method and the level of application of the fertilizers.*

(d) *In view of your answers to (a) – (c), under what conditions would a CRD design be appropriate for the experiment?* ◇

## 2.3 Analysis of the CRD

In this section we review the analysis of a CRD. The analysis of the design was previously discussed BMET 210 (ANOVA). Table 2.1 displays the layout of the data in a CRD with $t$ treatments and $r$ experimental units per treatment. The symbols in the table have the following meanings:

* $Y_{ij}$ is the $j^{th}$ response to the $i^{th}$ treatment;

* $Y_{i.} = \sum_{j=1}^{r} Y_{ij}$ and $\bar{Y}_{i.} = \frac{1}{r} Y_{i.}$ are the $i^{th}$ treatment total and treatment mean, respectively;

* $i = 1, 2, ..., t;\ j = 1, .., r;\ n = tr = \#(\text{observations})$.

*Table 2.1 Layout of the data in CRD*

| | Replication | | | | | | |
|---|---|---|---|---|---|---|---|
| Treatment | 1 | 2 | . | . | r | Total | Mean |
| $T_1$ | $Y_{11}$ | $Y_{12}$ | . | . | $Y_{1r}$ | $Y_{1.}$ | $\bar{Y}_{1.}$ |
| $T_2$ | $Y_{21}$ | $Y_{22}$ | . | . | $Y_{2r}$ | $Y_{2.}$ | $\bar{Y}_{2.}$ |
| . | . | . | . | . | . | . | . |
| $T_t$ | $Y_{t1}$ | $Y_{t2}$ | . | . | $Y_{tr}$ | $Y_{t.}$ | $\bar{Y}_{t.}$ |

We shall use the data in the example that follows to illustrate the analysis of a CRD.

**Example 2.1** *A CRD was used to compare the mileage (km) per litre obtained by competing brands of petrol. Fifteen identical cars were randomly assigned to brands I, II and III, with each brand assigned five cars. The cars were operated under the same conditions and the distance travelled by each car per litre of the assigned brand of petrol recorded. The results are displayed in table 2.2. Analyst the data and draw appropriate conclusions.* ◇

*Table 2.2 Mileage (km) per litre for 3 brands of petrol - I, II & III*

| | Replication | | | | | | |
|---|---|---|---|---|---|---|---|
| Brand | 1 | 2 | 3 | 4 | 5 | Total | Mean |
| *I* | 10.5 | 12.0 | 14.0 | 16.0 | 19.0 | 71.5 | 14.3 |
| *II* | 11.5 | 13.0 | 15.0 | 17.0 | 11.5 | 68.0 | 13.6 |
| *III* | 11.0 | 11.5 | 14.5 | 15.5 | 11.0 | 63.5 | 12.7 |

### 2.3.1 Models

The general linear statistical model for a CRD has the form:

$$\begin{aligned} Y_{ij} &= \mu + \tau_i + \epsilon_{ij} \\ &= \mu_i + \epsilon_{ij} \end{aligned} \tag{2.1}$$

where:

* $\mu$ is the overall population mean;
* $\tau_i$ is the effect of the $i^{th}$ treatment;
* $\mu_i = \mu + \tau_i$ is the mean of the $i^{th}$ treatment;
* the $\epsilon'_{ij}s$ are random errors which are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$;
* $i = 1, .., t;\ j = 1, .., r$.

We see that the model (2.1) decomposes each observation into three components. For example, in table 2.2, the mileage per litre of the $2^{nd}$ car assigned brand *II* petrol ($Y_{22} = 13$) decomposes into a constant ($\mu$), the brand *II* effect on mileage ($\tau_2$) and the random error associated with the $2^{nd}$ car ($\epsilon_{22}$). That is,

$$12.0 = \mu + \tau_2 + \epsilon_{22}.$$

We distinguish between the **fixed effects model** (model I) and the **random effects model** (model II).

#### Model I

Model I is model (2.1) but with the $\tau'_i s$ regarded as fixed **real constants** satisfying the constraint:

$$\sum_{i=1}^{t} \tau_i = 0.$$

Under this model the estimates of the $i^{th}$ treatment or factor level mean ($\mu_i$) and the overall mean ($\mu$) are:

$$\hat{\mu}_i = \bar{Y}_{i.} = \frac{1}{r} Y_{i.} = \frac{1}{r} \sum_{j=1}^{r} Y_{ij} \text{ and } \hat{\mu} = \bar{Y}_{..} = \frac{1}{n} Y_{..} = \frac{1}{n} \sum_{i=1}^{t} \sum_{j=1}^{r} Y_{ij},$$

respectively. Since $\tau_i = \mu_i - \mu$, it follows that the obvious estimate of the $i^{th}$ treatment effect ($\tau_i$) is:

$$\hat{\tau} = \hat{\mu}_i - \hat{\mu} = \bar{Y}_{i.} - \bar{Y}_{..}.$$

Our objective is to determine whether or not the $t$ treatment means are equal. That is, we wish to test the hypotheses:

$$H_o : \mu_1 = \mu_2 = ... = \mu_t \text{ versus } H_a : \text{ not all } \mu'_i s \text{ are equal},$$

or, equivalently,

$$H_o : \tau_1 = \tau_2 = ... = \tau_t = 0 \text{ versus } H_a : \text{ not all the } \tau_i's \text{ are equal to zero}$$

at the $\alpha$ level of significance. Furthermore, if treatment means are not all equal, we wish to determine which treatment means are different.

**Remark 2.1** *We use model I to analyst a CRD when the conclusions of the experiment are to pertain to the particular set of the t treatments included in the experiment. The conclusions can not be extended to any other treatments that were not included in the experiment.* $\diamond$

### Model II

Model II is model (2.1) but with the

* $\tau_i's$ **random variables** which are independent and normally distributed with mean 0 and variance $\sigma_\tau^2$;

* $\epsilon_{ij}'s$ and the $\tau_i's$ assumed to be **independent** random variables.

The assumptions of model II imply that the treatment means ($\mu_i = \mu + \tau_i$) are random variables which are independent and normally distributed with mean $\mu$ and variance $\sigma_\tau^2$. Our objective is to test the hypotheses:

$$H_o : \sigma_\tau^2 = 0 \text{ versus } H_a : \sigma_\tau^2 > 0,$$

or, equivalently,
$H_o$ : the treatments or factor levels have identical effects
versus
$H_a$ : the treatments or factor levels have different effects
at the $\alpha$ level of significance. Furthermore, if $\sigma_\tau^2 > 0$, then we wish to estimate it.

**Remark 2.2** *We use model II to analyst a CRD when the set of the t treatments included in the experiment are a random sample from a large population of treatments. In this case the conclusions of the experiment can be extended to all the treatments in the population.* $\diamond$

**Example 2.2** *Refer to exercise 2.1. If your conclusions are to pertain to the three types of fertilizers (X, Y, Z), then model I is the appropriate one. However, if the three types of fertilizer are a random sample from a large population (3 out of 25, say) and your interest is in comparing the fertilizers in this large population, then model II is the appropriate one.* $\diamond$

Next we discuss the ANOVA procedure for testing the hypotheses about the treatment effects under both models I and II.

## 2.3.2 ANOVA

### Sum of Squares

The observations ($Y_{ij}'s$) in a CRD vary in magnitude due to the treatment effects (if any) and the random errors. We measure this variability (called the **total variability**) in the observations using the **total sum of squares** which is given by:

$$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}Y_{ij}^2 - \frac{1}{tr}Y_{..}^2.$$

**Remark 2.3** *Note that if the treatment effects ($\tau_i's$) are equal and if the random errors ($\epsilon_{ij}'s$) are all equal to zero, then the observations ($Y_{ij}'s$) are identical. In this case $SSTO = 0$, otherwise, $SSTO > 0$.* $\diamond$

We decompose $SSTO$ into the following components:

* the **treatment sum of squares** ($SST$) which measure the variability in the observations that is due to differences in the treatment effects;

* the **error sum of squares** ($SSE$) which the measure the variability in the observations that is due to the random errors.

The decomposition of $SSTO$ is done as follows:

$$\begin{aligned} SSTO = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{..})^2 &= \sum_{i=1}^{t}\sum_{j=1}^{r}\left[(\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})\right]^2 \\ &= \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{Y}_{i.} - \bar{Y}_{..})^2}_{=SST} + \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})^2}_{=SSE} \end{aligned} \tag{2.2}$$

The computing formula for $SST$ is:

$$SST = \sum_{i=1}^{t}\frac{1}{r}Y_{i.}^2 - \frac{1}{tr}Y_{..}^2.$$

Equation (2.2) suggests that we get $SSE$ by subtraction after computing $SSTO$ and $SST$, i.e.,

$$SSE = SSTO - SST.$$

**Exercise 2.2** *Refer to example 2.1. Verify that* $SSTO = 92.23$, $SSBrand = 6.43$ *and* $SSE = 85.80$. ◇

### Degrees of Freedom

Each sum of squares is associated with some **degrees of freedom** (df). The df for the sum of squares are independent pieces of information that are used in its calculation. For example, we note that $SSTO$ is calculated from $tr$ deviations $Y_{ij} - \bar{Y}_{..}$ which sum to zero, i.e.,

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{..}) = 0.$$

This means that the $tr$ deviations are not independent since one deviation is completely determined by the other $tr - 1$ deviations. It is for this reason that we say $SSTO$ has $tr - 1$ df associated with it. This is to say $SSTO$ is calculated from $tr - 1$ independent pieces of information. Similarly, $SST$ has $t - 1$ df associated with it since it is calculated from $t$ deviations $\bar{Y}_{i.} - \bar{Y}_{..}$ which sum to zero. The df associated with $SST$ and $SSE$ add up to the df associated with $SSTO$. Therefore, we get the df associated with $SSE$ by subtraction, i.e., the df of SSE are:

$$(tr - 1) - (t - 1) = t(r - 1).$$

**Exercise 2.3** *Refer to example 2.1. What are the degrees of freedom for* $SSTO$, $SST$ *and* $SSE$? ◇

### Mean Sums of Squares

The **mean sum of squares** is the sum of squares divided by the corresponding df. For example, the **treatment mean sum of squares** ($MST$) and the **error mean sum of squares** ($MSE$) are given by:

$$MST = \frac{1}{t-1}SST \text{ and } MSE = \frac{1}{t(r-1)}SSE,$$

respectively.

**Exercise 2.4** *Refer to example 2.1. Verify that* $MST = 3.22$ *and* $MSE = 7.15$. ◇

$SSTO = SST + SSE$

$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{..})^2$

$= \sum_{i=1}^{t}\sum_{j=1}^{r}[(\bar{y}_{i.}-\bar{y}_{..})+(y_{ij}-\bar{y}_{i.})]^2$

$= \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{y}_{i.}-\bar{y}_{..})^2}_{SST} + \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i.})^2}_{SSE}$

**Hypotheses Testing**

Under both models I and II, we can be show that $MSE$ is an **unbiased** estimate of the experimental error ($\sigma^2$), i.e.,

$$E[MSE] = \sigma^2.$$

nicholas.hill@glasgow.ac.uk

G17

## Model I

Under model I, we can show that $MST$ is an **unbiased** estimate of:

$$E[MST] = \sigma^2 + \frac{1}{t-1}\sum_{i=1}^{t} r\tau_i^2.$$

$\frac{\sum_{i=1}^{r} var(\bar{y}_i - \bar{y}_{..})}{r-1} = var(\bar{y}_i) = \frac{\sigma^2}{r}$

$\sigma^2 \frac{(r-1)}{r}$

Now let us consider testing the hypotheses:

$$H_o : \mu_1 = \mu_2 = ... = \mu_t \text{ versus } H_a : \text{ not all } \mu_i \text{ are equal,}$$

or, equivalently,

$$H_o : \tau_1 = \tau_2 = ... = \tau_t = 0 \text{ versus } H_a : \text{ not all } \tau_i = 0$$

at the $\alpha$ level of significance. We note that if $H_o$ is true, $MST$ and $MSE$ are both unbiased estimates of $\sigma^2$. However, if $H_a$ is true $MSE$ is still an unbiased estimate of $\sigma^2$, but, $MST$ is an unbiased estimate of

$$\sigma^2 + \frac{1}{t-1}\sum_{i=1}^{t} r\tau_i^2 > \sigma^2.$$

$\exists \mu \neq \mu \neq$

$\exists \mu \neq \mu \neq \mu$

This suggests using the ratio:

$$F = \frac{MST}{MSE} \tag{2.3}$$

as a test statistic for the hypotheses. We expect the value of $F$ to be equal to 1 if $H_o$ is true and to be large if $H_a$ is true. How large is large? If $H_o$ is true, then $F$ is a random variable with an $F$ distribution with $t-1$ **numerator df** and $t(r-1)$ **denominator df**. (The numerator df and the denominator df for the $F$ distribution are those for $MST$ and $MSE$, respectively.) Therefore, the decision rule for the testing problem is to reject $H_o$ in favor of $H_a$ if $F > F^{\alpha}_{t-1,t(r-1)}$ - the **critical value** obtained from the $F-$*tables*.

## Model II

$E(MST) = \frac{E[SST]}{t-1}$

Under model II, we can show that $MST$ is an unbiased estimate of

$$E[MST] = \sigma^2 + r\sigma_\tau^2.$$

$= \sum_{i=1}^{t}\sum_{j=1}^{r} \tau_i^2$

$var(\tau_i^2)$

$\sum_{i=1} r\tau_i$

We wish to test the hypotheses:

$$H_o : \sigma_\tau^2 = 0 \text{ versus } H_a : \sigma_\tau^2 > 0,$$

or, equivalently,

$H_o$ : the treatment or factor levels have identical effects

versus

$H_a$ : the treatment factor levels have different effects

at the $\alpha$ level of significance. If $H_o$ is true, $MSE$ and $MST$ are both unbiased estimates of $\sigma^2$, and if $H_a$ is true $MSE$ is still an unbiased estimate $\sigma^2$, but, $MST$ is an unbiased estimate of $\sigma^2 + r\sigma_\tau^2 > \sigma^2$. It follows that the test statistic for this testing problem is also $F$ given by equation (2.3). Why? Again the decision rule is that we reject $H_o$ in favor of $H_a$ if $F > F^{\alpha}_{t-1,t(r-1)}$.

We conclude by summarizing the results of this section in table 2.3 below.

$\sum_i \sum_j \left\{(y_{ij} - \bar{y}_{..}) - (y_{ij} - \bar{y}_{i.})\right\}^2 \quad (y_{ij} + \bar{y}_{i.} - y_{ij} - \bar{y}_{..}$

$\sigma^2 + \sum \quad y_{ij} - \bar{y}$

*Table 2.3 ANOVA table for a CRD*

| Source of Variation | $SS$ | $df$ | $MS$ | $F$ |
|---|---|---|---|---|
| Treatment | $SST$ | $t-1$ | $MST = \frac{SST}{t-1}$ | $\frac{MST}{MSE}$ |
| Error | $SSE$ | $t(r-1)$ | $MSE = \frac{SSE}{t(r-1)}$ | |
| Total | $SSTO$ | $tr-1$ | | |

**Exercise 2.5** *Refer to example 2.1. The brands differed through the percentage of ethanol added to pure petrol. Suppose that the 3 percentages of ethanol added to pure petrol were randomly chosen in the interval [0;50].*

(a) *Which of the two models (I and II) is appropriate for the experiment? Give reasons for your answer.*

(b) *Test the relevant hypotheses at the 0.05 level of significance.* ◇

### 2.3.3 Estimation

#### Variance Component Estimation

Under model II, if we reject $H_o$ that means $\sigma_\tau^2 > 0$ at the $\alpha$ level of significance. The next question the we may ask is "How large is $\sigma_\tau^2$?". This is an estimation problem called **variance component estimation.** We estimate $\sigma_\tau^2$ as shown below.

Since

$$E[MSE] = \sigma^2 \text{ and } E[MST] = \sigma^2 + r\sigma_\tau^2$$

then

$$\sigma_\tau^2 = \frac{1}{r}(E[MST] - E[MSE]).$$

Therefore, the most obvious estimate of $\sigma_\tau^2$ is:

$$\hat{\sigma}_\tau^2 = \max\{0, \frac{1}{r}(MST - MSE)\}$$

since $MST$ and $MSE$ are estimates of $E[MST]$ and $E[MSE]$, respectively.

**Exercise 2.6** *Refer to example 2.1. Suppose that model II is appropriate for the experiment. Is it necessary to estimate the variance ($\sigma_\tau^2$) of the brand effects? Give reasons for your answer. If so, estimate $\sigma_\tau^2$ and interpret it.* ◇

#### Pairwise Comparisons

We compare pairs of treatment means if the ANOVA under model I concludes that at least two treatment means are different at the $\alpha$ level of significance. We perform pairwise comparisons of the treatment means in order to determine which means are different. The comparisons are performed at the same level of significance as the ANOVA. There are many methods of making such comparisons. We present one of the simplest of the methods which is called the **least significant difference** (LSD) method.

The hypotheses to be tested are:

$$H_o: \mu_i = \mu_{i'} \text{ versus } H_a: \mu_i \neq \mu_{i'} \text{ for all } i \neq i'$$

at the $\alpha$ level of significance. The hypotheses are tested using the $t - test$ whose test statistic is given by

$$t = \frac{\bar{Y}_{i.} - \bar{Y}_{i'}}{\sqrt{2MSE/r}}$$

where $\sqrt{2MSE/r}$ is the standard error of $\bar{Y}_{i.} - \bar{Y}_{i'}$. The test statistic $t$ has a **Student t** distribution with $t(r-1)$ degrees of freedom under $H_o$ and the model I assumptions holding. We declare means $\mu_i$ and $\mu_{i'}$ significantly different at the $\alpha$ level of significance if

$$\left| \frac{\bar{Y}_{i.} - \bar{Y}_{i'}}{\sqrt{2MSE/r}} \right| > t^{\alpha/2}_{t(r-1)},$$

or, equivalently if

$$|\bar{Y}_{i.} - \bar{Y}_{i'.}| > t^{\alpha/2}_{t(r-1)} \sqrt{2MSE/r}.$$

The quantity

$$LSD = t^{\alpha/2}_{t(r-1)} \sqrt{2MSE/r}$$

is called the **least significant difference.** The LSD procedure simply compares the observed absolute difference between each pair of averages to the corresponding $LSD$. If $|\bar{Y}_{i.} - \bar{Y}_{i'.}| > LSD$, then conclude that the factor level means $\mu_i$ and $\mu_{i'}$ differ at the $\alpha$ level of significance.

**Exercise 2.7** *Refer to example 2.1. Suppose that model I is appropriate for the experiment. Is it necessary to perform pairwise comparisons of the brand means? Give reasons for your answer. If so, use the LSD method to perform all pairwise comparisons of the brand means. Which is the best brand of petrol?* ◇

## 2.4 Additional Topics

### 2.4.1 Unbalanced CRD / Missing Observations

We consider the analysis of a CRD for which the number of responses within each treatment are different. This may occur by design or by accident. For example, if it is desired to estimate some treatment means / effects more precisely than the rest, then those treatments can be replicated more than the rest in the experiment. Accidents such as destruction of the experimental material before the end of the experiment can also lead to unequal number of responses within the treatments. A CRD for which the number of responses within each treatment are different is said to be **unbalanced** and the resultant data is called **unbalanced data.** Otherwise, the design is balanced and the resultant data is **balanced data.** The analysis of variance for the unbalanced CRD is the same as that for the balanced one with some slight modifications in the formulae. We show these modifications.

1. The general model for the unbalanced case is model (2.1) but with $j = 1, 2, ..., r_i$ where $r_i$ is the number of replications of the $i^{th}$ treatment or the number of responses within the $i^{th}$ treatment. Thus, $n = \sum_{i=1}^{t} r_i$ becomes the total number of experimental units or responses. The $i^{th}$ treatment total and the estimate of the overall mean ($\mu$) are given by:

$$Y_{i.} = \sum_{j=1}^{r_i} Y_{ij} \text{ and } \bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^{t} \sum_{j=1}^{r_i} Y_{ij},$$

respectively.

*Table 2.4 Residuals for the mileage data*

| | Replication | | | | |
|---|---|---|---|---|---|
| Brand | 1 | 2 | 3 | 4 | 5 |
| *I* | -3.8 | -2.3 | -0.3 | 1.7 | 4.7 |
| *II* | -2.1 | -0.6 | 1.4 | 3.4 | -2.1 |
| *III* | -1.7 | -1.2 | 1.8 | 2.8 | -1.7 |

### Normality Assumption

A plot of the **histogram** of the residuals can help detect skewness in the distribution of the residuals. If the distribution of the errors ($\epsilon'_{ij}s$) is indeed normal with mean zero, then we expect the histogram of the residuals to be approximately mound shaped (bell shaped) about (on) zero. Otherwise, the normality assumption would be violated.

Another way of detecting nonnormality of the error distribution is to construct a **normal probability plot**. If the error distribution is normal then this plot should be approximately a straight line. Otherwise, the normality assumption would be violated.

**Exercise 2.9** *Refer to example 2.1. Construct a histogram and a normal probability plot of the residuals for model (2.1). Comment on the normal error assumption.* ◇

Moderate departures from the normality of the error distribution do not seriously invalidate the conclusions of the analysis of variance of the fixed effects model. They only cause the true level of significance to differ slightly from the specified value and the tests ($F$ and $t$) to be slightly less powerful in detecting treatment differences. Conclusions of the analysis of variance of the random effects model are more seriously affected by nonnormality in the sense that the estimates of the variance components may be very inaccurate.

### Constant Variance Assumption

We can check the assumption of constant error variance using a plot of the residuals ($\hat{\epsilon}_{ij}$) versus the treatment averages / predicted responses ($\bar{Y}_{i.}$). If model (2.1) is correct and if the error assumptions are not violated, then the errors and hence the residuals should not be related to any other variable including the predicted responses ($\bar{Y}_{i.}$). The plot of the residuals versus the treatment averages shows the violation of the constant error variance assumption if the plot reveals some obvious pattern such as magnitudes of the residuals increasing or decreasing with treatment averages.

**Exercise 2.10** *Refer to example 2.1. Check the constant error variance assumption.* ◇

The conclusions of the $F$ tests are slightly affected by the violation of the constant error variance assumption in the case of a fixed effects model for balanced data. Conclusions become invalid if the data is unbalanced and /or some error variances are much larger than others. For the random effects model, the conclusions are generally invalid even for balanced data if the constant error variance assumption is violated.

### Independence Assumption

When the responses are measured at successive time points, the errors in the responses may become related through time. For example, the skill of the experimenter may change as the experiment progresses resulting in a systematic change in the error variance over time. If this happens then the assumption of independent errors would be violated. We can detect the violation of this assumption by plotting the residuals versus time. If the plot shows any obvious pattern, then this would imply that the assumption of independence is violated. Once the assumption is violated all the conclusions of the analysis of variance become invalid.

2. The ANOVA table is as follows:

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{t}\sum_{j=1}^{r_i} Y_{ij}^2 - \frac{1}{n}Y_{..}^2 \sim n-1\ df \\
SST &= \sum_{i=1}^{t} \frac{1}{r_i}Y_{i.}^2 - \frac{1}{n}Y_{..}^2 \sim t-1\ df \\
SSE &= SSTO - SST \sim n-t\ df
\end{aligned}
$$

| Source of Variation | $SS$ | $df$ | $MS$ | $F$ |
|---|---|---|---|---|
| Treatment | $SST$ | $t-1$ | $MST = \frac{SST}{t-1}$ | $\frac{MST}{MSE}$ |
| Error | $SSE$ | $n-t$ | $MSE = \frac{SSE}{n-t}$ | |
| Total | $SSTO$ | $n-1$ | | |

3. The estimates of $\mu_i$, $\tau_i$ and $\sigma^2$ are: $\bar{Y}_{i.} = \frac{1}{r_i}Y_{i.}$, $\bar{Y}_{i.} - \bar{Y}_{..}$ and $MSE$, respectively. The estimate of $\sigma_\tau^2$ can be derived from $E[MST]$ and $E[MSE]$ and is given by

$$\hat{\sigma}_\tau^2 = \max\{0, \frac{1}{r^*}(MST - MSE)\} \text{ where } r^* = \frac{1}{t-1}\left(\sum_{i=1}^{t} r_i - \sum_{i=1}^{t} r_i^2 / \sum_{i=1}^{t} r_i\right).$$

4. The $LSD$ for comparing two treatment means $\mu_i$ and $\mu_{i'}$ $(i \neq i')$ is:

$$LSD = t_{n-t)}^{\alpha/2}\sqrt{MSE(\frac{1}{r_i} + \frac{1}{r_{'i}})}.$$

We note that if $r_1 = r_2 = ... = r_t = r$ then the formulae for the unbalanced case reduce to those of the balanced case.

### 2.4.2 Checking Model Assumptions

The underlying assumptions of the analysis of variance of this chapter and the subsequent chapters are that the data are adequately described by the specified model (model (2.1) in this chapter) and that the errors $(\epsilon_{ij}'s)$ are independent and normally distributed with mean 0 and constant variance $\sigma^2$. There are graphical and formal statistical methods of checking these assumptions. In this section we discuss some common graphical methods of checking the error assumptions. These methods are based on the estimates of the errors called the **residuals**.

We deduce from model (2.1) that $\epsilon_{ij} = Y_{ij} - \mu_i$. Since the estimate of $\mu_i$ is $\bar{Y}_{i.}$, the $i^{th}$ treatment average (also called the predicted mean response for the $i^{th}$ treatment), then the residuals for model (2.1) are given by:

$$\hat{\epsilon}_{ij} = Y_{ij} - \bar{Y}_{i.}.$$

For example, in table 2.2, the residual associated with the mileage per litre of the $2^{nd}$ car assigned brand $II$ petrol is $\hat{\epsilon}_{22} = 12.0 - 14.3$.

**Exercise 2.8** *Refer to example 2.1. Verify that the residuals for model (2.1) are as given in table 2.4.* ◇

## 2.5 Exercises

1. Refer to example 2.1. You are to repeat the experiment but with 5 Mazda 323's, 5 Toyota Camry's and 5 Mercedes Benz's. Would a CRD still be appropriate for the experiment? Give reasons for your answer.

2. An experiment was conducted to compare the number of major defectives observed along each of five production lines in which changes were being instituted. Production was monitored continuously during the period of changes, and the number of major defectives was recorded per day for each line. These data are shown here.

| Production line | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 34 | 54 | 75 | 44 | 80 |
| 44 | 41 | 62 | 43 | 52 |
| 32 | 38 | 45 | 30 | 41 |
| 36 | 32 | 10 | 32 | 35 |
| 51 | 56 | 68 | 55 | 58 |

   (a) Write down the fixed effects model for the data.

   (b) Perform an ANOVA of these data and draw conclusions. Use $\alpha = 0.05$.

   (c) Is it necessary to perform all pairwise comparisons of the production line means? Give reasons for your answer.

3. A horticulturist used a CRD to investigate the phosphorous content of tree leaves form 3 different varieties (1,2,3) of apple trees. Random samples of 5 leaves of each variety were analyzed for phosphorous content. An ANOVA of the data was performed.

   (a) Complete the incomplete ANOVA table below.

   (b) Are the mean phosphorous levels equal for the three varieties? Use the 0.05 level of significance.

| Source of Variation | $SS$ | $df$ | $MS$ | $F$ |
|---|---|---|---|---|
| Variety | .276 | ... | ... | ... |
| Error | ... | ... | ... | |
| Total | .374 | 14 | | |

4. A textile manufacturing company weaves a fabric on a large number of looms. The company is interested in loom-to-loom variability in tensile strength. To investigate this possibility, a manufacturing engineer selects four looms at random and makes four strengths determinations on fabric samples chosen at random from each loom. Then the engineer performs an ANOVA of the tensile strengths.

   (a) Complete the incomplete ANOVA table below.

   (b) What should the engineer conclude? Use the 0.05 level of significance.

   (c) Find a reasonable estimate of $\sigma_\tau^2$.

| Source of Variation | $SS$ | $df$ | $MS$ | $F$ |
|---|---|---|---|---|
| Loom | 90 | ... | ... | ... |
| Error | ... | ... | ... | |
| Total | 100 | 14 | | |

# Chapter 3
# Randomized Block Designs

## 3.1 What this Chapter is About

In this chapter we discuss the design and the analysis of an experiment for comparing two or more treatment (factor level) means when we suspect that

* the experimental units differ with respect to one or more identifiable variables that may affect the response of interest;

* there are identifiable external variables that may also affect the response of interest.

The effects of these variables on the response, if unaccounted for, is to inflate the experimental error. A **randomized block design** (RBD) is the most widely used block design for reducing the experimental error through allowing the effects of the variables to be accounted for during the analysis of the RBD.

The analysis of variance model (ANOVA) model for the RBD may be one of the following:

**Fixed effects model (model I):** a model that is used when the inferences about the treatments are to pertain to the set of treatments and the set of blocks included in the experiment.

**Random effects model (model II):** a model that is used when inferences about the treatments are to be extended to the populations from which the set of treatments and the set of blocks were randomly selected.

**Mixed effects model (model III):** a mixture of the fixed effects and the random effects models.

We shall only consider RBD's for which the fixed effects and the random effects models are appropriate.

By the end of this chapter you should be able to

* describe a RBD and state the conditions under which it is applicable;

* state the advantages and the disadvantages of using a RBD;

* analyst a RBD and draw appropriate conclusions.

## 3.2 Randomized Block Designs (RBD's)

Refer to exercise 4 in section 1.3 of chapter 1. A one-way ANOVA ignoring **Plot** obtained:

$$SSE \approx 410 \sim 9\ df \text{ and } MSE \approx 46,$$

and a two-way ANOVA with **Plot** as a second factor obtained:

$$SSE \approx 24 \sim 6\ df \text{ and } MSE \approx 4,$$

and the **Plot** sums of squares ($SSP$) equal to 386 with 3 degrees of freedom. What could have caused the reduction of the estimate of the experimental error ($MSE$) from 46 to 4? The answer to this question is the subject of discussion in this chapter. For now, we note that:

$$\frac{46 \times 9 - SSP}{9-3} \approx \frac{410-386}{6} \approx 4.$$

That is, the two-way ANOVA reduced the experimental error by removing the variation due to plot differences ($SSP = 386$) from the variation due to the experimental errors ($SSE = 410$).

We construct a RBD as follows:

1. The experimental units are arranged into groups (called **blocks**) in such a way that within each block the experimental units are relatively homogeneous (identical / uniform) with respect to one or more characteristics of the units that may influence the response of interest.

2. Complete randomization is then done independently within each block as you did in exercise 1.5 in section 1.2.4 of chapter 1.

In exercise 1.5 (in section 1.2.4 of chapter 1), the breed of the steer affects growth and therefore should be used as a blocking factor. The layout of the RBD is shown in the table below.

| | Block (Breed) | |
|---|---|---|
| Treatment (Diet) | Afrikander | Jersey |
| A | Steers $A_4, A_2$ | Steers $J_2, J_3$ |
| B | Steers $A_1, A_2$ | Steers $J_4, J_1$ |
| | used Set 1 | used Set 2 |

We also use a RBD when there are some external variables that may influence the response of interest. Consider the design of the experiment in exercise 1.2 in section 1.2.2 of chapter 1. While the plots within station may be uniform, the environmental and management conditions which also affect yield can vary from station to station. This means that the station should be used as a blocking factor.

**Exercise 3.1** *Refer to exercise 1.2 in section 1.2.2 of chapter 1. Display the layout of the RBD as was done in the above table. Use the following sets of random numbers to randomize the experiment:*

$$Set\ 1 = \{7,1,3,4,8,9,6,5,2\};\ Set\ 2 = \{8,4,6,2,1,9,7,5,3\};\ Set\ 3 = \{1,2,9,3,5,7,4,6,8\}.\ \diamond$$

Blocking allows us to remove the variation due to the differences among the blocks from the experimental error. The end result is that the conclusions from the experiment become more precise than when we do not block the experiment. Complete randomization within each block offers insurance against subjective and systematic biases thereby making the conclusions from the experiment more accurate.

The advantages of a RBD are the following:

1. If blocking is effective, then the RBD can provide substantially more precise conclusions than a CRD that uses the same experimental units.

2. It can be used with any number of treatments, blocks and replications within each block.

3. In the design, different treatments need not be assigned the same number of experimental units.

4. The statistical analysis of the experiment (two-way ANOVA) is relatively easy even when an entire treatment or block is dropped.

5. The design allows variable experimental units to be used without sacrificing the precision of the conclusions. This has the effect of widening the range of validity of the conclusions of the experiment.

The major disadvantages of a RBD are the following:

1. The statistical analysis of the experiment (ANOVA) is complicated if there are some missing observations with blocks.

2. The degrees of freedom for the experimental error ($MSE$) are lost to blocking (one for each block). The consequence of this is that if blocking is not effective then the estimate of the experimental error is inflated. Why?

3. The model for the RBD experiment is more complicated than that for a CRD experiment, and requires more assumptions. For example, assumptions about the block effects and the joint effects of the blocks and the treatments have to be made.

**Example 3.1** *Consider an experiment to compare the effects of analysts on the DNA concentration of plaque in 18-to-20-year-old females. Three female subjects (ages 18-20) were chosen for the study. Each subject was allowed to maintain her usual diet, supplemented with 30 mg of sucrose per day. No toothbrushing or mouthwashing was allowed during the study. At the end of the week, plaque was scraped from the entire dentition of each subject and divided into three samples. Three analysts were chosen, and each given an unmarked sample of plaque from each of the subjects and asked to analyst them for DNA content.*

*We note that in this experiment,*

* *the* Analysts *are the factor levels of the factor* Analyst*;*
* *the* DNA *concentration of plaque inherently varies from subject to subject, hence the* Subjects *are the blocks;*
* *the experimental units are the* samples *of plaque from the subjects;*
* *if both the analysts and the subjects are chosen at* random *from their respective populations, then random effects model is appropriate for the RBD experiment, otherwise, if the conclusions are to pertain to the specific three analysts and three subjects, then the fixed effects model is the appropriate one for the experiment.* ◇

**Exercise 3.2** *An experiment is to be conducted to compare the mileage (km) per litre obtained by competing brands of petrol (I, II, III). Three new Mazda 323's, 3 new Toyota Camry's and 3 new Nissan Sentra's are available for experimentation. During the experiment the cars will be operated under the same conditions in order to eliminate the effects of external variables on the distance travelled per litre of the assigned brand of petrol.*

(a) *What are the experimental units in the proposed experiment?*

(b) *Which of the CRD and RBD is appropriate for the experiment? Give reasons for your answer.*

(c) *In case your answer to (b) is RBD, what is the blocking factor?* ◇

## 3.3 Analysis of the RBD

Table 3.1 displays the layout of the data from a RBD with $b$ blocks, $t$ treatments and $t$ experimental units per block. The symbols in the table have the following meanings:

* $Y_{ij}$ is the response to the $i^{th}$ treatment in the $j^{th}$ block;
* $Y_{i.} = \sum_{j=1}^{b} Y_{ij}$, $Y_{.j} = \sum_{i=1}^{t} Y_{ij}$ and $Y_{..} = \sum_{i=1}^{t}\sum_{j=1}^{b} Y_{ij}$; are the $i^{th}$ treatment total, $j^{th}$ block total and the overall total, respectively;
* $\bar{Y}_{i.} = \frac{1}{b}Y_{i.}$, $\bar{Y}_{.j} = \frac{1}{t}Y_{.j}$ and $\bar{Y}_{..} = \frac{1}{bt}Y_{..}$ are the $i^{th}$ treatment mean, $j^{th}$ block mean and the overall mean, respectively;
* $i = 1, .., t$; $j = 1, .., b$; $n = bt - \#(\text{observations})$.

We shall use the following example to illustrate the analysis of a RBD.

**Example 3.2** *Refer to exercise 3.2. A RBD with type of car as a blocking factor was used and the data in the table below recorded. Analyst the data and draw conclusions.* ◇.

| | *Type of Car* | | | | |
|---|---|---|---|---|---|
| *Brand of petrol* | *Mazda* | *Camry* | *Sentra* | *Total* | *Mean* |
| *I* | 10.6 | 12.0 | 11.0 | 33.6 | 11.2 |
| *II* | 9.0 | 15.0 | 12.0 | 36.0 | 12.0 |
| *III* | 12.0 | 17.4 | 13.0 | 42.4 | 14.1 |
| *Total* | 31.6 | 44.4 | 36.0 | 112.0 | |
| *Mean* | 10.5 | 14.8 | 12.0 | | 12.2 |

*Table 3.1 Layout of a RBD*

| | Block | | | | | |
|---|---|---|---|---|---|---|
| Treatment | 1 | 2 | ... | b | Total | Mean |
| 1 | $Y_{11}$ | $Y_{12}$ | ... | $Y_{1b}$ | $Y_{1.}$ | $\bar{Y}_{1.}$ |
| 2 | $Y_{21}$ | $Y_{22}$ | ... | $Y_{2b}$ | $Y_{2.}$ | $\bar{Y}_{2.}$ |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| t | $Y_{t1}$ | $Y_{t2}$ | ... | $Y_{tb}$ | $Y_{t.}$ | $\bar{Y}_{t.}$ |
| Total | $Y_{.1}$ | $Y_{.2}$ | ... | $Y_{.b}$ | $Y_{..}$ | |
| Mean | $\bar{Y}_{.1}$ | $\bar{Y}_{.2}$ | ... | $\bar{Y}_{.b}$ | | $\bar{Y}_{..}$ |

### 3.3.1 Models

Both the **fixed effects model (model I)** and the random effects model (**model II**) for the RBD (table 3.1) have the form:

$$\begin{aligned} Y_{ij} &= \mu + \tau_i + \beta_j + \epsilon_{ij} \\ &= \mu_{ij} + \epsilon_{ij} \\ &= \mu_{i.} + \beta_j + \epsilon_{ij} \end{aligned} \tag{3.1}$$

where:

* $\mu$ is the overall population mean;
* $\tau_i$ is the $i^{th}$ treatment effect;
* $\beta_j$ is the $j^{th}$ block effect;
* $\mu_{i.} + \tau_i$ is the $i^{th}$ treatment mean;
* the $\epsilon'_{ij}s$ are random errors are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$;
* $i = 1, 2, ..., t;\ j = 1, 2, ..., b.$

**Remark 3.1** *The model also assumes that there are no joint treatment and block effects, i.e., there are no treatment by block* **interaction effects** *- an assumption which does not hold for some RBD experiments.* ◇

#### Model I

Model I is model (3.1) but with the $\tau'_i s$ and the $\beta'_j s$ regarded as **fixed real constants** satisfying the constraints:

$$\sum_{i=1}^{t} \tau_i = \sum_{j=1}^{b} \beta_j = 0.$$

Under this model, the $i^{th}$ treatment mean and the $j^{th}$ block mean are

$$\mu_{i.} = \frac{1}{b}\sum_{j=1}^{b} \mu_{ij} = \mu + \tau_i \text{ and } \mu_{.j} = \frac{1}{t}\sum_{i=1}^{t} \mu_{ij} = \mu + \beta_j,$$

respectively. The estimates of $\mu$, $\mu_{i.}$, $\mu_{.j}$, $\tau_i$ and $\beta_j$ are

$$\hat{\mu} = \bar{Y}_{..},\ \hat{\mu}_{i.} = \bar{Y}_{i.},\ \hat{\mu}_{.j} = \bar{Y}_{.j},\ \hat{\tau} = \bar{Y}_{i.} - \bar{Y}_{..} \text{ and } \hat{\beta} = \bar{Y}_{.j} - \bar{Y}_{..},$$

respectively. Our objective is to test the following hypotheses at the $\alpha$ level of significance.

$$H_o: \ \mu_{1.} = \mu_{2.} = ... = \mu_{t.} \text{ versus } H_a: \text{ not all } \mu_{i.} \text{ are equal,}$$

or, equivalently,

$$H_o: \ \tau_1 = \tau_2 = ... = \tau_t = 0 \text{ versus } H_a: \text{ not all } \tau_i = 0.$$

If we reject $H_0$, we wish to compare the treatment means to determine which means are different.

We may also want to test the hypotheses about the effectiveness of blocking. We test the hypotheses are about the effectiveness of blocking in order to determine whether or not blocking indeed increased the precision of the experiment, i.e., to determine whether or not future similar experiments should be blocked. The hypotheses to be tested are:

$$H_o: \text{ Blocking was not effective versus } H_a: \text{ Blocking was effective,}$$

or, equivalently,

$$H_o: \ \mu_{.1} = \mu_{.2} = ... = \mu_{.b} \text{ versus } H_a: \text{ not all } \mu_{.j}, \text{ are equal}$$

or, equivalently,

$$H_o: \ \beta_1 = \beta_2 = ... = \beta_b = 0 \text{ versus } H_a: \text{ not all } \beta_j = 0.$$

**Remark 3.2** *We use model I when the conclusions of the RBD experiment are to pertain to the particular set of treatments and the set of blocks used in the experiment.* ◇

### Model II

Model II is also model (3.1) but with the

* $\tau_i's$ random variables which are independent and normally distributed with mean 0 and variance $\sigma_\tau^2$;
* $\beta_j's$ random variables which are independent and normally distributed with mean 0 and variance $\sigma_\beta^2$;
* $\epsilon_{ij}'s$, $\tau_i's$ and $\beta_j's$ assumed to be independent random variables.

Under this model we wish to test the hypotheses:

$$H_o: \ \sigma_\tau^2 = 0 \text{ versus } H_a: \ \sigma_\tau^2 > 0,$$

or, equivalently,

$H_o$ : the treatment effects are identical

versus

$H_a$ : the treatment effects are not identical

If we reject $H_o$, we want to estimate $\sigma_\tau^2$.

We may also want to test the following hypotheses about the effectiveness of blocking.

$$H_o: \text{ Blocking was not effective versus } H_a: \text{ Blocking was effective,}$$

or, equivalently,

$$H_o: \ \sigma_\beta^2 = 0 \text{ versus } H_a: \ \sigma_\beta^2 > 0.$$

**Remark 3.3** *We use model II when the set of treatments and the set of the blocks used in the RBD experiment are random samples from their respective treatment and block populations, and the conclusions of the experiment are to be extended to these populations.* ◇

### 3.3.2 ANOVA

**Sums of Squares**

We measure the total variability in the observations ($Y'_{ij}s$) using the total sum of squares given by:

$$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{b} Y_{ij}^2 - \frac{1}{n}Y_{..}^2 \sim n-1\ df,\ n = br.$$

We partition $SSTO$ into the following three components:

* the **treatment sum of squares** ($SST$) which measure the variability in the observations that is due to treatment differences;
* the **block sum of squares** ($SSBlk$) which measure the variability in the observations that is due to block differences;
* the **error sum of squares** ($SSE$) which measure the variability in the observations that is due to random errors.

We decompose $SSTO$ as follows:

$$\begin{aligned}
\sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{..})^2 &= \sum_{i=1}^{t}\sum_{j=1}^{b}\left[(\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})\right]^2 \\
&= \underbrace{b\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2}_{SST} + \underbrace{t\sum_{j=1}^{b}(\bar{Y}_{.j} - \bar{Y}_{..})^2}_{SSBlk} + \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}_{SSE}
\end{aligned}$$

The computing formulae for $SST$, $SSBlk$ and $SSE$ follow.

$$\begin{aligned}
SST &= \sum_{i=1}^{t} b(\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^{t}\frac{1}{b}Y_{i.}^2 - \frac{1}{n}Y_{..}^2 \sim t-1\ df \\
SSBlk &= \sum_{j=1}^{b} t(\bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_{j=1}^{b}\frac{1}{t}Y_{.j}^2 - \frac{1}{n}Y_{..}^2 \sim b-1\ df \\
SSE &= SSTO - SST - SSBlk \sim n-t-b+1 = (t-1)(b-1)\ df
\end{aligned}$$

**Exercise 3.3** *Refer to example 3.2. Verify that* $SSTO = 49.34$, $SSBrand = 13.80$, $SSCar = 28.20$ *and* $SSE = 7.35$. *What are the degrees of freedom for these sums of squares.* ◇

**Exercise 3.4** *Refer to example 3.2. Calculate* $MSBrand$, $MSCar$ *and* $MSE$. ◇

**Hypotheses Testing**

Under both models I and II, $MSE = \dfrac{SSE}{(t-1)(b-1)}$ is an unbiased estimate of the experimental error ($\sigma^2$), i.e., $E[MSE] = \sigma^2$.

**Model I**

Under this model $MSBlk$ and $MST$ are unbiased estimates of:

$$\sigma^2 + \frac{1}{b-1}\sum_{j=1}^{b} t\beta_j^2 \text{ and } \sigma^2 \frac{1}{t-1}\sum_{i=1}^{t} b\tau_i^2,$$

### 3.3.2 ANOVA

**Sums of Squares**

We measure the total variability in the observations ($Y_{ij}'s$) using the total sum of squares given by:

$$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{b}Y_{ij}^2 - \frac{1}{n}Y_{..}^2 \sim n-1\ df,\ n = br.$$

We partition $SSTO$ into the following three components:

* the **treatment sum of squares** ($SST$) which measure the variability in the observations that is due to treatment differences;
* the **block sum of squares** ($SSBlk$) which measure the variability in the observations that is due to block differences;
* the **error sum of squares** ($SSE$) which measure the variability in the observations that is due to random errors.

We decompose $SSTO$ as follows:

$$\begin{aligned}\sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{..})^2 &= \sum_{i=1}^{t}\sum_{j=1}^{b}\left[(\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + Y_{..})\right]^2 \\ &= \underbrace{b\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2}_{SST} + \underbrace{t\sum_{j=1}^{b}(\bar{Y}_{.j} - \bar{Y}_{..})^2}_{SSBlk} + \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{b}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2}_{SSE}\end{aligned}$$

The computing formulae for $SST$, $SSBlk$ and $SSE$ follow.

$$\begin{aligned}SST &= \sum_{i=1}^{t}b(\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_{i=1}^{t}\frac{1}{b}Y_{i.}^2 - \frac{1}{n}Y_{..}^2 \sim t-1\ df \\ SSBlk &= \sum_{j=1}^{b}t(\bar{Y}_{.j} - \bar{Y}_{..})^2 = \sum_{j=1}^{b}\frac{1}{t}Y_{.j}^2 - \frac{1}{n}Y_{..}^2 \sim b-1\ df \\ SSE &= SSTO - SST - SSBlk \sim n-t-b+1 = (t-1)(b-1)\ df\end{aligned}$$

**Exercise 3.3** *Refer to example 3.2. Verify that* $SSTO = 49.34$, $SSBrand = 13.80$, $SSCar = 28.20$ *and* $SSE = 7.35$. *What are the degrees of freedom for these sums of squares.* ◇

**Exercise 3.4** *Refer to example 3.2. Calculate* $MSBrand$, $MSCar$ *and* $MSE$. ◇

**Hypotheses Testing**

Under both models I and II, $MSE = \dfrac{SSE}{(t-1)(b-1)}$ is an unbiased estimate of the experimental error ($\sigma^2$), i.e., $E[MSE] = \sigma^2$.

**Model I**

Under this model $MSBlk$ and $MST$ are unbiased estimates of:

$$\sigma^2 + \frac{1}{b-1}\sum_{j=1}^{b}t\beta_j^2 \text{ and } \sigma^2\frac{1}{t-1}\sum_{i=1}^{t}b\tau_i^2,$$

respectively. We can derive the test statistics of the hypotheses which follow from these expectations ($E[MST]$, $E[MSBlk]$ and $E[MSE]$). (See how it is done in section 2.3.2) of chapter 2.)

The hypotheses about the treatment means (or effects) are:

$$H_o : \mu_{1.} = \mu_{2.} = ... = \mu_{t.} \text{ versus } H_a : \text{ not all } \mu_{i.} \text{ are equal,}$$

or, equivalently,

$$H_o : \tau_1 = \tau_2 = ... = \tau_t = 0 \text{ versus } H_a : \text{ not all } \tau_i \text{ are equal.}$$

The test statistic for these hypotheses is:

$$F = \frac{MST}{MSE} \tag{3.2}$$

$F$ has an $F$ distribution with $t-1$ **numerator** degrees of freedom and $(t-1)(b-1)$ **denominator** degrees of freedom if $H_o$ is true and the model error assumptions hold. We note that under $H_o$ the value of $F$ is expected to be equal to 1 and under $H_a$ the value is expected to be large. Why? Therefore, we reject $H_o$ at the $\alpha$ level of significance in favor of $H_a$ if $F > F^{\alpha}_{t-1,(t-1)(b-1)}$ - the critical value.

We may wish to test the hypotheses are about the effectiveness of blocking. The hypotheses are:

$H_o$ : Blocking was not effective, i.e., $\beta_1 = \beta_2 = ... = \beta_b = 0$.

versus

$H_a$ : Blocking was effective, i.e., not all $\beta_j = 0$.

The test statistic for these hypotheses is:

$$F = \frac{MSBlk}{MSE} \tag{3.3}$$

This $F$ has an $F$ distribution with $b-1$ numerator degrees of freedom and $(t-1)(b-1)$ denominator degrees of freedom if $H_o$ is true and the model assumptions hold. We conclude that blocking was effective at the $\alpha$ level of significance if $F > F^{\alpha}_{b-1,(t-1)(b-1)}$ - the critical value.

$b + t + 1$

$tb - b + t - 1$

$b(t-1) -$

$tb - t - b - 1$

$t(b-1) - 1(b+1)$

$tb - (t-1) - (b-1)$

$tb - t + 1 - b + 1$

$tb - t - b + 1$

$t(b-1) - 1(b-1)$

$(t-1)(b-1)$

Model II

Under this model we have:

$$E[MSBlk] = \sigma^2 + t\sigma_\beta^2 \text{ and } E[MST] = \sigma^2 + b\sigma_\tau^2.$$

We can derive the test statistics which follow from these expectations.

The test statistic for:

$$H_o : \sigma_\tau^2 = 0 \text{ versus } H_a : \sigma_\tau^2 > 0,$$

or, equivalently,

$$H_o : \text{ the treatment effects are identical versus } H_a : \text{ the treatment effects are different}$$

is given by (3.2). Why? We reject $H_o$ in favor of $H_a$ if $F > F^{\alpha}_{t-1,(b-1)(t-1)}$.

The test statistic and the decision rule for the test of the effectiveness of blocking under model II are also the same as those under the fixed effects model (see formula (3.3)).

### 3.3.3 Estimation

#### Variance Component Estimation

Suppose that from the ANOVA test we conclude that $\sigma_\tau^2 > 0$ and $\sigma_\beta^2 > 0$. Then we may want to estimate both $\sigma_\beta^2$ and $\sigma_\tau^2$. We note that since

$$E[MST] = \sigma^2 + b\sigma_\tau^2 \text{ and } E[MSBlk] = \sigma^2 + t\sigma_\beta^2$$

then

$$\sigma_\tau^2 = \frac{1}{b}(E[MST] - \sigma^2) \text{ and } \sigma_\beta^2 = \frac{1}{t}(E[MSBlk] - \sigma^2).$$

Therefore, we can estimate $\sigma_\tau^2$ and $\sigma_\beta^2$ by

$$\max\{0, \frac{1}{b}(MST - MSE)\} \text{ and } \max\{0, \frac{1}{t}(MSB - MSE)\},$$

respectively.

**Exercise 3.5** *Refer to example 3.2. Assume that model II is appropriate for the experiment.*

(a) *Perform an ANOVA on the data and draw conclusions. Use* $\alpha = 0.05$.

(b) *Estimate the variance components if appropriate.* $\diamond$

#### Pairwise the Treatment Means

Pairwise comparisons of the treatment means are done only if the ANOVA test under the fixed effects model concludes that some factor level (treatment) means are different. In this case the hypotheses to be tested are:

$$H_o : \mu_{i.} = \mu_{i'.} \text{ versus } H_a : \mu_{i.} \neq \mu_{i'.} \text{ for all } i \neq i'.$$

We note that the comparisons should be made at the same level of significance as that used in the ANOVA. If we choose to use the least significant difference method for making the comparisons, then the least significant difference is:

$$LSD = t^{\alpha/2}_{(t-1)(b-1)}\sqrt{2MSE/b}.$$

The means $\mu_{i.}$ and $\mu_{i'.}$ are declared significantly different if

$$|\bar{Y}_{i.} - \bar{Y}_{i'.}| > LSD.$$

**Exercise 3.6** *Refer to example 3.2. Assume that model I is appropriate for the experiment. Is it necessary to perform pairwise comparisons of the brand means? If so, use the LSD method to compare the means.* $\diamond$.

### 3.3.4 Checking Model Assumptions

In section 2.4.2 we discussed the importance of checking the error assumptions and also presented some graphical methods of checking the assumptions. In this case and subsequent cases it is still important that we check the error assumptions in order to make sure that our inferences are valid. The graphical methods presented in section 2.4.2 apply in this case and subsequent cases. However, in this case the residuals for model (3.1) are given by:

$$\hat{\epsilon}_{ij} \doteq Y_{ij} - \hat{Y}_{ij}, \ \hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} + \bar{Y}_{..}$$

$\hat{Y}_{ij}$ is the predicted mean response to the $i^{th}$ treatment in the $j^{th}$ block, i.e., is an estimate of $\mu_{ij}$.

## 3.4 Exercises

1. Refer to example 3.1. Suppose that the analysts and the subjects were chosen at random. Furthermore, suppose that the plaque collected from each woman was subdivided into six samples and each analyst asked to analyst two unmarked samples (for DNA content) for each woman. The data obtained are in the table below.

   (a) Which model is appropriate for analyzing the data and why?

   (b) Are the effects of the combinations of the Analysts and the Subjects important? Use $\alpha = 0.05$.

   (c) In view of your conclusions in part $(c)$, is it necessary to test for the presence of the Analyst by Subject interactions? If so, are the Analyst by Subject interactions present? Use $\alpha = 0.05$.

   (d) In view of your answer or conclusions in part $(d)$, does it make sense to test for the significance of the subject effects or the analyst effects. If so, test the relevant hypotheses at the 0.05 level of significance.

   (e) In view of your answers or conclusions in parts $(b) - (e)$, which parameters of your model are worth estimating and why? Estimate the parameters.

| | Subject | | | |
|---|---|---|---|---|
| Analyst | 1 | 2 | 3 | Totals |
| 1 | 13.2, 12.3 | 10.6, 9.8 | 8.5, 8.9 | 63.3 |
| 2 | 12.5, 12.9 | 9.6, 10.7 | 7.9, 8.4 | 62.0 |
| 3 | 13.0, 12.4 | 9.9, 10.3 | 8.3, 8.6 | 62.5 |
| Totals | 76.3 | 60.9 | 50.6 | 187.8 |

2. An experiment was conducted to determine the effects of 4 different pesticides on the yield of fruit from 3 different varieties $(B_1, B_2, B_3)$ of a citrus tree. Eight trees from each variety were randomly selected from an orchard. The 4 pesticides were each randomly assigned to 2 trees of a particular variety. Yields of fruit, in bushels per tree, were obtained after the test period. The data follow.

| | Pesticide | | | |
|---|---|---|---|---|
| Variety | 1 | 2 | 3 | 4 |
| $B_1$ | 49,39 | 50,55 | 43,38 | 53,48 |
| $B_2$ | 55,41 | 67,58 | 53,42 | 85,73 |
| $B_3$ | 66,68 | 85,92 | 69,62 | 85,99 |

   (a) Identify the design of the experiment.

   (b) Perform an ANOVA of these data and draw conclusions. Use $\alpha = 0.05$.

   (c) Based on your conclusions in part $(c)$, is it appropriate to perform pairwise comparisons of the true mean yields for the 4 pesticides? *Give reasons for your answer.* If so, perform all pairwise comparisons of the true mean yields for the pesticides.

# Chapter 4
# Latin Square and Crossover Designs

## 4.1 What this Chapter is About

In chapter 3 we introduced the randomized block design (RBD) as the most widely used block design for reducing the experimental error. Another popular block design for reducing the experimental error is the **latin square design**. We use the latin square design to compare treatments when

* the experimental units can be grouped according to two identifiable characteristics that may affect the response of interest, or,

* the experimental units themselves are the blocks and there is an identifiable external variable (such as time) that may also affect the response of interest.

In this chapter we discuss the latin square design and its analysis. We also discuss the **crossover design** and its analysis. We assume that the treatment effects are fixed and hence adopt either the **fixed effects model** or the **mixed effects model** for analyzing the designs. The mixed effects model is simply a mixture of the fixed effects and the random effects models.

By the end of this chapter you should be able to

* describe a latin square design and state the conditions under which it is applicable;

* randomize a given standard latin square design;

* state the advantages and the disadvantages of using a latin square design;

* state the purposes of replicating latin squares;

* describe a crossover design and state the conditions under which it is applicable;

* analyst a latin square or crossover experiments and draw appropriate conclusions;

## 4.2 Latin Square Designs

In some experiments it is feasible to block the experiment by two blocking factors in order to reduce the experimental error. For example, if the experimental units are animals, and both the age and the sex of the animal affect the response of interest, then the age and the sex of the animal can be used as blocking factors in the experiment. As another example, let us consider an experiment in which treatments are applied on different days and the responses are measured by different observers. In such an experiment they may be a day to day variation in the responses as well as the observer to observer variation besides the variation in the responses due to the treatment effects and the random errors. Thus, they would be need to block such an experiment by day and observer. The appropriate design for eliminating the two sources of variability in the responses is the **latin square design**.

Suppose that we wish to design an experiment to compare the mean weight gains of steers that are fed diets A, B and C over a specified period. Furthermore, suppose that we have the following steers available for the experiment:

* 1-year olds: 1 Afrikander, 1 Tuli, 1 Jersey.

* 2-year olds: 1 Afrikander, 1 Tuli, 1 Jersey.

* 3-year olds: 1 Afrikander, 1 Tuli, 1 Jersey.

Animal scientists have proved that the weight gain of a steer also depends on the age and the breed of a steer besides its diet. Thus, we would need to block our experiment by the age and the breed of the steers in order to eliminate the variability in the weight gains that is due to the age and the breed of the steers. We can block the experiment using the latin square design that is displayed in table 4.1.

The design in table 4.1 is called a latin square design because it is a square arrangement (# columns = # rows = # diets) and the treatments (diets) are denoted by Latin letters A, B, C. Furthermore, the design is a **standard** latin square because the letters in the first row and first column are arranged alphabetically. **We note that in the design each diet appears once in each age group and once in each breed.**

*Table 4.1 A standard* $3 \times 3$ *latin square design*

| | | Breed | | |
|---|---|---|---|---|
| | | Afrikander | Tuli | Jersey |
| | 1 | A | B | C |
| Age | 2 | B | C | A |
| | 3 | C | A | B |

A general latin square design has the following features:

1. There are $t$ treatments denoted by Latin letters.
2. There are 2 blocking factors (called the column and the row blocking factors) each with $t$ blocks or levels.
3. The $t$ treatments are randomly assigned to the experimental units within the rows and columns so that each treatment appears once in every row and once in every column.
4. If the letters in the first row and first column of the square are arranged alphabetically, then the latin square design is called a **standard latin square design.**

There exist many latin squares for a given number of treatments. For example, for $t = 3$ there are 12 different latin square designs and for $t = 5$ there are 161280 possible latin square designs. For $t = 3$, 4 of the 12 possible latin square designs are the standard latin square design (see table 4.1) and

| | | Breed | | | Breed | | | Breed | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Afri. | Tuli | Jers. | Afri. | Tuli | Jers. | Afri. | Tuli | Jers. |
| | 1 | C | B | A | A | C | B | B | A | C |
| Age | 2 | A | C | B | B | A | C | C | B | A |
| | 3 | B | A | C | C | B | A | A | C | B |

If there exist many latin squares for a given number of treatments, then which one should we use for our experiment? The answer to this question is that we should choose the one to use at random. The process of choosing a latin square design ~~at~~ random is called **randomization of the latin square design**. Because it is difficult to list all possible latin squares for a given number of treatments, we randomly choose the latin square design to use with the help of computers.

A latin square design has the advantages that follow.

1. Double blocking an experiment often results in substantial reductions in the experimental error.
2. The design uses a small number of experimental units to study the treatment effects.
3. In experiments where the $t$ treatments are applied to each experimental unit in succession, the design is used to account for "Period" effects or the effects of the order of running the treatments. (We discuss this design, called a crossover design, in section 4.5 of this chapter.)

4. The analysis of the experiment is simple.

Refer to table 4.1. The latin square design uses 9 steers. If we were to use a RBD with the combination of the levels of age and breed as the blocks, then we would need more than 9 steers (at least 3 steers in each age-breed combination). Why?

We could conduct the same experiment using three steers of the same age, one from each breed. In this case the rows of the latin square design would be the periods within which each diet is fed the same steer. We note that randomizing the rows of this latin square design is actually randomizing the order in which the diets are to be fed each steer.

The following are the disadvantages of a latin square design when compared with a RBD.

1. The number of levels of each blocking factor must be equal to the number of treatments $t$. We show, in section 4.3.2, that this implies that the degrees of freedom for the experimental error ($MSE$) are equal to $(t-2)(t-1)$. Thus, the degrees of freedom for the experimental error are too few when a few treatments are studied and too large when many treatments are studied.

2. The latin square design can only be used when there are no joint block and treatment effects on the response.

3. The two blocking factors can not have different numbers of levels.

4. Randomizing a latin square experiment is more complicated than randomizing a CRD or a RBD experiment.

## 4.3 Analysis of a Latin Square Design

Let us consider the $4 \times 4$ latin square design that is displayed in table 4.2. For any number of treatments $t$ the symbols in the table have the following meanings:

* $Y_{ijk}$ is the response to the $k^{th}$ treatment in the $i^{th}$ row and $j^{th}$ column;

* $Y_{i..} = \sum_{j=1}^{t} Y_{ijk}$, $Y_{.j.} = \sum_{i=1}^{t} Y_{ijk}$ and $Y_{...} = \sum_{i=1}^{t}\sum_{j=1}^{t} Y_{ijk}$ are the $i^{th}$ row total, $j^{th}$ column total and the overall total, respectively;

* $\bar{Y}_{i..} = \frac{1}{t}Y_{i..}$, $\bar{Y}_{.j.} = \frac{1}{t}Y_{.j.}$ and $\bar{Y}_{...} = \frac{1}{t^2}Y_{...}$ are the $i^{th}$ row mean, $j^{th}$ column mean and the overall mean, respectively;

* $i = 1, 2, .., t$; $j = 1, 2, .., t$; $k = 1, 2, ..., t$ ($1 = \text{A}, 2 = \text{B}, ...,$ etc.).

**Remark 4.1** *Pay attention to the meanings of the symbols in table 4.2. In calculating the row, the column and the overall totals (and hence the means), the subscript for the treatments ($k$) is redundant. That is, we do not sum over $k$ because the treatments are uniquely specified by the rows and the columns of a latin square design. We shall use the following notation for the $k^{th}$ treatment total, the $k^{th}$ treatment mean and the* **treatment sum of squares**, *respectively:*

$$Y_{..k},\ \bar{Y}_{..k} = \frac{1}{t}Y_{..k} \text{ and } SST = \sum_{i=1}^{t}\sum_{j=1}^{t}(\bar{Y}_{..k} - \bar{Y}_{...})^2 \equiv \sum_{k=1}^{t}(\bar{Y}_{..k} - \bar{Y}_{...})^2. \diamond$$

The example that follows illustrates the calculations of the row, column, treatment and overall totals (and means) for a latin square experiment. We shall use the same data to illustrate the analysis the experiment.

**Example 4.1** *A supervisor wanted to determine whether or not there is a difference in the number of defective items produced by three different shifts: A, B, and C. He suspected that malfunctioning machines and operator errors were also responsible for the production of defective items besides the shifts. In order to get a clear picture about the effects of the shifts on the number of defective items produced he conducted an experiment using a latin square design with operator and machine as the blocking factors. The data collected are tabulated in table 4.3. Analyst the data and draw conclusions.* $\diamond$

*Table 4.2 Layout of a $4 \times 4$ latin square design*

| | | Column 1 | 2 | 3 | 4 | Total | Mean |
|---|---|---|---|---|---|---|---|
| Row | 1 | $Y_{111}$ (A) | $Y_{122}$ (B) | $Y_{134}$ (D) | $Y_{143}$ (C) | $Y_{1..}$ | $\bar{Y}_{1..}$ |
| | 2 | $Y_{212}$ (B) | $Y_{223}$ (C) | $Y_{231}$ (A) | $Y_{244}$ (D) | $Y_{2..}$ | $\bar{Y}_{2..}$ |
| | 3 | $Y_{313}$ (C) | $Y_{324}$ (D) | $Y_{332}$ (B) | $Y_{341}$ (A) | $Y_{3..}$ | $\bar{Y}_{3..}$ |
| | 4 | $Y_{414}$ (D) | $Y_{421}$ (A) | $Y_{433}$ (C) | $Y_{442}$ (B) | $Y_{4..}$ | $\bar{Y}_{4..}$ |
| | Total | $Y_{.1.}$ | $Y_{.2.}$ | $Y_{.3.}$ | $Y_{.4.}$ | $Y_{...}$ | |
| | Mean | $\bar{Y}_{.1.}$ | $\bar{Y}_{.2.}$ | $\bar{Y}_{.3.}$ | $\bar{Y}_{.4.}$ | | $\bar{Y}_{...}$ |
| Treatments | | A | B | C | D | | |
| | Total | $Y_{..1}$ | $Y_{..2}$ | $Y_{..3}$ | $Y_{..4}$ | | |
| | Mean | $\bar{Y}_{..1}$ | $\bar{Y}_{..2}$ | $\bar{Y}_{..3}$ | $\bar{Y}_{..4}$ | | |

**Solution 4.1** *We illustrate the calculation of the row, column, treatment and overall totals and means which are displayed in table 4.3.*

*Machine 1 total and mean:* $Y_{1..} = \sum_{j=1}^{3} Y_{ijk} = 23 + 31 + 51 = 105$; $\bar{Y}_{1..} = \frac{1}{3} Y_{1..} = 105/3 = 35$.

*Operator 1 total and mean:* $Y_{.1.} = \sum_{i=1}^{3} Y_{ijk} = 23 + 71 + 34 = 128$; $\bar{Y}_{.1.} = \frac{1}{3} Y_{.1.} = 128/3 = 42.67$.

*Overall total and mean:* $Y_{...} = \sum_{i=1}^{3} \sum_{j=1}^{3} Y_{ijk} = 23 + 31 + ... + 67 + 29 = 383$; $\bar{Y}_{...} = \frac{1}{9} Y_{...} = 383/9 = 42.55$.

*Shift A(=1) total and mean:* $Y_{..1} = 71 + 67 + 51 = 189$; $\bar{Y}_{..1} = \frac{1}{3} Y_{..1} = 189/3 = 63$.

*We perform the analysis of variance of the data in the section that follows.* ◇

*Table 4.3 A $3 \times 3$ latin square design for the #(defective) items produced*

| | | Operator 1 | 2 | 3 | Total | Mean |
|---|---|---|---|---|---|---|
| Machine | 1 | 23 (B) | 31 (C) | 51 (A) | 105 | 35.00 |
| | 2 | 71 (A) | 42 (B) | 35 (C) | 148 | 39.33 |
| | 3 | 34 (C) | 67 (A) | 29 (B) | 130 | 43.33 |
| | Total | 128 | 140 | 115 | 383 | |
| | Mean | 42.67 | 46.67 | 38.33 | | 42.55 |
| Shifts | | A | B | C | | |
| | Total | 189 | 94 | 100 | | |
| | Mean | 63.00 | 31.33 | 33.33 | | |

### 4.3.1 Models

The **fixed effects model (model I)** and the **mixed effects model (model III)** for a latin square experiment with $t$ treatments have the form:

$$\begin{aligned} Y_{ijk} &= \mu + \rho_i + \beta_j + \tau_k + \epsilon_{ijk} \\ &= \mu_{ijk} + \epsilon_{ijk} \\ &= \mu_{..k} + \rho_i + \beta_j + \epsilon_{ijk} \end{aligned} \tag{4.1}$$

where:

* $\mu$ is the overall population mean;
* $\rho_i$ is the $i^{th}$ row effect;
* $\beta_j$ is the $j^{th}$ column effect;
* $\tau_k$ is the $k^{th}$ treatment effect;
* $\mu_{..k} = \mu + \tau_k$ is the $i^{th}$ treatment mean;
* the $\epsilon'_{ijk}s$ are random error errors which are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$.

**Remark 4.2** *Another assumption of the model (or the latin square design) is that there is no interactions between the rows, columns and the treatments.* $\diamond$

**Remark 4.3** *The residuals for checking the error assumptions of model 4.1 are given by:*

$$\hat{\epsilon}_{ijk} = Y_{ijk} - \hat{Y}_{ijk}, \ \hat{Y}_{ijk} = \bar{Y}_{i..} + \bar{Y}_{.j.} + \bar{Y}_{..k} - 2\bar{Y}_{...} \diamond$$

We distinguish between models I and III.

### Model I

Model I is model (4.1) but with the $\rho'_i s$, $\beta'_j s$ and $\tau'_k s$ regarded as fixed real constants satisfying the constraints:

$$\sum_{i=1}^{t} \rho_i = \sum_{j=1}^{t} \beta_j = \sum_{k=1}^{t} \tau_k = 0.$$

Under this model the estimates of the overall mean $\mu$ and the $k^{th}$ treatment mean are:

$$\hat{\mu} = \bar{Y}_{...} \text{ and } \hat{\mu}_{..k} = \bar{Y}_{..k},$$

respectively. Since $\mu_{..k} = \mu + \tau_k$, then the obvious estimate of $\tau_k$ is:

$$\hat{\tau}_k = \bar{Y}_{..k} - \bar{Y}_{...}.$$

Our objective is to test the following hypotheses:

$$H_o: \ \mu_{..1} = \mu_{..2} = ... = \mu_{..t} \text{ versus } H_a: \text{ not all } \mu'_{..k}s \text{ are equal}$$

or, equivalently,

$$H_o: \ \tau_1 = \tau_2 = ... = \tau_t = 0 \text{ versus } H_a: \text{ not all the } \tau'_k s \text{ are equal to zero.}$$

Furthermore, if we reject $H_o$, then we wish to determine which means are different.

### Model III

Model III is model (4.1) but with some effects regarded as fixed real constants and some regarded as random variables. We consider the case when treatment effects are fixed and both the rows and the columns of the latin square are random selections from their respective populations. For this latin square design the model III assumptions are that the

* $\rho'_i s$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma^2_\rho$;

* $\beta_j's$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma_\beta^2$;

* $\sum_{k=1}^{t} \tau_k = 0$; the $\epsilon'_{ijk}s$, the $\rho'_i s$ and the $\beta'_j s$ are independent random variables.

Under model III the hypotheses about the treatment effects to be tested are the same as those under model I.

**Exercise 4.1** *Refer to example 4.1. The company operates with the three shifts (A, B C), 20 machines, and 24 operators. The supervisor randomly chose the machines and the operators for his experiment. Which of the two models I and III is appropriate for his experiment and why?* ◇

### 4.3.2 ANOVA

Sums of squares

We attribute the total variability in the observations from a latin square design to the **row factor effects**, **column factor effects**, **treatment effects** and the **random errors**. As usual we measure the total variability in the observations using the total sums of squares ($SSTO$) given by:

$$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{t}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{t}\sum_{j=1}^{t} Y_{ijk}^2 - \frac{1}{t^2}Y_{...}^2 \sim t^2 - 1 \; df.$$

We note that the subscript $k$ in the formula for $SSTO$ is redundant for the reasons that were given in remark 4.1.

We decompose the total sum of squares into

* the **row sum of squares** ($SSRow$) which measure the variability in the responses that is due to the row effects;

* the **column sum of squares** ($SSCol$) which measure the variability in the responses that is due to the column effects;

* the **treatment sum of squares** ($SST$) which measure the variability in the responses that is due to the treatment effects;

* the **error sum of squares** ($SSE$) which measure the variability in the responses that is due to the random errors.

$$\begin{aligned} SSTO &= \sum_{i=1}^{t}\sum_{j=1}^{t}\left[(\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{..k} - \bar{Y}_{...}) + (Y_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}_{...})\right]^2 \\ &= \underbrace{t\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2}_{SSRow} + \underbrace{t\sum_{j=1}^{t}(\bar{Y}_{.j} - \bar{Y}_{..})^2}_{SSCol} + \underbrace{t\sum_{k=1}^{t}(\bar{Y}_{..k} - \bar{Y}_{...})^2}_{SST(\text{see remark 4.1})} \qquad (4.2) \\ &+ \underbrace{\sum_{i=1}^{t}\sum_{j=1}^{t}(Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}_{...})^2}_{SSE} + \underbrace{\text{cross product terms}}_{=0} \end{aligned}$$

The computing formula for the sums of squares follow.

$$
\begin{aligned}
SSRow &= \frac{1}{t}\sum_{i=1}^{t} Y_{i..}^2 - \frac{1}{t^2}Y_{...}^2 \sim t-1\ df \\
SSCol &= \frac{1}{t}\sum_{j=1}^{t} Y_{.j.}^2 - \frac{1}{t^2}Y_{...}^2 \sim t-1\ df \\
SST &= \frac{1}{t}\sum_{k=1}^{t} Y_{..k}^2 - \frac{1}{t^2}Y_{...}^2 \sim t-1\ df \\
SSE &= SSTO - SSRow - SSCol - SST \sim (t-2)(t-1)\ df
\end{aligned}
$$

The degrees of freedom for $SSRow$, $SSCol$ and $SST$ are all equal to $t-1$ because there are $t$ treatments and $t$ levels for the row blocking factor and $t-1$ levels for the column blocking factors. As usual the degrees of freedom for $SSE$ are obtained by subtraction. That is,

$$(t^2-1)-(t-1)-(t-1)-(t-1) = (t-2)(t-1).$$

**Remark 4.4** *This formula for the degrees of freedom for $SSE$ was used to explain one of the disadvantages of using a latin square design (see section 4.2).* $\diamond$

**Exercise 4.2** *One of the disadvantages of using a latin square design is that it can only be used when there are no interactions between the rows, the columns and the treatments (see section 4.2). Show that if we were to include the row by column interaction term in model (4.1), then the degrees of freedom for $SSE$ would be negative which means that both the row by column interaction effects or the experimental error would be unestimable.* $\diamond$

### Hypotheses Testing

We obtain the mean sum of squares ($MS$) of the factors by dividing the sum of squares by their corresponding degrees of freedom. These are displayed in table 4.4. The column $E[MS]$ contains the model parameters being estimated by $MS$ under each of models I and III. It is clear from the E[MS] of both models I and III that to test for the treatment effects:

$$H_o:\ \tau_1 = \tau_2 = ... = \tau_t = 0 \text{ versus } H_a:\ \text{not all the } \tau_k's \text{ are equal to zero}$$

we should use

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

as the test statistic. If $H_o$ is true and the model error assumptions hold, then $F$ has an $F$ distribution with $t-1$ numerator degrees of freedom and $(t-1)(t-2)$ denominator degrees of freedom. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{t-1,(t-1)(t-2)}$ - the critical value obtained from the $F$ tables.

The test statistics for checking the effectiveness of blocking by the row blocking factor and the column blocking factor are:

$$F = \frac{MSRow}{MSE} \text{ and } F = \frac{MSCol}{MSE},$$

respectively. Both $F$ ratios have an $F$ distribution with $t-1$ numerator degrees of freedom and $(t-1)(t-2)$ denominator degrees of freedom if blocking is not effective and the model error assumptions hold.

**Example 4.2** *We perform an ANOVA of the data in table 4.3.*

$$\sum_{i=1}^{3}\sum_{j=1}^{3} Y_{ijk}^2 = (23)^2 + (31)^2 + (51)^2 + ... + (29)^2 = 18607.$$

*Table 4.4 ANOVA table for a latin square design.*

| Source of Var. | $MS$ | $F$ | $E[MS]$ Model I | $E[MS]$ Model III |
|---|---|---|---|---|
| Treatment | $MST = \frac{SST}{t-1}$ | $\frac{MST}{MSE}$ | $\sigma^2 + \frac{t}{t-1}\sum_{k=1}^{t}\tau_k^2$ | $\sigma^2 + \frac{t}{t-1}\sum_{k=1}^{t}\tau_k^2$ |
| Rows | $MSRow = \frac{SSRow}{t-1}$ | | $\sigma^2 + \frac{t}{t-1}\sum_{i=1}^{t}\rho_i^2$ | $\sigma^2 + t\sigma_\rho^2$ |
| Column | $MSCol = \frac{SSCol}{t-1}$ | | $\sigma^2 + \frac{r}{t-1}\sum_{j=1}^{t}\beta_j^2$ | $\sigma^2 + t\sigma_\beta^2$ |
| Error | $MSE = \frac{SSE}{(t-1)(t-2)}$ | | $\sigma^2$ | $\sigma^2$ |

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{3}\sum_{j=1}^{3} Y_{ijk}^2 - \frac{1}{9}Y_{...}^2 = 18607 - (383)^2/9 = 2308.22 \sim 8\ df. \\
SSMach. &= \frac{1}{3}\sum_{i=1}^{3} Y_{i..}^2 - \frac{1}{9}Y_{...}^2 = ((105)^2 + (148)^2 + (130)^2)/3 - (383)^2/9 = 310.89 \sim 2\ df. \\
SSOper. &= \frac{1}{3}\sum_{j=1}^{3} Y_{.j.}^2 - \frac{1}{9}Y_{...}^2 = ((128)^2 + (140)^2 + (115)^2)/3 - (383)^2/9 = 104.22 \sim 2\ df. \\
SSShifts &= \frac{1}{3}\sum_{k=1}^{3} Y_{..k}^2 - \frac{1}{9}Y_{...}^2 = ((189)^2 + (94)^2 + (100)^2)/3 - (383)^2/9 = 1886.89 \sim 2\ df. \\
SSE &= SSTO - SSMach. - SSOper. - SSShifts \\
&= 2308 - 311 - 104 - 1887 = 6.22 \sim 2\ df.
\end{aligned}
$$

*The results are summarized in the ANOVA table below.*

| *Source of Variation* | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| *Machines* | 2 | 310.89 | 155.45 | 49.98 |
| *Operators* | 2 | 104.22 | 52.11 | 16.76 |
| *Shifts* | 2 | 1886.89 | 943.45 | 303.34 |
| *Error* | 2 | 6.22 | 3.11 | |
| *Total* | 8 | 2308.22 | | |

*Under models I and III the hypotheses to be tested are*

$$H_o:\ \mu_{..A} = \mu_{..B} = \mu_{..C} \text{ versus } H_a:\ \text{not all the } \mu'_{..k}s \text{ are equal}$$

*or*

$$H_o:\ \tau_A = \tau_B = \tau_C = 0 \text{ versus } H_a:\ \text{not all the } \tau'_k s \text{ are equal to zero}$$

*at the 0.05 level of significance, where $\mu_{..k}$ $(k = A, B, C)$ is the mean number of defects for shift $k$.*

*For $\alpha = 0.05$, the critical value is $F_{2,2}^{0.05} = 19$. Therefore, we reject $H_o$ at the 0.05 level of significance since $F = 303.34 > F_{2,2}^{0.05} = 19$. This means the mean number of defective items produced by the three shifts are different.* ◇

**Exercise 4.3** *Refer to example 4.2. Check the effectiveness of blocking by the operator and the machines. Use $\alpha = 0.05$.* ◇

### 4.3.3 Estimation

**Pairwise comparisons of the means**

We perform pairwise comparisons of the treatment means only if the treatment effects are fixed and if the ANOVA test about the treatment means concludes that some treatment means are different. The least significant difference for comparing the means is

$$LSD = t^{\alpha/2}_{(t-1)(t-2)}\sqrt{2MSE/t}.$$

Why?

**Exercise 4.4** *Refer to example 4.2. Under the fixed effects model, which shift means are different?* ◇

## 4.4 Replication of Latin Squares

In section 4.2 we stated that one of the disadvantages of using a latin square design is that if the number of treatments $t$ is small, then the number of degrees of freedom for the experimental error are too few. This also means that the number of replications for each treatment are too few since each treatment is replicated $t$ times in the whole experiment. Consequently, a latin square design with a small number of treatments obtains

* an inaccurate estimate of the experimental error;
* unreliable estimates of the model parameters;
* tests which are insensitive to treatment differences (see section 1.2.3 of chapter 1).

The other disadvantage of using a latin square design that we stated in section 4.2 is that the design can not be used when there are interactions between the rows, columns and the treatments. We can circumvent or partially circumvent these disadvantages of using a latin square design by **replicating within the cells** of a latin square. This is feasible when

* two or more experimental units can be obtained for each cell defined by levels of the row blocking factor and the column blocking factor, or,
* the latin square experiment can be repeated two or more times using the same experimental units.

For example, consider the latin square design displayed in table 4.1. Replication with the cells of the square is feasible since it is possible to obtain at least two steers for each age-breed cell. Alternatively, replication within the cells can be done by repeating the experiment using the same nine steers. This is feasible if the weight gains are measurable over a short period of time.

Suppose that the number of replications within each cell is $r$. Then the model for the data is:

$$\begin{aligned} Y_{ijkm} &= \mu + \rho_i + \beta_j + \tau_k + \epsilon_{ijkm} \\ &= \mu_{ijk} + \epsilon_{ijkm} \\ &= \mu_{..k} + \rho_i + \beta_j + \epsilon_{ijkm} \end{aligned} \tag{4.3}$$

where:

* $Y_{ijkm}$ is the $m^{th}$ $(m = 1, 2, ..., r)$ response to the $k^{th}$ treatment in the $i^{th}$ row and $j^{th}$ column;
* $\mu$ is the overall population mean;
* $\rho_i$ is the $i^{th}$ row effect;
* $\beta_j$ is the $j^{th}$ column effect;
* $\tau_k$ is the $k^{th}$ treatment effect;

* $\mu_{..k} = \mu + \tau_k$ is the $k^{th}$ treatment mean;

* the $\epsilon_{ijkm}$ is the random error.

We redefine the totals and the means as follows:

* $Y_{ijk.} = \sum_{m=1}^{r} Y_{ijkm}$, $Y_{i...} = \sum_{j=1}^{t}\sum_{m=1}^{r} Y_{ijkm}$, $Y_{.j..} = \sum_{i=1}^{t}\sum_{m=1}^{r} Y_{ijkm}$ and $Y_{....} = \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{m=1}^{r} Y_{ijkm}$ are the $(ij)^{th}$ cell total, $i^{th}$ row total, $j^{th}$ column total and the overall total, respectively;

* $\bar{Y}_{ijk.} = \frac{1}{r}Y_{ijk.}$, $\bar{Y}_{i...} = \frac{1}{tr}Y_{i...}$, $\bar{Y}_{.j..} = \frac{1}{tr}Y_{.j..}$ and $\bar{Y}_{....} = \frac{1}{t^2r}Y_{....}$ are the $(ij)^{th}$ cell mean, $i^{th}$ row mean, $j^{th}$ column mean and the overall mean, respectively;

* $i = 1, 2, .., t$; $j = 1, 2, .., t$; $k = 1, 2, .., t$ $(1 = \text{A}, 2 = \text{B}, ..., \text{etc.})$.

The $k^{th}$ treatment total and the $k^{th}$ treatment mean are: $Y_{..k.}$ and $\bar{Y}_{..k.} = \frac{1}{tr}Y_{..k.}$, respectively.

The total sum of squares ($SSTO$) is as usual:

$$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{m=1}^{r}(Y_{ijkm} - \bar{Y}_{....})^2 = \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{m=1}^{r} Y_{ijkm}^2 - \frac{1}{t^2r}Y_{....}^2 \sim t^2r - 1\ df.$$

The row sum of squares, the column sum of squares, the treatment sum of squares and the error sum of squares are, respectively:

$$\begin{aligned}
SSRow &= \frac{1}{tr}\sum_{i=1}^{t} Y_{i...}^2 - \frac{1}{t^2r}Y_{....}^2 \sim t - 1\ df \\
SSCol &= \frac{1}{tr}\sum_{j=1}^{t} Y_{.j..}^2 - \frac{1}{t^2r}Y_{....}^2 \sim t - 1\ df \\
SST &= \frac{1}{tr}\sum_{k=1}^{t} Y_{..k.}^2 - \frac{1}{t^2r}Y_{....}^2 \sim t - 1\ df \\
SSE &= SSTO - SSRow - SSCol - SST \sim t^2r - 3t + 2\ df
\end{aligned}$$

**Remark 4.5** *We note that replicating the latin square design has increased the df for $SSE$ by $(r-1)t^2$.* ◇

We have summarized the results in table 4.5. The $F$ ratios are test statistics for testing hypotheses about the corresponding factor effects. Hypotheses testing under models I and III is as described in section 4.3.2.

**Example 4.3** *A team of educators were interested in determining the relative effectiveness of instruction methods A (Video instruction), B (traditional classroom) and C (programmed study) on student learning. They felt that the IQ and the age of the student could influence their study too much. To take these two factors into account they used a $3 \times 3$ latin square design with IQ and Age as the blocking factors. They had two students in each IQ - Age cell of the latin square. The scores (out of 100) for the instruction methods are in the table below. Analyst the data and draw conclusions. Use $\alpha = 0.05$.* ◇

| | | IQ | | | | |
|---|---|---|---|---|---|---|
| | | *High* | *Average* | *Low* | *Total* | *Mean* |
| *Age* | *20* | *40,50 (C)* | *40,40 (B)* | *50,40 (A)* | *260* | *43.33* |
| | *30* | *70,60 (B)* | *30,20 (A)* | *55,50 (C)* | *285* | *47.50* |
| | *40* | *20,30 (A)* | *70,80 (C)* | *25,25 (B)* | *250* | *41.67* |
| | *Total* | *270* | *280* | *245* | *795* | |
| | *Mean* | *45.00* | *46.67* | *40.83* | | *44.17* |
| *Method* | | *A* | *B* | *C* | | |
| | *Total* | *190* | *260* | *345* | | |
| | *Mean* | *31.67* | *43.33* | *57.50* | | |

*Table 4.5 ANOVA for a latin square design with r replications per cell*

| Source of Variation | df | SS | $MS$ | $F$ |
|---|---|---|---|---|
| Row | $t-1$ | $SSRow$ | $MSRow = \frac{SSRow}{t-1}$ | $\frac{MSRow}{MSE}$ |
| Column | $t-1$ | $SSCol$ | $MSCol = \frac{SSCol}{t-1}$ | $\frac{MSCol}{MSE}$ |
| Treatments | $t-1$ | $SST$ | $MST = \frac{SST}{t-1}$ | $\frac{MST}{MSE}$ |
| Error | $t^2r-3t+2$ | $SSE$ | $MSE = \frac{SSE}{t^2r-3t+2}$ | |
| Total | $t^2r-1$ | $SSTO$ | | |

**Solution 4.2**

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{m=1}^{2} Y_{ijkm}^2 = (40)^2+(50)^2+(40)^2+...+(25)^2 = 40575.$$

$$\begin{aligned}
SSTO &= \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{m=1}^{2} Y_{ijkm}^2 - \frac{1}{18}Y_{....}^2 = 40575-(795)^2/18 \\
&= 5462.5 \sim 17\ df. \\
SSAge &= \frac{1}{6}\sum_{i=1}^{3} Y_{i...}^2 - \frac{1}{18}Y_{....}^2 = ((260)^2+(285)^2+(250)^2)/6-(795)^2/18 \\
&= 108.33 \sim 2\ df. \\
SSIQ &= \frac{1}{6}\sum_{j=1}^{3} Y_{.j..}^2 - \frac{1}{18}Y_{....}^2 = ((270)^2+(280)^2+(245)^2)/6-(795)^2/18 \\
&= 108.33 \sim 2\ df. \\
SSMethod &= \frac{1}{6}\sum_{k=1}^{3} Y_{..k.}^2 - \frac{1}{18}Y_{....}^2 = ((190)^2+(260)^2+(345)^2)/6-(795)^2/18 \\
&= 2008.33 \sim 2\ df. \\
SSE &= SSTO - SSAge - SSIQ - SSMethod \\
&= 5462.5-108.33-108.33-2008.33 = 3237.51 \sim 11\ df.
\end{aligned}$$

*The results are summarized in the ANOVA table below.*

| *Source of Variation* | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| *Age* | *2* | *108.33* | *54.17* | *0.185* |
| *IQ* | *2* | *108.33* | *54.17* | *0.185* |
| *Method* | *2* | *2008.33* | *1004.16* | *3.412* |
| *Error* | *11* | *3237.51* | *294.32* | |
| *Total* | *17* | *5462.50* | | |

*We test the hypotheses about the effectiveness of the methods under models I and III. The hypotheses to be tested are*

$$H_o:\ \mu_{..A} = \mu_{..B} = \mu_{..C} \text{ versus } H_a:\ \text{not all the } \mu'_{..k}s \text{ are equal}$$

*or*

$$H_o: \ \tau_A = \tau_B = \tau_C = 0 \text{ versus } H_a: \text{ not all the } \tau_k's \text{ are equal to zero}$$

*at the 0.05 level of significance, where* $\mu_{..k}$ $(k = A, B, C)$ *is the mean score for method* $k$.

*For* $\alpha = 0.05$, *the critical value is* $F^{0.05}_{2,11} = 3.98$. *Therefore, we fail to reject* $H_o$ *at the 0.05 level of significance since* $F = 3.41 < F^{0.05}_{2,11} = 3.98$. *This means the instruction methods are equally effective (or ineffective).* ◇

Replication within the cells of a latin square allows us to test whether or not model (4.3) is appropriate for modelling the data. The model may not be appropriate for modelling the data for the reason that it does not include the interaction effects between the rows, columns and the treatments. If these interactions are present there are part of the experimental error. Replication within the cells allows us to account for the variation in the responses due to these interactions effects (if present) and other effects which are not in model (4.3).

With $r$ replications within the cells of a latin square we are able to estimate the pure experimental error even if model (4.3) is not appropriate for modelling the data. The **pure error sum of squares** ($SSPE$) are given by

$$SSPE = \sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{m=1}^{r} Y^2_{ijkm} - \frac{1}{r}\sum_{i=1}^{t}\sum_{j=1}^{r} Y^2_{ijk.} \sim t^2(r-1) \ df.$$

The error sum of squares ($SSE$) are a sum of the pure error sum of squares ($SSPE$) and the **lack of fit sum of squares** ($SSLF$). $SSLF$ measures the variation in the responses that is due to the effects that are not in model (4.3). Having calculated $SSE$ and $SSPE$,

$$SSLF = SSE - SSPE \sim t^2r - 3t + 2 - t^2(r-1) = (t-1)(t-2) \ df.$$

Note that the degrees of freedom associated with $SSLF$ are also found by subtraction as usual.

The **mean pure error sum of squares** ($MSPE$) and the **mean lack of fit sum of squares** ($MSLF$) are obtained by dividing the sum of squares by their corresponding degrees of freedom.

$$MSPE = \frac{1}{t^2(r-1)}SSPE \text{ and } MSLF = \frac{1}{(t-1)(t-2)}SSLF.$$

We wish to test the hypotheses:
$H_o$ : Model (4.3) is appropriate for the experiment
versus
$H_a$ : Model (4.3) is not appropriate for the experiment
The test statistic for the hypotheses is

$$F = \frac{MSLF}{MSPE}$$

which has an $F$ distribution with $(t-1)(t-2)$ numerator degrees of freedom and $t^2(r-1)$ degrees of freedom if $H_o$ is true. We reject $H_o$ if $F > F^{\alpha}_{(t-1)(t-2),t^2(r-1)}$.

**Exercise 4.5** *Refer to example 4.3. Test for the lack of fit of model (4.3) at the 0.05 level of significance.* ◇

**Remark 4.6** *There are other methods of replicating latin squares which we did not discuss in this section. These involve adding one or more latin squares to the row blocking factor or the column blocking factor. We discuss one of these in the next section.*

## 4.5 Crossover Designs

We consider a $t \times t$ latin square experiment in which the columns are the experimental units and the rows are the time periods at which the $t$ treatments under the unit are administered sequentially to the unit. In this experiment, each experimental unit receives a sequence of all the $t$ treatments in $t$ successive time periods, the treatment sequences being different for different experimental units. At the end of each time period, the

response to the treatment is measured on the experimental unit, and a period of time is allowed to pass in order to eliminate the effect of the current treatment before the next treatment is administered to the unit. This type of a latin square design is called a **crossover design**. For example, suppose we want to study the effects of diets A, B and C on the weight gain of steers. The layout of a crossover design might be as shown in table 4.6. At the end of each period, the weight gain of each steer would be recorded, and a period of time allowed to pass in which any physiological effect of the current diet would be eliminated before the steer is fed its next assigned diet.

*Table 4.6 A $3 \times 3$ latin square crossover design.*

| | | Steer | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| | 1 | A | B | C |
| Period | 2 | B | C | A |
| | 3 | C | A | B |

**Remark 4.7** *A simple crossover experiment such as described above is a special latin square experiment hence it is randomized the same way as a latin square experiment.* ◇

A crossover design can be a group of two or more $t \times t$ latin squares. With $rt$ experimental units, the crossover design is a group of $r$ $t \times t$ latin square designs with $t$ common time periods and $t$ different experimental units forming the columns of each latin square. For example, consider the experiment in which we want to study the effects of diets A, B and C on the weight gain of steers using $3r$ steers. For $r = 2$, the layout of the crossover experiment might be as shown in table 4.7.

*Table 4.7 A crossover design with two $3 \times 3$ latin square design.*

| | | Square | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | | | 2 | | |
| | | Steer | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| | 1 | A | B | C | A | B | C |
| Period | 2 | B | C | A | B | C | A |
| | 3 | C | A | B | C | A | B |

The model for a crossover design with $r$ $t \times t$ latin squares is

$$\begin{aligned} Y_{ijk} &= \mu + \rho_i + \beta_j + \tau_k + \epsilon_{ijk} \\ &= \mu_{ijk} + \epsilon_{ijk} \\ &= \mu_{..k} + \rho_i + \beta_j + \epsilon_{ijk} \end{aligned} \tag{4.4}$$

where:

* $Y_{ijk}$ is the response to the $k^{th}$ treatment administered to experimental unit $j$ during period $i$;

* $\mu$ is the overall population mean;

* $\rho_i$ is the period $i$ effect;

* $\beta_j$ is the experimental unit $j$ effect;

* $\tau_k$ is the $k^{th}$ treatment effect;

* $\mu_{..k} = \mu + \tau_k$ is the $k^{th}$ treatment mean;

* the $\epsilon_{ijk}$ is the random error;

* $i = 1, 2, .., t;\ j = 1, 2, .., rt;\ k = 1, 2, .., t.$

The totals and the means are as follows:

* $Y_{i..} = \sum_{j=1}^{rt} Y_{ijk},\ Y_{.j.} = \sum_{i=1}^{t} Y_{ijk}$ and $Y_{...} = \sum_{i=1}^{t}\sum_{j=1}^{rt} Y_{ijk}$ are the $i^{th}$ period total, $j^{th}$ experimental unit total and the overall total, respectively;

* $\bar{Y}_{i..} = \frac{1}{rt} Y_{i..},\ \bar{Y}_{.j.} = \frac{1}{t} Y_{.j.}$ and $\bar{Y}_{...} = \frac{1}{t^2 r} Y_{...}$ are the $i^{th}$ period mean, $j^{th}$ experimental unit mean and the overall mean, respectively;

* $i = 1, 2, .., t;\ j = 1, 2, .., rt;\ k = 1, 2, ..., t$ (1 = A, 2 = B, ..., etc.).

The $k^{th}$ treatment total and the $k^{th}$ treatment mean are $Y_{..k},\ \bar{Y}_{..k} = \frac{1}{tr} Y_{..k}$, respectively.

The total sum of squares ($SSTO$) is as usual:

$$SSTO = \sum_{i=1}^{t}\sum_{j=1}^{rt}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{t}\sum_{j=1}^{rt} Y_{ijk}^2 - \frac{1}{t^2 r} Y_{...}^2 \sim t^2 r - 1\ df.$$

The **period sum of squares**, the **experimental unit** sum of squares, the treatment sum of squares, and the error sum of squares are, respectively:

$$\begin{aligned}
SSPeriod &= \frac{1}{tr}\sum_{i=1}^{t} Y_{i..}^2 - \frac{1}{t^2 r} Y_{...}^2 \sim t - 1\ df \\
SSUnit &= \frac{1}{t}\sum_{j=1}^{rt} Y_{.j.}^2 - \frac{1}{t^2 r} Y_{....}^2 \sim rt - 1\ df \\
SST &= \frac{1}{tr}\sum_{k=1}^{t} Y_{..k}^2 - \frac{1}{t^2 r} Y_{...}^2 \sim t - 1\ df \\
SSE &= SSTO - SSPeriod - SSUnit - SST \sim (rt-2)(t-1)\ df
\end{aligned}$$

We have summarized results in table 4.8. The $F$ ratios in table 4.8 are test statistics for testing hypotheses about the corresponding factor effects. Hypotheses testing under models I and III is as described in section 4.3.2.

**Example 4.4** *A crossover design was used to study the effects of three diets (A,B,C) on the daily weight gain of steers. A sufficiently long period of time was allowed to pass before a steer was fed its new diet in order to eliminate the effect of its previous diets on the response to the new diet. The following data were recorded (in grams per day).*

*Table 4.8 ANOVA for a latin square design with r replications per cell.*

| Source of Variation | df | SS | $MS$ | $F$ |
|---|---|---|---|---|
| Period | $t-1$ | $SSPeriod$ | $MSPeriod = \frac{SSPeriod}{t-1}$ | $\frac{MSPeriod}{MSE}$ |
| Unit | $rt-1$ | $SSUnit$ | $MSUnit = \frac{SSUnit}{rt-1}$ | $\frac{MSUnit}{MSE}$ |
| Treatments | $t-1$ | $SST$ | $MST = \frac{SST}{t-1}$ | $\frac{MST}{MSE}$ |
| Error | $(rt-2)(t-1)$ | $SSE$ | $MSE = \frac{SSE}{(rt-2)(t-1)}$ | |
| Total | $t^2r-1$ | $SSTO$ | | |

| | | Steer | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *1* | *2* | *3* | *4* | *5* | *6* | *Total* | *Mean* |
| *Period* | *1* | *50.5 (C)* | *40.0 (B)* | *40.5 (A)* | *40.5 (C)* | *40.5 (B)* | *50.0 (A)* | *262.0* | *43.67* |
| | *2* | *55.0 (B)* | *30.5 (A)* | *50.0 (C)* | *70.5 (B)* | *20.0 (A)* | *55.5 (C)* | *281.5* | *46.92* |
| | *3* | *30.0 (A)* | *75.5 (C)* | *30.0 (B)* | *30.5 (A)* | *70.5 (C)* | *40.0 (B)* | *276.5* | *46.08* |
| | *Total* | *136.0* | *146.0* | *120.5* | *141.5* | *131.0* | *145.5* | *820.0* | |
| | *Mean* | *45.33* | *48.67* | *40.17* | *47.17* | *43.67* | *48.50* | | *45.56* |
| *Diet* | | *A* | *B* | *C* | | | | | |
| | *Total* | *201.5* | *276.0* | *342.5* | | | | | |
| | *Mean* | *37.17* | *46.00* | *57.08* | | | | | |

*Perform an ANOVA on the data and draw conclusions. Use* $\alpha = 0.05$. $\diamond$

**Solution 4.3**

$$\sum_{i=1}^{3}\sum_{j=1}^{6} Y_{ijk}^2 = (50.5)^2 + (40)^2 + (40.5)^2 + ... + (40.5)^2 = 41463.$$

$$\begin{aligned}
SSTO &= \sum_{i=1}^{3}\sum_{j=1}^{6} Y_{ijk}^2 - \frac{1}{18}Y_{...}^2 = 41463 - (820)^2/18 \\
&= 4106.95 \sim 17\ df. \\
SSPeriod &= \frac{1}{6}\sum_{i=1}^{3} Y_{i..}^2 - \frac{1}{18}Y_{...}^2 = ((262)^2 + (281.5)^2 + (276.5)^2)/6 - (820)^2/18 \\
&= 34.20 \sim 2\ df. \\
SSSteer &= \frac{1}{3}\sum_{j=1}^{6} Y_{.j.}^2 - \frac{1}{18}Y_{...}^2 = ((136)^2 + (146)^2 + ... + (145.5)^2)/3 - (820)^2/18 \\
&= 206.36 \sim 5\ df. \\
SSDiet &= \frac{1}{6}\sum_{k=1}^{3} Y_{..k}^2 - \frac{1}{18}Y_{...}^2 = ((201.5)^2 + (276)^2 + (342.5)^2)/6 - (820)^2/18 \\
&= 1658.53 \sim 2\ df. \\
SSE &= SSTO - SSPeriod - SSSteer - SSDiet \\
&= 4106.95 - 34.19 - 206.36 - 1658.53 = 2207.86 \sim 8\ df.
\end{aligned}$$

*The results are summarized in the ANOVA table below.*

| *Source of Variation* | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| *Period* | *2* | *34.19* | *17.10* | *0.062* |
| *Steer* | *5* | *206.36* | *41.27* | *0.149* |
| *Diet* | *2* | *1658.53* | *829.26* | *3.000* |
| *Error* | *8* | *2207.86* | *275.98* | |
| *Total* | *17* | *4106.95* | | |

*We compare the diets under model III. The hypotheses to be tested are*

$$H_o: \ \mu_{..A} = \mu_{..B} = \mu_{..C} \text{ versus } H_a: \text{ not all the } \mu'_{..k}s \text{ are equal}$$

*or*

$$H_o: \ \tau_A = \tau_B = \tau_C = 0 \text{ versus } H_a: \text{ not all the } \tau'_k s \text{ equal are equal to zero}$$

*at the 0.05 level of significance, where* $\mu_{..k}$ $(k = A, B, C)$ *is the mean weight gain for diet* $k$.

*For* $\alpha = 0.05$, *the critical value is* $F^{0.05}_{2,8} = 4.44$. *Therefore, we fail to reject* $H_o$ *at the 0.05 level of significance since* $F = 3.00 < F^{0.05}_{2,8} = 4.44$. *This means that the three diets are not significantly different in terms of increasing the weight gain of steers.* ◇

In this section we discussed a simple crossover design, simple in the that sense that effect of one treatment administered to an experimental unit is eliminated before another treatment is administered to the unit. In some experiments it is not possible to totally eliminate the effects of previous treatments, called **carry-over effects**, on the response to a new a treatment by increasing times between treatments. However, there are special crossover designs that allow us to account for the variation in the responses that is due to the carry-over effects of the treatments. We shall not discuss these in this module.

## 4.6 Exercises

1. Randomize the $5 \times 5$ standard latin square below using the following sets of random numbers:

$$Set\ 1 = \{4, 2, 5, 3, 1\} \text{ and } Set\ 2 = \{5, 2, 4, 1, 3\}.$$

| | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| B | C | D | E | A |
| C | D | E | A | B |
| D | E | A | B | C |
| E | A | B | C | D |

2. The effect of three different catalysts (A,B,C) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit 6 runs of the experiment. Each run takes 1 1/2 hours, so that only 6 runs can be made in one day. The experimenter decides to run the experiment as a Latin square, so that day and batch effects may be accounted for. The data follow.

| | | Day 1 | Day 2 | Day 3 |
|---|---|---|---|---|
| Batch | 1 | 11,10 (C) | 7,6 (B) | 7,10 (A) |
| | 2 | 4,7 (B) | 8,10 (A) | 10,8 (C) |
| | 3 | 8,9 (A) | 8,8 (C) | 3,6 (B) |

(a) Write an appropriate statistical model for the data.

(b) Are the reaction times for the catalysts significantly different? If so, which are different? Use $\alpha = 0.05$.

(c) Test for the lack of fit of the model in part ($a$).

3. The effects of two drugs (A,B) on the duration of sleep were studied using a group of 8 patients. Four patients were randomly assigned to drug A during period 1 and the other four to drug B. During period 2, the two groups of patients switched drugs. The sleep duration data follow.

| | | Patient | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Period | 1 | 8.6(A) | 7.1(B) | 8.3(A) | 7.3(B) | 7.9(A) | 7.5(B) | 6.3(A) | 6.8(B) |
| | 2 | 8.0(B) | 7.5(A) | 7.4(B) | 8.4(A) | 7.3(B) | 7.6(A) | 6.4(B) | 7.5(A) |

(a) Identify the design.

(b) Give a model for this design. State all the relevant assumptions of the model (design).

(c) Analyst the data and draw conclusions. Use $\alpha = 0.05$.

# Chapter 5
# Factorial Designs

## 5.1 What this Chapter is About

In this chapter we discuss the design and the analysis of experiments whose objective is to study the effects of the **combination** of the levels of two or more factors on a response of interest. **Factorial designs** and their variations are most efficient for these types of experiments. In a factorial design all the possible combinations of the levels of two or more factors are investigated. For example, if there are two factors $A$ and $B$ with $a$ and $b$ levels, respectively, then there are $ab$ factor level combinations (treatments) to be investigated. Table 5.1 displays these $ab$ factor level combinations.

*Table 5.1 A two factor factorial design with factors A and B*

| | Factor $B$ | | | |
|---|---|---|---|---|
| Factor $A$ | $B_1$ | $B_2$ | ..... | $B_b$ |
| $A_1$ | $A_1B_1$ | $A_1B_2$ | ..... | $A_1B_b$ |
| $A_2$ | $A_2B_1$ | $A_2B_2$ | ..... | $A_2B_b$ |
| . | . | . | ..... | . |
| . | . | . | ..... | . |
| $A_a$ | $A_aB_1$ | $A_aB_2$ | ..... | $A_aB_b$ |

We regard each combination of the factor levels in table 5.1 as a treatment. This means a completely randomized design (see chapter 2) or any block design such as a randomized block design and a latin square design (see chapters 3 and 4) can be used with a factorial design. For example, if the experimental units are identical and the experimental conditions are uniform, then a completely randomized design is appropriate for a factorial experiment. Otherwise, the factorial experiment has to be blocked.

By the end of this chapter you should be able to

* state the advantages of using a factorial design;
* analyst a two-factor factorial experiment in a completely randomized design and draw appropriate conclusions;
* analyst a two-factor factorial experiment in a randomized block design and draw appropriate conclusions.

## 5.2 Factorial Designs

We describe a simple factorial design in either a completely randomized design (CRD) or a block design (BD). Suppose that we wish to study the effects of the combination of the levels of factors $A$ and $B$ and that each factor has two 2 levels. Let the levels of factor $A$ be $A_1$ and $A_2$, and the levels of factor $B$ be $B_1$ and $B_2$. The combinations of the levels of the two factors that are to be investigated are:

$$A_1B_1, \quad A_1B_2, \quad A_2B_1, \quad A_2B_2.$$

The factorial designs for the experiment are as follows:

**CRD:** If we have homogenous experimental units and the experimental conditions are uniform, then a completely randomized design is appropriate for the factorial experiment. In this case we would randomly assign the experimental units to the 4 factor level combinations. We would require a minimum of 4 experimental units for us to be able to compare the factor level means.

**BD:** If we have heterogeneous experimental units or / and if there are external variables that may influence the response, then a block design (randomized block design (RBD) or latin square design) is appropriate for the factorial experiment. If the randomized block design is appropriate, then we would randomly assign the experimental units within each block to the 4 factor level combinations. The design requires a minimum of 4 experimental units per block. On the other hand, if the latin square design is appropriate, then we would randomize the latin square as was described in section 4.2 of chapter 4. In the square the treatments would be the 4 factor level combinations. The design may require a minimum of 4 or 16 experimental units (see chapter 4).

We discuss the advantages of using a factorial design in the context of a factorial design in a completely randomized design. An alternative to a factorial design is an experiment whereby we investigate the effects of the two factors one factor at a time as follows:

* To compare $mean(B_1)$ and $mean(B_2)$, we would have to assign 2 experimental units to each of the $A_1B_1$ and the $A_1B_2$ combinations. Why? (see section 1.2.3 of chapter 1)

* To compare $mean(A_1)$ and $mean(A_2)$, we would have to assign an additional 2 experimental units to the $A_2B_1$ combination. Why?

Thus, the "cost" of comparing the factor level means is 6 experimental units compared to 4 experimental units in a factorial experiment. The comparisons of the factor level means using the 4 experimental units in a factorial experiment are just as precise as those using the 6 experimental units in a one factor at a time experiment. To see this compare the standard errors of the differences between the sample factor level means for the two experiments.

A factorial experiment allows us to check or test for the presence of the interaction effects between the factors because observations are taken at all the possible combinations of the factor levels. The one factor at a time experiment does not obtain information on the $A_2B_2$ combination which is required to check whether or not the $A$ by $B$ interactions are present. In the presence of the interactions an attempt at guessing the response to the $A_2B_2$ combination can result in misleading conclusions. For example, it does not necessarily follow that the response to $A_2B_2$ is the largest if the responses to $A_1B_2$ and $A_2B_1$ are larger than the response to $A_1B_1$.

## 5.3 Analysis of Factorial Designs

In this section we discuss the analysis of a two-factor factorial design when in a completely randomized design and also when in a randomized block design. In both cases we shall assume that the **fixed effects models** are appropriate for the factorial designs.

**Remark 5.1** *The analysis of multifactor factorial experiments in which some factor effects are fixed and some are random, or, all factor effects are random can be done but are too complicated.* ◇

### 5.3.1 Completely Randomized Design (CRD)

Let us consider the analysis of a two-factor factorial design in a completely randomized design. Suppose that the two factors are $A$ and $B$ with $a$ and $b$ levels, respectively. In this design observations are taken at all the $ab$ factor level combinations. Table 5.2 displays the layout of the factorial design when replicated $r$ times.

The symbols in table 5.2 have the following meanings:

* $Y_{ijk}$ is the $k^{th}$ response to the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$;

* $Y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}$, $Y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{r} Y_{ijk}$, $Y_{ij.} = \sum_{k=1}^{r} Y_{ijk}$ and $Y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}$ are totals of all the responses under the $i^{th}$ level of factor $A$, the $j^{th}$ level of factor $B$, the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$ and the whole experiment, respectively;

*Table 5.2 The two-factor factorial design in a CRD and with r replications*

| Factor $A$ | Factor $B$: $B_1$ | $B_2$ | ... | $B_b$ | Total | Mean |
|---|---|---|---|---|---|---|
| $A_1$ | $Y_{111}, Y_{112}, .., Y_{11r}$ | $Y_{121}, Y_{122}, .., Y_{12r}$ ... | $Y_{1b1}, Y_{1b2}, .., Y_{1br}$ | $Y_{1..}$ | $Y_{1..}$ | |
| $A_2$ | $Y_{211}, Y_{212}, .., Y_{21r}$ | $Y_{221}, Y_{222}, .., Y_{22r}$ ... | $Y_{2b1}, Y_{2b2}, .., Y_{2br}$ | $Y_{2..}$ | $Y_{2..}$ | |
| . | . | . | ... | . | . | . |
| . | . | . | ... | . | . | . |
| $A_a$ | $Y_{a11}, Y_{a12}, .., Y_{a1r}$ | $Y_{a21}, Y_{a22}, .., Y_{a2r}$ ... | $Y_{ab1}, Y_{ab2}, .., Y_{abr}$ | $Y_{a..}$ | $Y_{a..}$ | |
| Total | $Y_{.1.}$ | $Y_{.2.}$ | ... | $Y_{.b.}$ | $Y_{...}$ | |
| Mean | $\bar{Y}_{.1.}$ | $\bar{Y}_{.2.}$ | ... | $\bar{Y}_{.b.}$ | $\bar{Y}_{...}$ | |

* $\bar{Y}_{i..} = \frac{1}{br}Y_{i..}$, $\bar{Y}_{.j.} = \frac{1}{ar}Y_{.j.}$, $\bar{Y}_{ij.} = \frac{1}{r}Y_{ij.}$ and $\bar{Y}_{...} = \frac{1}{abr}Y_{...}$ are the sample means of the $i^{th}$ level of factor $A$, the $j^{th}$ level of factor $B$, the combination of $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$ and the overall mean, respectively;

* $i = 1, .., a;\ j = 1, .., b;\ k = 1, .., r.$

The following is a practical example of a two-factor factorial experiment in a CRD. We shall use the data to illustrate the analysis of the experiment.

**Example 5.1** *The yield of a chemical process is being studied. The two most important variables that are thought to affect the yield of the chemical process are the pressure and the temperature. Three levels of each factor are selected and a factorial experiment in a CRD and with two replications performed. The yield data follow. The numbers in ( ) are the cell totals. Analyst the data and draw conclusions.* ◇

| *Temperature* | *Pressure*: *200* | *215* | *230* | *Total* |
|---|---|---|---|---|
| *20* | *90.4*<br>*90.2*<br>*(180.6)* | *90.7*<br>*90.6*<br>*(181.3)* | *90.2*<br>*90.4*<br>*(180.6)* | *542.5* |
| *30* | *90.1*<br>*90.3*<br>*(180.4)* | *90.5*<br>*90.6*<br>*(181.1)* | *89.9*<br>*90.1*<br>*(180.0)* | *541.5* |
| *40* | *90.5*<br>*90.7*<br>*(181.2)* | *90.8*<br>*90.9*<br>*(181.7)* | *90.4*<br>*90.1*<br>*(180.5)* | *543.4* |
| *Total* | *542.2* | *544.1* | *541.1* | *1627.4* |

**Remark 5.2** *The analysis of the data usually begins by checking the presence of the A by B interactions using a graph of $\bar{Y}_{ij.}$ versus j for each i (or $\bar{Y}_{ij.}$ versus i for each j) plotted on the same graph paper. If the curves of the graphs of $\bar{Y}_{ij.}$ versus j for each i are almost parallel, then we conclude that there may be no A by B interactions.* ◇

**Example 5.2** *Refer to example 5.1. We check for the presence of the pressure by temperature interaction using a graph of the cell means versus the pressure levels for each temperature level. Figure 5.1 displays the graph. The curves are not parallel which suggests the presence of the pressure by temperature interaction effects. This means the model for the data should include a pressure by temperature interaction term.* ◇

## Model

The ANOVA model for a two-factor factorial design in a CRD (data in table 5.2) has the form:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

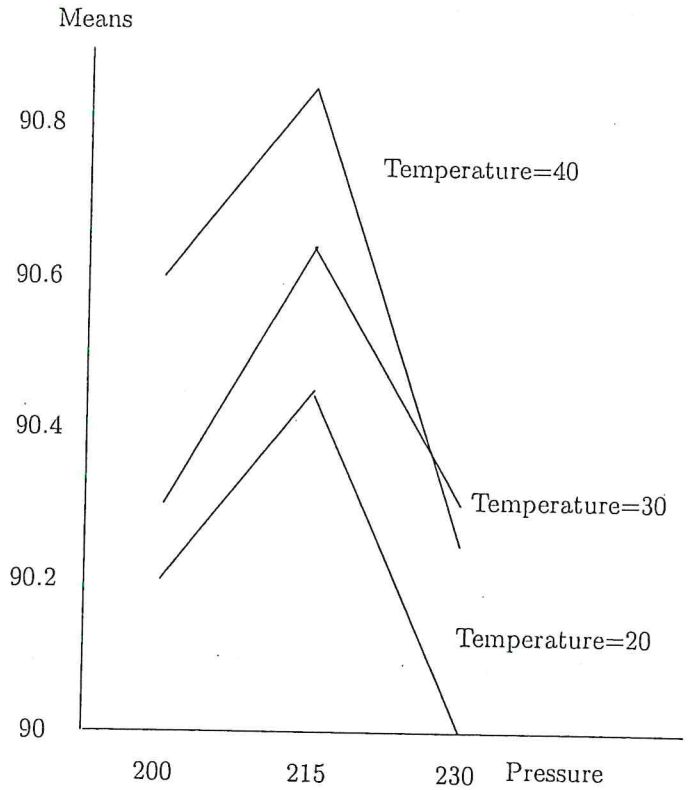


*Figure 5.1 Pressure - temperature plot for example 5.1*

$$= \mu_{ij} + \epsilon_{ijk} \tag{5.1}$$

where:

* $\mu$ is the overall population mean;
* $\tau_i$ is the effect of the $i^{th}$ level of factor $A$;
* $\beta_j$ is the effect of the $j^{th}$ level of factor $B$;
* $(\tau\beta)_{ij}$ is the interaction effect of the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$;
* $\mu_{ij}$ is the mean of the combination of the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$, i.e., the $A_iB_j$ treatment mean;
* the $\epsilon_{ijk}$ are the random errors which are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$.

The **fixed effects model** (model I) assumes that the $\tau_i's$, the $\beta_j's$ and the $(\tau\beta)_{ij}'s$ are fixed real constants satisfying the constraints:

$$\sum_{i=1}^{a}\tau_i = \sum_{j=1}^{b}\beta_j = \sum_{i=1}^{a}(\tau\beta)_{ij} = \sum_{j=1}^{b}(\tau\beta)_{ij} = 0.$$

Under this model the $i^{th}$ factor $A$ level mean and the $j^{th}$ factor $B$ level mean are:

$$\mu_{i.} = \sum_{j=1}\mu_{ij} = \mu + \tau_i \text{ and } \mu_{.j} = \sum_{i=1}\mu_{ij} = \mu + \beta_j,$$

respectively. The estimates of $\mu$, $\mu_{ij}$, $\mu_{i.}$ and $\mu_{.j}$ are:

$$\bar{Y}_{...},\ \bar{Y}_{ij.},\ \bar{Y}_{i..} \text{ and } Y_{.j.}$$

respectively. Hence, the estimates of $\tau_i$, $\beta_j$ and $(\tau\beta)_{ij}$ are:

$$\hat{\tau}_i = \bar{Y}_{i..} - \bar{Y}_{...},\ \hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...} \text{ and } \widehat{(\tau\beta)}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...},$$

respectively.

Our objective is to test the hypotheses which follow.

### Treatment Effects

The first hypotheses that we test are about the treatment effects. The hypotheses are:
$H_o$ : The treatments (factor level combinations) have no effect on the response
versus
$H_a$ : The treatments (factor level combinations) have a significant effect on the response,
or, equivalently,
$H_o$ : All the $\mu_{ij}$ are equal *versus* $H_a$ : Not all the $\mu_{ij}$ are equal.

### A by B Interaction Effects

Suppose that we reject $H_o$ (All the $\mu_{ij}$ are equal) in favor of $H_a$ (Not all $\mu_{ij}$ are equal). Then the next hypotheses that we test are about the $A$ by $B$ interaction effects. The hypotheses are:
$H_o$ : All the $(\tau\beta)'_{ij}s$ are equal to zero - There are no $A$ by $B$ interaction effects
versus
$H_a$ : Some $(\tau\beta)'_{ij}s$ are significantly different from zero - There are $A$ by $B$ interaction effects.
If we reject $H_o$ we compare the treatment (factor level combinations) means.

### A Effects and B Effects

Tests about the $A$ factor level effects or the $B$ factor level effects are meaningful only if there are no $A$ by $B$ interaction effects. Therefore, we test hypotheses about the $A$ factor level effects or the $B$ factor level effects only if the test about the $A$ by $B$ interaction effects concludes that the interaction effects are not present.

The hypotheses about $A$ factor level effects are:
$H_o$ : All the $\tau'_i s$ are equal to zero - The $A$ factor level means $(\mu'_{i.}s)$ are equal
versus
$H_a$ : Some $\tau'_i s$ are significantly different from zero - Some $A$ factor level means $(\mu'_{i.}s)$ are not equal.
If we reject $H_o$ we compare the $A$ factor level means.

The hypotheses about $B$ factor level effects are:
$H_o$ : All the $\beta'_j s$ are equal to zero - The $B$ factor level means $(\mu'_{.j}s)$ are equal
versus
$H_a$ : Some $\beta'_j s$ are significantly different from zero - Some $B$ factor level means $(\mu'_{.j}s)$ are not equal.
If we reject $H_o$ we compare the $B$ factor level means.

We discuss the analysis of the experiment.

ANOVA

## Sums of Squares

We measure the total variability in the observations using the total sum of squares which is given by:

$$SSTO = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{1}{abr}Y_{...}^2 \sim abr - 1 \ df$$

We regard the $ab$ factor level combinations as treatments. Hence, the total variability in the observations is attributed to the

* treatment (factor level combination) effects;

* random errors.

Consequently, we decompose $SSTO$ into

* the **treatment sum of squares** ($SST$) which measure the variability in the observations that is due to the treatment (factor level combination) effects;

* the **error sum of the squares** ($SSE$) which measure the variability in the observations that is due to the random errors.

We obtain $SST$ by performing a one-way ANOVA of the data with the factor level combinations $(A_iB_j)$ as the treatments. We note that the $Y'_{ij.}s$ and the $\bar{Y}'_{ij.}s$ are the treatment totals and the treatment means, respectively. This means the usually formula for the treatment sum of squares in a one-way ANOVA (see section 2.3.2 in chapter 2) obtains:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{...})^2 = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij.}^2 - \frac{1}{abr}Y_{...}^2 \sim ab - 1 \ df$$

We obtain the error sum of squares ($SSE$) and its degrees of freedom by subtraction, i.e.,

$$SSE = SSTO - SST \sim (abr - 1) - (ab - 1) = ab(r - 1) \ df.$$

We attribute the variability in the observations that is due to the treatments effects to:

* $A$ factor level effects;

* $B$ factor level effects;

* $A$ by $B$ interaction effects.

Thus, we decompose $SST$ into

* the $A$ sum of squares ($SSA$) which measure the variation due to the factor $A$ effects;

* the $B$ sum of squares ($SSB$) which measure the variation due to the factor $B$ factor effects;

* the $A$ by $B$ interaction sum of squares ($SSAB$) which measure the variation due to the $A$ by $B$ interaction effects.

We decompose $SST$ as follows:

$$\begin{aligned}
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{...})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left[(\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})\right]^2 \\
&= \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{i..} - \bar{Y}_{...})^2}_{=SSA} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{.j.} - \bar{Y}_{...})^2}_{=SSB} \\
&+ \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2}_{=SSAB} + \underbrace{cross\ product\ term}_{=0}
\end{aligned} \tag{5.2}$$

The computing formulae for $SSA$ and $SSB$ are:

$$SSA = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = \frac{1}{br}\sum_{i=1}^{a}\bar{Y}_{i..}^2 - \frac{1}{abr}Y_{...}^2 \sim a-1\ df \tag{5.3}$$

$$SSB = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 = \frac{1}{ar}\sum_{j=1}^{b}\bar{Y}_{.j.}^2 - \frac{1}{abr}Y_{...}^2 \sim b-1\ df \tag{5.4}$$

We obtain the $A$ by $B$ interaction sum of squares and its degrees of freedom by subtraction, i.e.,

$$\begin{aligned}
SSAB &= SST - SSA - SSB \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{...})^2 \sim (a-1)(b-1)\ df
\end{aligned} \tag{5.5}$$

We have summarized the results in table 5.3. The column $MS$ contains the **mean sum of squares** and the column $F$ contains **test statistics** for testing hypotheses about the corresponding effects in the column "Sources of Variation".

*Table 5.3 ANOVA table a two-factor factorial design in a CRD*

| Source of Variation | $df$ | $SS$ | $MS$ | $F$ |
|---|---|---|---|---|
| Treatment | $ab-1$ | $SST$ | $MST = \frac{SST}{ab-1}$ | $\frac{MST}{MSE}$ |
| A | $a-1$ | $SSA$ | $MSA = \frac{SSA}{a-1}$ | $\frac{MSA}{MSE}$ |
| B | $b-1$ | $SSB$ | $MSB = \frac{SSB}{b-1}$ | $\frac{MSB}{MSE}$ |
| AB | $(a-1)(b-1)$ | $SSAB$ | $MSAB = \frac{SSAB}{(a-1)(b-1)}$ | $\frac{MSAB}{MSE}$ |
| Error | $ab(r-1)$ | $SSE$ | $MSE = \frac{SSE}{ab(r-1)}$ | |
| Total | $abr-1$ | $SSTO$ | | |

**Example 5.3** *Refer to example 5.1. We show the construction of the ANOVA table for the data in example 5.1.*

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{1}{abk}Y_{...}^2 \qquad (5.6)\\
&= (90.4)^2 + (90.2)^2 + (90.2)^2 + ... + (90.1)^2 - (1627.4)^2/18\\
&= 1.29778 \sim 17\ df\\
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{1}{r}Y_{ij.}^2 - \frac{1}{abr}Y_{...}^2\\
&= (180.6)^2/2 + (181.3)^2/2 + ... + (180.5)^2/2 - (1627.4)^2/18\\
&= 1.13778 \sim 8\ df\\
SSE &= SSTO - SST = 0.1600 \sim 17 - 8 = 9\ df\\
SSTemperature &= \frac{1}{br}\sum_{i=1}^{a} \bar{Y}_{i..}^2 - \frac{1}{abr}Y_{...}^2\\
&= (542.5)^2/6 + (541.5)^2/6 + (543.4)^2/6 - (1627.4)^2/18\\
&= 0.30111 \sim 2\ df\\
SSPressure &= \frac{1}{ar}\sum_{j=1}^{b} \bar{Y}_{.j.}^2 - \frac{1}{abr}Y_{...}^2\\
&= (542.2)^2/6 + (544.1)^2/6 + (541.1)^2/6 - (1627.4)^2/18\\
&= 0.76778 \sim 2\ df\\
SSPressure \times Temperature &= SST - SSTemperature - SSPressure\\
&= 0.06889 \sim 8 - 2 - 2 = 4\ df
\end{aligned}
$$

$MSE = SSE/9 = 0.01778$; $MST = SST/8 = 0.14222$; $MSPressure = SSPressure/2 = 0.38389$; $MSTemperature = SSTemperature/2 = 0.15056$; $MSPressure \times Temperature = SSPressure \times Temperature/4 = 0.01722$.◇

## Hypotheses Testing

**Treatment Effects:** The first hypotheses that we always test are:

$H_o$ : The treatments (factor level combinations) have no effect on the response

versus

$H_a$ : The treatments (factor level combinations) have a significant effect on the response,

or, equivalently,

$H_o$ : All the $\mu_{ij}$ are equal *versus* $H_a$ : Not all the $\mu_{ij}$ are equal.

The test statistic for the hypotheses is:

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

which has an $F$ distribution with $ab - 1$ numerator degrees of freedom and $ab(r-1)$ denominator degrees of freedom if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance in favor of $H_a$ if $F > F^{\alpha}_{ab-1,ab(r-1)}$ - the critical value.

**A by B Interaction Effects:** The next hypotheses that we test after rejecting $H_o$ (All the $\mu_{ij}$ are equal) in favor of $H_a$ (Not all $\mu_{ij}$ are equal) are:

$H_o$ : All the $(\tau\beta)'_{ij}s$ are equal to zero - There are no $A$ by $B$ interaction effects

versus

$H_a$ : Some $(\tau\beta)'_{ij}s$ are significantly different from zero - There are $A$ by $B$ interaction effects.

The test statistic for the hypotheses is:

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

which has an $F$ distribution with $(a-1)(b-1)$ numerator degrees of freedom and $ab(r-1)$ denominator degrees of freedom if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance in favor of $H_a$ if $F > F^{\alpha}_{(a-1)(b-1),ab(r-1)}$.

**A Effects and B Effects:** We test hypotheses about the $A$ factor level effects or the $B$ factor level effects only if there are no $A$ by $B$ interaction effects. The hypotheses about $A$ factor level effects are:

$H_o$ : All the $\tau'_i s$ are equal to zero - The $A$ factor level means $(\mu'_{i.}s)$ are equal

versus

$H_a$ : Some $\tau'_i s$ are significantly different from zero - Some $A$ factor level means $(\mu'_{i.}s)$ are not equal.

The test statistic for the hypotheses is:

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

which has $a-1$ numerator degrees of freedom and $ab(r-1)$ denominator degrees of freedom if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance in favor of $H_a$ if $F > F^{\alpha}_{(a-1),ab(r-1)}$.

The hypotheses about $B$ factor level effects are:

$H_o$ : All the $\beta'_j s$ are equal to zero - The $B$ factor level means $(\mu'_{.j}s)$ are equal

versus

$H_a$ : Some $\beta'_j s$ are significantly different from zero - Some $B$ factor level means $(\mu'_{.j}s)$ are not equal.

The test statistic for the hypotheses is:

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

which has $b-1$ numerator degrees of freedom and $ab(r-1)$ denominator degrees of freedom if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance in favor of $H_a$ if $F > F^{\alpha}_{(b-1),ab(r-1)}$.

**Example 5.4** *We refer to examples 5.1 and 5.3. We test hypotheses about the pressure and the temperature effects under the assumption that the conclusions of the tests are to pertain to the specific pressure and temperature levels. That is, we assume that the fixed effects model holds for the experiment.*

**Pressure and Temperature Effects:**

$H_o$ : *The pressure and the temperature have no effect on yield*

*versus*

$H_a$ : *The pressure and the temperature have significant a effect on yield*

*at the 0.05 level of significance.*

*The test statistic for the hypotheses is* $F = \frac{MST}{MSE}$.

*We reject $H_o$ at the 0.05 level of significance and conclude that pressure and temperature have a significant effect on yield since $F = 7.999 > F^{0.05}_{8,9} = 3.23$ - the critical value.*

**Pressure by Temperature Interaction Effects:** *Since the pressure and the temperature have a significant effect on yield, then the next hypotheses to be tested is about the pressure by temperature interaction effects.*

$H_o$ : *There are no pressure by temperature interaction effects on yield*

*versus*

$H_a$ : *There are significant pressure by temperature interaction effects on yield*

*at the 0.05 level of significance.*

*The test statistic for the hypotheses is* $F = \dfrac{MSPressure \times Temperature}{MSE}$.

*We fail to reject* $H_o$ *at the 0.05 level of significance and conclude that there are no significant pressure by temperature interaction effects on yield since* $F = 0.968 < F_{4,9}^{0.05} = 3.63$ *- the critical value.*

**Pressure Effects and Temperature Effects:** *Tests about the pressure effects or the temperature effects on yield are meaningful since there are no significant pressure by temperature interaction effects on yield.*

*The hypotheses about the pressure effects are:*

$H_o$ : *There are no pressure effects on yield - The pressure level means are equal*

*versus*

$H_a$ : *There are significant pressure effects on yield - Some pressure level means are not equal*

*at the 0.05 level of significance.*

*The test statistic for the hypotheses is* $F = \dfrac{MSPressure}{MSE}$.

*We reject* $H_o$ *at the 0.05 level of significance and conclude that some pressure level means are not equal since* $F = 21.591 > F_{2,9}^{0.5} = 4.26$ *- the critical value.*

*The hypotheses about the temperature effects are:*

$H_o$ : *There are no temperature effects on yield - The temperature level means are equal*

*versus*

$H_a$ : *There are significant temperature effects on yield - Some temperature level means are not equal.*

*The test statistic for the hypotheses is* $F = \dfrac{MSTemperature}{MSE}$.

*We reject* $H_o$ *at the 0.05 level of significance and conclude that some temperature level means are not equal since* $F = 8.468 > F_{2,9}^{0.05} = 4.26$ *- the critical value.* $\diamond$

### Pairwise Comparisons

We perform pairwise comparisons of the treatment means or factor level means only if the ANOVA tests conclude that some treatment means are not equal. Why?

Suppose that the ANOVA tests conclude that some treatment means are not equal as well as that the $A$ by $B$ interaction effects are present. In this case, the pairwise comparisons of the $A$ or the $B$ factor level means does not make sense. This is because the presence of the $A$ by $B$ interactions implies that certain $A$ factor level means differ for some levels of factor $B$ and are the same for other levels factor $B$. Also, certain $B$ factor level means differ for some levels of factor $A$ and are the same for other levels of factor $A$ (e.g., see figure 5.1). This suggests that in the presence of the $A$ by $B$ interactions we should compare the $A$ factor level means at each level of $B$, or, compare the $B$ factor level means at each level of $A$.

Let us consider the pairwise comparisons of the $A$ factor level means at the $j^{th}$ level of factor $B$. For $i = 1, 2, .., a$ and $i' = 1, 2, .., a$, the hypotheses are:

$$H_o: \ \mu_{ij} = \mu_{i'j} \ versus \ H_a: \ \mu_{ij} \neq \mu_{i'j}, \ i \neq i'.$$

The means $\mu_{ij}$ and $\mu_{i'j}$ are estimated by $\bar{Y}_{ij.}$ and $\bar{Y}_{i'j.}$. The standard error of $\bar{Y}_{ij.} - \bar{Y}_{i'j.}$ is $\sqrt{2MSE/r}$. Why? The least significant difference for the tests is:

$$LSD = t^{\alpha/2}_{ab(r-1)}\sqrt{2MSE/r}.$$

Thus, means $\mu_{ij}$ and $\mu_{i'j}$ are significantly different if

$$\left|\bar{Y}_{ij.} - \bar{Y}_{i'j.}\right| > LSD.$$

Similarly, for $j = 1, 2, .., b$ and $j' = 1, 2, .., b$, the hypotheses for pairwise comparisons of the $B$ factor level means at the $i^{th}$ level of factor $A$ are:

$$H_o: \ \mu_{ij} = \mu_{ij'} \ versus \ H_a: \ \mu_{ij} \neq \mu_{ij'}, \ j \neq j'.$$

The least significant difference for the tests is the same as above, and the means are significantly different if

$$\left|\bar{Y}_{ij.} - \bar{Y}_{ij'.}\right| > LSD.$$

Suppose that the $ANOVA$ test concludes that the $A$ by $B$ interactions are absent. In this case pairwise comparisons of the $A$ factor level means or/and the $B$ factor level means can be done provided the ANOVA tests concluded that some $A$ or/and $B$ factor level means are not equal. Suppose that it is so for factor $A$. Then for $i = 1, 2, .., a$ and $i' = 1, 2, .., a$, the hypotheses to be tested are:

$$H_o: \ \mu_{i.} = \mu_{i'.} \ versus \ H_a: \ \mu_{i.} \neq \mu_{i'.}, \ i \neq i'$$

where:

$$\mu_{i.} = \mu + \tau_i \text{ and } \mu_{i'.} = \mu + \tau_{i'}$$

are the $i^{th}$ and $i'^{th}$ $A$ factor level means, respectively. The means $\mu_{i.}$ and $\mu_{i'.}$ are estimated by $\bar{Y}_{i..}$ and $\bar{Y}_{i'..}$, respectively. The standard error of $\bar{Y}_{i..} - \bar{Y}_{i'..}$ is $\sqrt{2MSE/br}$. Why? The least significant difference for the tests is:

$$LSD = t^{\alpha/2}_{ab(r-1)}\sqrt{2MSE/br}.$$

Thus, the means $\mu_{i.}$ and $\mu_{i'.}$ are significantly different if

$$\left|\bar{Y}_{i..} - \bar{Y}_{i'..}\right| > LSD.$$

Similarly, for $j = 1, 2, .., b$ and $j' = 1, 2, .., b$, the hypotheses for pairwise comparisons of the $B$ factor level means are:

$$H_o: \ \mu_{.j} = \mu_{.j'} \ versus \ H_a: \ \mu_{.j} \neq \mu_{.j'}, \ j \neq j'$$

where:

$$\mu_{.j} = \mu + \beta_j \text{ and } \mu_{.j'} = \mu + \beta_{j'}$$

are the $j^{th}$ and $j'^{th}$ $B$ factor level means, respectively. The least significant difference for the tests is:

$$LSD = t^{\alpha/2}_{ab(r-1)}\sqrt{2MSE/ar}.$$

The means $\mu_{.j}$ and $\mu_{.j'}$ are significantly different if

$$\left|\bar{Y}_{.j.} - \bar{Y}_{.j'.}\right| > LSD.$$

**Exercise 5.1** *Refer to examples 5.1 and 5.3. Perform pairwise comparisons of the relevant means. Which combination of the pressure and the temperature levels obtains maximum yield?* ◇

*Table 5.4 The layout of two-factor factorial design in a RBD.*

| Treatment | Block 1 | 2 | ... | $r$ | Total | Mean |
|---|---|---|---|---|---|---|
| $A_1B_1$ | $Y_{111}$ | $Y_{112}$ | ... | $Y_{11r}$ | $Y_{11.}$ | $Y_{11.}$ |
| $A_1B_2$ | $Y_{121}$ | $Y_{122}$ | ... | $Y_{12r}$ | $Y_{12.}$ | $Y_{12.}$ |
| . | . | . | ... | . | . | . |
| . | . | . | ... | . | . | . |
| $A_1B_b$ | $Y_{1b1}$ | $Y_{1b2}$ | ... | $Y_{1br}$ | $Y_{1b.}$ | $Y_{1b.}$ |
| $A_2B_1$ | $Y_{211}$ | $Y_{212}$ | ... | $Y_{21r}$ | $Y_{21.}$ | $Y_{21.}$ |
| $A_2B_2$ | $Y_{221}$ | $Y_{222}$ | ... | $Y_{22r}$ | $Y_{22.}$ | $Y_{12.}$ |
| . | . | . | ... | . | . | . |
| . | . | . | ... | . | . | . |
| $A_2B_b$ | $Y_{2b1}$ | $Y_{2b2}$ | ... | $Y_{2br}$ | $Y_{2b.}$ | $Y_{2b.}$ |
| . | . | . | ... | . | . | . |
| . | . | . | ... | . | . | . |
| $A_aB_1$ | $Y_{a11}$ | $Y_{a12}$ | ... | $Y_{a1r}$ | $Y_{a1.}$ | $Y_{a1.}$ |
| $A_aB_2$ | $Y_{a21}$ | $Y_{a22}$ | ... | $Y_{a2r}$ | $Y_{a2.}$ | $Y_{a2.}$ |
| . | . | . | ... | . | . | . |
| . | . | . | ... | . | . | . |
| $A_aB_b$ | $Y_{ab1}$ | $Y_{ab2}$ | ... | $Y_{abr}$ | $Y_{ab.}$ | $Y_{ab.}$ |
| Total | $Y_{..1}$ | $Y_{..2}$ | ... | $Y_{..r}$ | $Y_{...}$ | |
| Mean | $Y_{..1}$ | $Y_{..2}$ | ... | $Y_{..r}$ | $Y_{...}$ | |

### 5.3.2 Randomized Block Design (RBD)

Let us consider the analysis of a two-factor factorial design in a RBD. Suppose that the two factors are $A$ and $B$ with $a$ and $b$ levels, respectively. Furthermore, suppose that there are $r$ blocks. In this design measurements are made at all the $ab$ factor level combinations within each of the $r$ blocks. Table 5.4 displays the layout of the experiment.

The symbols in table 5.4 have the following meanings:

* $Y_{ijk}$ is the response to the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$ in the $k^{th}$ block;

* $Y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}$, $Y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{r} Y_{ijk}$, $Y_{..k} = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ijk}$, $Y_{ij.} = \sum_{k=1}^{r} Y_{ijk}$ and $Y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ijk}$ are totals of all the responses under the $i^{th}$ level of factor $A$, the $j^{th}$ level of factor $B$, the $k^{th}$ block, the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$ and in the whole experiment, respectively;

* $\bar{Y}_{i..} = \frac{1}{br}Y_{i..}$, $\bar{Y}_{.j.} = \frac{1}{ar}Y_{.j.}$, $\bar{Y}_{..k} = \frac{1}{ab}Y_{..k}$, $\bar{Y}_{ij.} = \frac{1}{r}Y_{ij.}$ and $\bar{Y}_{...} = \frac{1}{abr}Y_{...}$ are sample means for the $i^{th}$ level of factor $A$, the $j^{th}$ level of factor $B$, the $k^{th}$ block, the combination of $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$, and the overall mean, respectively;

* $i = 1, .., a$; $j = 1, .., b$; $k = 1, .., r$.

We shall use the example that follows to illustrate the analysis of a two-factor factorial design in a RBD.

**Example 5.5** *An experiment was set up to compare the effect of soil ph (factor A with levels 4, 5, 6, 7) and calcium additives (factor B with levels 100, 200, 300) on the increase in diameters for orange trees. All factor level combinations of these two factors were used at each of Limpopo, Mpumalanga and KZN Research Stations. Thus, the experiment was a 4 × 3 factorial design in a randomized block design - with research stations as the blocks. At the end of a two-year period the diameter of a randomly chosen tree at*

*each station and factor level combination was measured. The data appear in table 5.5. Assume that there are no interactions between the blocks and ph, and between the blocks and calcium levels. Analyst the data and draw conclusions.* ◇

| | Block | | | | |
|---|---|---|---|---|---|
| *Treatment*$(A, B)$ | *Mpumalanga* | *Limpopo* | *KZN* | *Total* | *Mean* |
| $(4, 100)$ | 5.2 | 5.9 | 6.3 | 17.4 | 5.800 |
| $(4, 200)$ | 7.4 | 7.0 | 7.6 | 22.0 | 7.333 |
| $(4, 300)$ | 6.3 | 6.7 | 6.1 | 19.1 | 6.367 |
| $(5, 100)$ | 7.1 | 7.4 | 7.5 | 22.0 | 7.333 |
| $(5, 200)$ | 7.4 | 7.3 | 7.1 | 21.8 | 7.267 |
| $(5, 300)$ | 7.3 | 7.5 | 7.2 | 22.0 | 7.333 |
| $(6, 100)$ | 7.6 | 7.2 | 7.4 | 22.2 | 7.400 |
| $(6, 200)$ | 7.6 | 7.5 | 7.8 | 22.9 | 7.633 |
| $(6, 300)$ | 7.2 | 7.3 | 7.0 | 21.5 | 7.167 |
| $(7, 100)$ | 7.2 | 7.5 | 7.2 | 21.9 | 7.300 |
| $(7, 200)$ | 7.4 | 7.0 | 6.9 | 21.3 | 7.100 |
| $(7, 300)$ | 6.8 | 6.6 | 6.4 | 19.8 | 6.600 |
| *Total* | 84.5 | 84.9 | 84.5 | 253.9 | |
| *Mean* | 7.042 | 7.075 | 7.042 | | 7.053 |

**Exercise 5.2** *Refer to example 5.5. If we can assume that there are no interactions between the blocks and ph, and between the blocks and calcium levels, then we can begin the analysis of the data by checking the presence of the ph by calcium level interactions using a graph of* $\bar{Y}_{ij.}$ *versus* $j$ *for each* $i$ *(or* $\bar{Y}_{ij.}$ *versus* $i$ *for each* $j$*) plotted on the same graph paper. Plot the graph. Does it suggest the presence of the ph by calcium interaction effects? Explain* ◇

### Model

The ANOVA model for a two-factor factorial design in a RBD (data in table 5.4) has the form:

$$\begin{aligned} Y_{ijk} &= \mu + \tau_i + \beta_j + \rho_k + (\tau\beta)_{ij} + \epsilon_{ijk} \\ &= \mu_{ij.} + \rho_k + \epsilon_{ijk} \\ &= \mu_{ijk} + \epsilon_{ijk} \end{aligned} \tag{5.7}$$

where:

* $\mu$ is the overall population mean;
* $\tau_i$ is the effect of the $i^{th}$ level of factor $A$;
* $\beta_j$ is the effect of the $j^{th}$ level of factor $B$;
* $\rho_k$ is the effect of the $k^{th}$ block;
* $(\tau\beta)_{ij}$ is the interaction effect of the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$;
* $\mu_{ij.}$ is the mean of the combination of the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$, i.e., the $A_iB_j$ treatment mean;
* $\mu_{ijk}$ is the mean of the combination of the $A_iB_j$ treatment and the $k^{th}$ block;
* the $\epsilon_{ijk}$ are the random errors which are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$.

**Remark 5.3** *Model (5.7) assumes that there are no interactions between the treatments and the blocks - an assumption which may not hold for some two-factor factorial designs in RBD's.* $\diamond$

We assume that the $\tau_i's$, the $\beta_j's$, the $(\tau\beta)_{ij}'s$ and the $\rho_k's$ are fixed real constants satisfying the constraints:

$$\sum_{i=1}^{a}\tau_i = \sum_{j=1}^{b}\beta_j = \sum_{i=1}^{a}(\tau\beta)_{ij} = \sum_{j=1}^{b}(\tau\beta)_{ij} = \sum_{k=1}^{r}\rho_k = 0.$$

**Remark 5.4** *Mixed and random effects models may be appropriate for some RBD two-factor factorial experiments. However, the analysis of such experiments is complicated and hence is beyond the scope of this module as was mentioned before.* $\diamond$

**Remark 5.5** *The estimates of the A and B factor level means as well as the estimates of the A factor level effects, the B factor level effects and the A by B interaction effects are the same as those in section 5.3.1. Furthermore, the hypotheses to be tested and the order in which they are tested are/is also the same as those/that in section 5.3.1.* $\diamond$

ANOVA

The total sum of squares is given by:

$$SSTO = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}Y_{ijk}^2 - \frac{1}{abr}Y_{...}^2 \sim abr - 1\ df \tag{5.8}$$

The total variability in the observations is attributed to the

* treatment (factor level combination) effects;

* block effects;

* random errors.

Consequently, we decompose $SSTO$ into

* the treatment sum of squares ($SST$);

* the block sum of squares ($SSBlk$);

* the error sum of the squares ($SSE$).

We obtain $SST$, $SSBlk$ and $SSE$ by performing a two-way ANOVA of the data with the factor level combinations $(A_iB_j)$ as the treatments. We note that the $Y_{ij.}'s$ and the $\bar{Y}_{ij.}'s$ are the respective treatment totals and the treatment means, and the $Y_{..k}'s$ and the $\bar{Y}_{..k}'s$ are the respective block totals and block means. Thus, the usually two-way ANOVA formulae for the treatment and the block sum of squares (see section 3.3.2 in chapter 3) give:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{...})^2 = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b}Y_{ij.}^2 - \frac{1}{abr}Y_{...}^2 \sim ab - 1\ df$$

and

$$SSBlk = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{..k} - \bar{Y}_{...})^2 = \frac{1}{ab}\sum_{k=1}^{r}Y_{..k}^2 - \frac{1}{abr}Y_{...}^2 \sim r - 1\ df.$$

We obtain the error sum of squares ($SSE$) by subtraction, i.e.,

$$SSE = SSTO - SST - SSBLK \sim (abr - 1) - (ab - 1) - (r - 1) = (ab - 1)(r - 1)\ df.$$

The variability in the observations that is due to the treatment effects is actually due to the

* $A$ factor level effects;

* $B$ factor level effects;

* $A$ by $B$ interaction effects.

Thus, we decompose $SST$ into

* the $A$ sum of squares ($SSA$);

* the $B$ sum of squares ($SSB$);

* the $A$ by $B$ interaction sum of squares ($SSAB$).

We decompose $SST$ as shown by equation (5.2) in section 5.3.1 of this chapter. The respective computing formulae for $SSA$, $SSB$ and $SSAB$ are also the same as those given by equations (5.3), (5.4) and (5.5) in the same section. The results of this section are summarized in table 5.5.

*Table 5.5 ANOVA table for a two-factor factorial design in a RBD*

| Source of Variation | $df$ | $SS$ | $MS$ | $F$ |
|---|---|---|---|---|
| Treatment | $ab-1$ | $SST$ | $MST = \frac{SST}{ab-1}$ | $\frac{MST}{MSE}$ |
| A | $a-1$ | $SSA$ | $MSA = \frac{SSA}{a-1}$ | $\frac{MSA}{MSE}$ |
| B | $b-1$ | $SSB$ | $MSB = \frac{SSB}{b-1}$ | $\frac{MSB}{MSE}$ |
| AB | $(a-1)(b-1)$ | $SSAB$ | $MSAB = \frac{SSAB}{(a-1)(b-1)}$ | $\frac{MSAB}{MSE}$ |
| Blocks | $r-1$ | $SSBlk$ | $MSBlk = \frac{SSBlk}{r-1}$ | $\frac{MSBlk}{MSE}$ |
| Error | $(ab-1)(r-1)$ | $SSE$ | $MSE = \frac{SSE}{(ab-1)(r-1)}$ | |
| Total | $abr-1$ | $SSTO$ | | |

**Example 5.6** *Refer to example 5.5. We show the construction of the ANOVA table for the data in example 5.5.*

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{1}{abk}Y_{...}^2 \qquad (5.9)\\
&= (5.2)^2 + (5.9)^2 + (6.3)^2 + ... + (6.8)^2 - (253.9)^2/36\\
&= 10.8098 \sim 35\ df\\
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{1}{r}Y_{ij.}^2 - \frac{1}{abr}Y_{...}^2\\
&= (17.4)^2/3 + (22.0)^2/3 + ... + (19.8)^2/3 - (253.9)^2/36\\
&= 9.18311 \sim 11\ df\\
SSStation &= \sum_{k=1}^{r} \frac{1}{ab}Y_{..k} - \frac{1}{abr}Y_{...}^2\\
&= (84.5)^2/12 + (84.9)^2/12 + (84.5)^2/12 - (253.9)^2/36\\
&= 0.00891 \sim 2\ df
\end{aligned}
$$

$$
\begin{aligned}
SSE &= SSTO - SST - SSStation = 1.6178 \sim 22\ df \\
SSPh &= \frac{1}{br}\sum_{i=1}^{a}\bar{Y}_{i..}^2 - \frac{1}{abr}Y_{...}^2 \\
&= (58.5)^2/9 + (65.8)^2/9 + (66.6)^2/9 + (63.0)^2/9 - (253.9)^2/36 \\
&= 4.4608 \sim 3\ df \\
SSCalcium &= \frac{1}{ar}\sum_{j=1}^{b}\bar{Y}_{.j.}^2 - \frac{1}{abr}Y_{...}^2 \\
&= (83.5)^2/12 + (88)^2/12 + (82.4)^2/12 - (253.9)^2/36 \\
&= 1.4672 \sim 2\ df \\
SSPh \times Calcium &= SST - SSPh - SSCalcium \\
&= 3.2551 \sim 6\ df
\end{aligned}
$$

$MSE = SSE/22 = 0.07353$; $MST = SST/11 = 0.83483$; $MSPh = SSPh/3 = 1.48694$; $MSCalcium = SSCalcium/2 = 0.73358$; $MSPh \times Calcium = SSPh \times Calcium/6 = 0.5425$; $MStation = SSStation/3 = 0.004455$. $\diamond$

## Hypotheses Testing

The hypotheses to be tested and the corresponding test statistics are the same as those in section 5.3.1 of this chapter. Furthermore, the order in which the hypotheses have to be tested is also the same as that in section5.3.1.

**Example 5.7** *We refer to examples 5.5 and 5.6. We test hypotheses about the effects of the ph and the amount of calcium on the diameters of orange trees under the model (5.7) assumptions.*

**Ph and Calcium Effects:** *The first hypotheses to be tested are:*

$H_o$ : *The ph and the amount of calcium have no effect on the diameters of orange trees (All the $\mu_{ij.}$ are equal)*

*versus*

$H_a$ : *The ph and the amount of calcium have significant a effect on the diameters of orange trees (Not all the $\mu_{ij.}$ are equal)*

*at the 0.05 level of significance.*

*The test statistic for the hypotheses is:*

$$F = \frac{MST}{MSE} \sim F_{11,22}$$

*if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the 0.05 level of significance and conclude that ph and the amount of calcium have a significant effect on the diameters of orange trees since $F = 11.35 > F_{11,22}^{0.05} = 2.26$ - the critical value.*

**Ph by Calcium Interaction Effects:** *Since the ph and the amount of calcium have a significant effect on the diameters of orange trees, then the next hypotheses to be tested is about the ph by calcium interaction effects. The hypotheses are:*

$H_o$ : *There are no ph by calcium interaction effects on the diameters of orange trees (All the $(\tau\beta)_{ij}$ are equal to zero)*

*versus*

*$H_a$ : There are significant ph by calcium interaction effects on the diameters of orange trees (Some $(\tau\beta)_{ij}$ are not equal to zero)*

*at the 0.05 level of significance.*

*The test statistic for the hypotheses is:*

$$F = \frac{MSPh \times Calcium}{MSE} \sim F_{6,22}$$

*if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the 0.05 level of significance and conclude that there are significant ph by calcium interaction effects on the diameters of orange trees since $F = 3.255 > F^{0.05}_{6,22} = 2.55$ - the critical value.* $\diamond$

**Exercise 5.3** *Refer to example 5.7.*

(a) *Is it necessary to test the hypotheses about the ph or amount of calcium effects on the diameters of orange trees? Explain.*

(b) *Was blocking by research station effective? Use $\alpha = 0.05$.* $\diamond$

## Pairwise Comparisons

Pairwise comparisons of the treatment means or the factor level means are performed only after the ANOVA test has concluded that some treatment means are not equal as was mentioned in section 5.3.2 of this chapter. In the presence of the $A$ by $B$ interactions the appropriate pairwise comparisons are the $A$ factor level means at each level of factor $B$, or the $B$ factor level means at each level of factor $A$ for the same reasons that we gave in section 5.3.1 of this chapter. For example, the hypotheses for pairwise comparisons of the $A$ factor level means at the $j^{th}$ level of factor $B$ are:

$$H_o: \ \mu_{ij.} = \mu_{i'j.} \ versus \ H_a: \ \mu_{ij.} \neq \mu_{i'j.}, \ i \neq i', \ i = 1, 2, .., a, \ i' = 1, 2, .., a.$$

The means $\mu_{ij.}$ and $\mu_{i'j.}$ are estimated by $\bar{Y}_{ij.}$ and $\bar{Y}_{i'j.}$. The standard error of $\bar{Y}_{ij.} - \bar{Y}_{i'j.}$ is $\sqrt{2MSE/r}$. Why? The least significant difference for the tests is:

$$LSD = t^{\alpha/2}_{(ab-1)(r-1)}\sqrt{2MSE/r}.$$

Thus, means $\mu_{ij.}$ and $\mu_{i'j.}$ are significantly different if

$$|\bar{Y}_{ij.} - \bar{Y}_{i'j.}| > LSD.$$

Similarly, for $j = 1, 2, .., b$ and $j' = 1, 2, .., b$, the hypotheses for pairwise comparisons of the $B$ factor level means at the $i^{th}$ level of factor $A$ are

$$H_o: \ \mu_{ij.} = \mu_{ij'.} \ versus \ H_a: \ \mu_{ij.} \neq \mu_{ij'.}, \ j \neq j'.$$

The least significant difference for the tests is the same as above, and the means are significantly different if

$$|\bar{Y}_{ij.} - \bar{Y}_{ij'.}| > LSD.$$

In the absence of the $A$ by $B$ interactions, pairwise comparisons of the $A$ factor level means or/and the $B$ factor level means are performed only if the ANOVA tests concluded that some $A$ or/and $B$ factor level means are not equal. If some $A$ factor level means are not equal, then for $i = 1, 2, .., a$ and $i' = 1, 2, .., a$, the hypotheses to be tested are:

$$H_o: \ \mu_{i..} = \mu_{i'..} \ versus \ H_a: \ \mu_{i..} \neq \mu_{i'..}, \ i \neq i'$$

where: $\mu_{i..} = \mu + \tau_i$ and $\mu_{i'..} = \mu + \tau_{i'}$. are the $i^{th}$ and $i'^{th}$ $A$ factor level means, respectively. The means $\mu_{i..}$ and $\mu_{i'..}$ are estimated by $\bar{Y}_{i..}$ and $\bar{Y}_{i'..}$, respectively. The standard error of $\bar{Y}_{i..} - \bar{Y}_{i'..}$ is $\sqrt{2MSE/br}$. Why? The least significant difference for the tests is:

$$LSD = t^{\alpha/2}_{(ab-1)(r-1)}\sqrt{2MSE/br}.$$

Thus, the means $\mu_{i..}$ and $\mu_{i'..}$ are significantly different if

$$|\bar{Y}_{i..} - \bar{Y}_{i'..}| > LSD.$$

Similarly, for $j = 1, 2, .., b$ and $j' = 1, 2, .., b$, the hypotheses for pairwise comparisons of the $B$ factor level means are:

$$H_o: \ \mu_{.j.} = \mu_{.j'.} \ versus \ H_a: \ \mu_{.j.} \neq \mu_{.j'.}, \ j \neq j'$$

where $\mu_{.j.} = \mu + \beta_j$ and $\mu_{.j'.} = \mu + \beta_{j'}$ are the $j^{th}$ and $j'^{th}$ $B$ factor level means, respectively. The least significant difference for the tests is:

$$LSD = t^{\alpha/2}_{(ab-1)(r-1)}\sqrt{2MSE/ar}.$$

The means $\mu_{.j.}$ and $\mu_{.j'.}$ are significantly different if

$$|\bar{Y}_{.j.} - \bar{Y}_{.j'.}| > LSD.$$

**Exercise 5.4** *Refer to examples 5.5 and 5.7. Perform pairwise comparisons of the relevant means. Which combination of the ph and the amount of calcium yields the largest increase in the diameter of orange trees?* ◇

## 5.4 Exercises

1. The factors that influence the breaking strength of synthetic fiber are being studied. Four machines and three operators are chosen at random and a factorial experiment is run using fiber from the same production batch. The results are as follows.

| | Machine | | | |
|---|---|---|---|---|
| Operator | 1 | 2 | 3 | 4 |
| 1 | 109 | 110 | 108 | 110 |
| | 110 | 115 | 109 | 108 |
| 2 | 110 | 110 | 111 | 114 |
| | 112 | 111 | 109 | 112 |
| 3 | 116 | 112 | 114 | 120 |
| | 114 | 115 | 119 | 117 |

   (a) Which of the mixed effects or the random effects model is appropriate for the experiment.

   (b) Analyst the data and draw conclusions.

   (c) Estimate the variance components.

2. An experiment is conducted to study the influence of operating temperature and three types of face-plate glass in the light output of a TV tube. The following data are collected.

| Glass Type | Temperature 100 | 125 | 150 |
|---|---|---|---|
| 1 | 580 | 1090 | 1392 |
| | 570 | 1085 | 1386 |
| 2 | 530 | 1035 | 1312 |
| | 579 | 1000 | 1299 |
| 3 | 546 | 1045 | 867 |
| | 575 | 1053 | 904 |

Assume that both factors are fixed. Analyst the data and draw conclusions.

3. A food-processing company has tested 6 different formulations of a new soft drink obtained from combining one of 3 levels of sweetness and one of 2 levels of energy content. Each of six subjects rated the 6 formulations when bottled in brown bottles and also when bottled in green bottles. The average ratings of the formulations follow.

| | | Bottle: Brown Energy Low | Brown Energy High | Green Energy Low | Green Energy High |
|---|---|---|---|---|---|
| Sweetness | 1 | 59.5 | 42.5 | 54.5 | 40.1 |
| | 2 | 66.8 | 49.6 | 64.7 | 50.1 |
| | 3 | 52.0 | 39.3 | 35.1 | 30.2 |

Suppose that the color of the bottle was regarded as a nuisance factor in experiment.

(a) Identify the design.

(b) Assume that the only interactions present in the ratings are the Sweetness by Energy interactions. Write an appropriate model for the data.

(c) Analyst the data and draw conclusions.

# Chapter 6
# Nested Designs

## 6.1 What this Chapter is About

In a factorial design (see chapter 5) each level of a factor appears with all the levels of the other factors. That is, the levels of a factor are the same (identical) for different levels of the other factors. When such is the case we say that the factors are crossed and the design is a factorial design. In other designs the levels of a factor are similar but not the same for different levels of the other factors. Such designs are called **nested** designs, and are the subject of discussion in this chapter.

By the end of this chapter you should be able to

* identify the appropriate model (fixed effects, random effects or mixed effects) for a given nested design;
* analyst a 2-stage nested design and draw appropriate conclusions.

## 6.2 Nested Designs

We explain a nested design by means of an example. Suppose that the effectiveness of two teaching methods 1 and 2 are being studied using the following two designs.

**Design I:** Three randomly chosen teachers use both methods and the responses of the students are recorded. The layout of the design is shown below.

| | Teacher | | |
|---|---|---|---|
| Method | 1 | 2 | 3 |
| 1 | $Y_{111}, Y_{112}, ..., Y_{11r}$ | $Y_{121}, Y_{122}, ..., Y_{12r}$ | $Y_{131}, Y_{132}, ..., Y_{13r}$ |
| 2 | $Y_{211}, Y_{212}, ..., Y_{21r}$ | $Y_{221}, Y_{222}, ..., Y_{22r}$ | $Y_{131}, Y_{132}, ..., Y_{13r}$ |

**Design II:** A group of the three randomly chosen teachers use method 1 and another group of three randomly chosen teachers use method 2. The responses of the students are recorded. The layout of the design is shown below.

| Method | | | | | |
|---|---|---|---|---|---|
| $A_1$ | | | $A_2$ | | |
| Teacher | | | Teacher | | |
| 1 | 2 | 3 | 1 | 2 | 3 |
| $Y_{111}$ | $Y_{121}$ | $Y_{131}$ | $Y_{211}$ | $Y_{221}$ | $Y_{231}$ |
| $Y_{112}$ | $Y_{122}$ | $Y_{132}$ | $Y_{212}$ | $Y_{222}$ | $Y_{232}$ |
| ... | ... | ... | ... | ... | ... |
| $Y_{11r}$ | $Y_{12r}$ | $Y_{13r}$ | $Y_{21r}$ | $Y_{22r}$ | $Y_{23r}$ |

Clearly design I is a factorial design. The two factors Method and Teacher are crossed, i.e., each teacher uses both methods of teaching. This is not the case in design II. Teacher 1 who uses method 1 is not the same teacher 1 who uses method 2. That is, teachers using each teaching method are different. It is for this reason that design II is called a **nested design**. The factor Teacher is **nested** under each teaching method.

They are many types of nested designs. Design II is called a **two-stage** nested design which means that there are $m$-**stage** ($m > 2$) nested designs. Indeed a design with factor $A$, factor $B$ nested under each level of $A$ and factor $C$ nested under each level of factor $B$ is a 3-stage nested design. In this chapter we shall only discuss the analysis of 2-stage nested design.

## 6.3 2-stage Nested Design

Suppose that the 2-stage design consists of factor $A$ with $a$ levels, factor $B$ (with $b$ levels) nested under each level of factor $A$ and $r$ observations taken at each combination of the levels of factors $A$ and $B$. The layout of the design is shown in table 6.1.

*Table 6.1 Layout of a 2-stage nested design*

| | $A$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | | 2 | | | | ... | $a$ | | |
| | $B$ | | | | $B$ | | | | ... | $B$ | | |
| | 1 | 2 | ... | $b$ | 1 | 2 | ... | $b$ | ... | 1 | 2 | ... | $b$ |
| | $Y_{111}$ | $Y_{121}$ | ... | $Y_{1b1}$ | $Y_{211}$ | $Y_{221}$ | ... | $Y_{2b1}$ | ... | $Y_{a11}$ | $Y_{a21}$ | ... | $Y_{ab1}$ |
| | $Y_{112}$ | $Y_{122}$ | ... | $Y_{1b2}$ | $Y_{212}$ | $Y_{222}$ | ... | $Y_{2b2}$ | ... | $Y_{a12}$ | $Y_{a22}$ | ... | $Y_{ab2}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| | $Y_{111}$ | $Y_{121}$ | ... | $Y_{1b1}$ | $Y_{211}$ | $Y_{221}$ | ... | $Y_{2b1}$ | ... | $Y_{a11}$ | $Y_{a21}$ | ... | $Y_{ab1}$ |
| Total | $Y_{11.}$ | $Y_{12.}$ | ... | $Y_{1b.}$ | $Y_{21.}$ | $Y_{22.}$ | ... | $Y_{2b.}$ | ... | $Y_{a1.}$ | $Y_{a2.}$ | ... | $Y_{ab.}$ |
| Mean | $\bar{Y}_{11.}$ | $\bar{Y}_{12.}$ | ... | $\bar{Y}_{1b.}$ | $\bar{Y}_{21.}$ | $\bar{Y}_{22.}$ | ... | $\bar{Y}_{2b.}$ | ... | $\bar{Y}_{a1.}$ | $\bar{Y}_{a2.}$ | ... | $\bar{Y}_{ab.}$ |

We present a practical example of a nested design. We shall use this example to illustrate the analysis of a 2-stage nested design.

**Example 6.1** *An experiment was conducted to determine the content uniformity of film-coated tablets for a cardiovascular drug used to lower blood pressure. A random sample of three batches was obtained from each of two blending sites; within each batch a random sample of five tablets was assayed to determine content uniformity. These data are shown in table 6.2 below. We have to investigate the effects of Site and Batch nested in Site on the content uniformity of the tablets.* ◇

*Table 6.2 Content uniformity of cardiovascular tablets from batches nested under blending sites*

| | Site | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | | 2 | | |
| | Batch | | | Batch | | |
| | 5.03 | 4.64 | 5.10 | 5.05 | 5.46 | 4.90 |
| | 5.10 | 4.73 | 5.15 | 4.96 | 5.15 | 4.95 |
| | 5.25 | 4.82 | 5.20 | 5.12 | 5.18 | 4.86 |
| | 4.98 | 4.95 | 5.08 | 5.12 | 5.18 | 4.86 |
| | 5.05 | 5.06 | 5.14 | 5.05 | 5.11 | 5.07 |
| Total | 25.41 | 24.20 | 25.67 | 25.30 | 26.08 | 23.74 |
| Mean | 5.082 | 4.840 | 5.134 | 5.060 | 5.216 | 5.348 |

## 6.4 Analysis of 2-stage Nested Design

### 6.4.1 Models

The ANOVA model for a 2-stage nested design (data in table 6.1) has the form:

$$\begin{aligned} Y_{ijk} &= \mu + \tau_i + \beta_j(i) + \epsilon_{ijk} \\ &= \mu_{ij} + \epsilon_{ijk} \end{aligned} \tag{6.1}$$

where:

* $\mu$ is the overall population mean;
* $\tau_i$ is the effect of the $i^{th}$ level of factor $A$;
* $\beta_j(i)$ is the effect of the $j^{th}$ level of factor $B$ nested under the $i^{th}$ level of factor $A$;
* $\mu_{ij}$ is the mean of the $j^{th}$ level of factor $B$ nested under the $i^{th}$ level of factor $A$;
* the $\epsilon_{ijk}$ are the random errors which are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$.

**Remark 6.1** *The residuals for checking the model error assumptions are given by:*

$$\hat{\epsilon}_{ijk} = Y_{ijk} - \bar{Y}_{ij.}.\diamond$$

**Model I**

Model I is the fixed effects model. The model assumes that the $\tau_i's$ and the $\beta_j(i)'s$ are fixed real constants satisfying the constraints:

$$\sum_{i=1}^{a} \tau_i = \sum_{j=1}^{b} \beta_j(i) = 0.$$

Under model I the means of $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$ nested under the $i^{th}$ level of factor $B$ are:

$$\mu_{i.} = \sum_{j=1} \mu_{ij} = \mu + \tau_i \text{ and } \mu_{ij},$$

respectively. The estimates of $\mu$, $\mu_{ij}$ and $\mu_{i.}$ are:

$$\bar{Y}_{...},\ \bar{Y}_{ij.} \text{ and } \bar{Y}_{i..},$$

respectively. We deduce that the estimates of $\tau_i$ and $\beta_j(i)$ must be:

$$\hat{\tau}_i = \bar{Y}_{i..} - \bar{Y}_{...} \text{ and } \hat{\beta}_j(i) = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.}$$

respectively. We wish to test the hypotheses which follow.

**Factor A Effects**

$H_o$ : All the $\tau_i's$ are equal to zero - The $A$ factor level means ($\mu_{i.}'s$) are equal
versus
$H_a$ : Some $\tau_i's$ are significantly different from zero - Some $A$ factor level means ($\mu_{i.}'s$) are not equal.

If we reject $H_o$ we wish to compare the $A$ factor level means.

**Factor B Nested in Factor A Effects**

$H_o$ : All the $\beta_j(i)'s$ are equal to zero - The means of the levels of factor $B$ nested under the levels of factor $A$ ($\mu_{ij}'s$) are equal
versus
$H_a$ : Some $\beta_j's$ are significantly different from zero - Some means of the levels of factor $B$ nested under the levels of factor $A$ ($\mu_{ij}'s$) are not equal.

If we reject $H_o$ we wish to compare the means of the levels of factor $B$ nested under the levels of factor $B$.

### Model II

Model II is random effects model. The model assumes that the

* $\tau_i's$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma_\tau^2$;
* $\beta_j(i)'s$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma_\beta^2$;
* $\epsilon_{ij}'s$, $\tau_i's$ and $\beta_j's$ are independent random variables.

We wish to test the hypotheses which follow.

**Factor $A$ Effects**

$H_o: \ \sigma_\tau^2 = 0$ - The effects of the levels of factor $A$ are identical
versus
$H_o: \ \sigma_\tau^2 > 0$ - The effects of the levels of factor $A$ are different.
If we reject $H_o$ we wish to estimate $\sigma_\tau^2$.

**Factor B Nested in Factor A Effects**

$H_o: \ \sigma_\beta^2 = 0$ - The effects of the levels of factor $B$ nested under the levels of factor $A$ are identical
$H_a: \ \sigma_\beta^2 > 0$ - The effects of the levels of factor $B$ nested under the levels of factor $A$ are different.
If we reject $H_o$ we wish to estimate $\sigma_\beta^2$.

### Model III

Model III is a mixed effects model. Model III usually assumes that the factor effects $A$ are fixed and that the factor $B$ effects nested in factor $A$ are random. Thus, the model assumes that the

* $\tau_i's$ are fixed real constants satisfying the constraint:

$$\sum_{i=1}^{a} \tau_i = 0;$$

* $\beta_j(i)'s$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma_\beta^2$;
* $\epsilon_{ij}'s$ and $\beta_j's$ are independent random variables.

The hypotheses about the factor $A$ and factor $B$ effects to be tested are the same as those under model I and model II, respectively. If the null hypotheses are rejected, we wish to compare the $A$ factor level means and also to estimate the variance component $\sigma_\beta^2$.

## 6.4.2 ANOVA

### Sums of Squares

The total sum of squares given by the usual formula:

$$SSTO = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} Y_{ijk}^2 - \frac{1}{abr}Y_{...}^2 \sim abr - 1 \ df.$$

The total sum of squares decompose into

* the factor $A$ sum of squares ($SSA$) which measure the variability due to factor $A$;

* the factor $B$ nested under factor $A$ sum of squares $SSB(A)$ which measure the variability due to factor $B$ nested under factor $A$;

* the error sum of the squares ($SSE$) which measure the variability due to the random errors.

We decompose $SSTO$ as follows:

$$\begin{aligned}
SSTO &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ijk} - \bar{Y}_{...})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left[(\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{ij.} - \bar{Y}_{i..}) + (\bar{Y}_{ijk} - \bar{Y}_{ij.})\right]^2 \\
&= \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{i..} - \bar{Y}_{...})^2}_{=SSA} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2}_{=SSB(A)} \\
&+ \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ijk} - \bar{Y}_{ij.})^2}_{=SSE}
\end{aligned}$$

The computing formulae for $SSA$, $SSB(A)$ and SSE are:

$$\begin{aligned}
SSA &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = \frac{1}{br}\sum_{i=1}^{a}\bar{Y}_{i..}^2 - \frac{1}{abr}Y_{...}^2 \sim a-1\ df \\
SSB(A) &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2 = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b}Y_{ij.}^2 - \frac{1}{br}Y_{i..}^2 \sim a(b-1)\ df \\
SSE &= SSTO - SSA - SSB(A) \sim ab(r-1)
\end{aligned}$$

**Exercise 6.1** *Refer to example 6.1. Verify that:*

$$SSTO = 0.76246 \sim 29\ df;\ SSA = 0.01825 \sim 1\ df;\ SSB(A) = 0.45401 \sim 4\ df. \diamond$$

### Mean Sums of Squares

As usual we compute the mean sum of squares for a factor by dividing the sum of squares for a factor by the corresponding degrees of freedom. The mean sums of squares for factor $A$ and factor $B$ nested in factor $A$ are:

$$MSA = \frac{SSA}{a-1} \text{ and } MSB(A) = \frac{SSB(A)}{a(b-1)},$$

respectively. Table 6.3 gives the expected values of these mean sums of squares under models I, II and III. We use these expectations to derive the test statistics for the hypotheses under the three models as well as for deriving formulae for the estimates of the variance components.

### Hypotheses Testing

We refer to table 6.3.

Model I

*Table 6.3 Expected mean sum of squares in a 2-stage nested model*

| $E[MS]$ | Model I | Model II | Model III |
|---|---|---|---|
| $E[MSA]$ | $\sigma^2 + \frac{br}{a-1}\sum_{i=1}^{a}\tau_i^2$ | $\sigma^2 + r\sigma_\beta^2 + \frac{br}{a-1}\sum_{i=1}^{a}\tau_i^2$ | $\sigma^2 + r\sigma_\beta^2 + br\sigma_\tau^2$ |
| $E[MSB(A)]$ | $\sigma^2 + \frac{r}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}^2$ | $\sigma^2 + r\sigma_\beta^2$ | $\sigma^2 + r\sigma_\beta^2$ |
| $E[MSE]$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ |

**Factor A Effects:** We note that if the null hypotheses is true, then $E[MSA] = E[MSE]$, otherwise, $E[MSA] > E[MSE$. This means that the test statistic about the factor $A$ effects under model I is:

$$F = \frac{MSA}{MSE} \sim F_{a-1,ab(r-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a-1,ab(r-1)}$.

**Factor B Nested in A Effects:** Similarly, the test statistic is:

$$F = \frac{MSB(A)}{MSE} \sim F_{a(b-1),ab(r-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a(b-1),ab(r-1)}$.

## Model II

**Factor A Effects:** If $H_o$ is true, then $E[MSA] = E[MSB(A)]$, and if $H_a$ is true, then $[MSA] > E[MSB(A)]$. Thus, the test statistic about the factor $A$ effects must be:

$$F = \frac{MSA}{MSB(A)} \sim F_{a-1,a(b-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a-1,a(b-1)}$.

**Factor B Nested in A Effects:** The test statistic for the hypotheses about factor $B$ nested in factor $A$ effects is:

$$F = \frac{MSB(A)}{MSE} \sim F_{a(b-1),ab(r-1)}.$$

Why? We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a(b-1),ab(r-1)}$.

**Exercise 6.2** *Use table 6.3 to argue that under model III the test statistics for testing the hypotheses about the factor A effects and the factor B nested in factor A effects are:*

$$F = \frac{MSA}{MSB(A)} \sim F_{a-1,a(b-1)} \text{ and } F = \frac{MSB(A)}{MSE} \sim F_{a(b-1),ab(r-1)},$$

*respectively.* ◇

**Exercise 6.3** *Refer to example 6.1. Assume that the Site effects are fixed. Test the relevant hypotheses about the Site effects and the Batch nested in Site effects at the 0.05 level of significance.* ◇

*Table 6.3 Expected mean sum of squares in a 2-stage nested model*

| $E[MS]$ | Model I | Model II | Model III |
|---|---|---|---|
| $E[MSA]$ | $\sigma^2 + \frac{br}{a-1}\sum_{i=1}^{a}\tau_i^2$ | $\sigma^2 + r\sigma_\beta^2 + \frac{br}{a-1}\sum_{i=1}^{a}\tau_i^2$ | $\sigma^2 + r\sigma_\beta^2 + br\sigma_\tau^2$ |
| $E[MSB(A)]$ | $\sigma^2 + \frac{r}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}^2$ | $\sigma^2 + r\sigma_\beta^2$ | $\sigma^2 + r\sigma_\beta^2$ |
| $E[MSE]$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ |

**Factor A Effects:** We note that if the null hypotheses is true, then $E[MSA] = E[MSE]$, otherwise, $E[MSA] > E[MSE$. This means that the test statistic about the factor $A$ effects under model I is:

$$F = \frac{MSA}{MSE} \sim F_{a-1,ab(r-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a-1,ab(r-1)}$.

**Factor B Nested in A Effects:** Similarly, the test statistic is:

$$F = \frac{MSB(A)}{MSE} \sim F_{a(b-1),ab(r-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a(b-1),ab(r-1)}$.

## Model II

**Factor A Effects:** If $H_o$ is true, then $E[MSA] = E[MSB(A)]$, and if $H_a$ is true, then $[MSA] > E[MSB(A)]$. Thus, the test statistic about the factor $A$ effects must be:

$$F = \frac{MSA}{MSB(A)} \sim F_{a-1,a(b-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a-1,a(b-1)}$.

**Factor B Nested in A Effects:** The test statistic for the hypotheses about factor $B$ nested in factor $A$ effects is:

$$F = \frac{MSB(A)}{MSE} \sim F_{a(b-1),ab(r-1)}.$$

Why? We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a(b-1),ab(r-1)}$.

**Exercise 6.2** *Use table 6.3 to argue that under model III the test statistics for testing the hypotheses about the factor A effects and the factor B nested in factor A effects are:*

$$F = \frac{MSA}{MSB(A)} \sim F_{a-1,a(b-1)} \text{ and } F = \frac{MSB(A)}{MSE} \sim F_{a(b-1),ab(r-1)},$$

*respectively.* ◇

**Exercise 6.3** *Refer to example 6.1. Assume that the Site effects are fixed. Test the relevant hypotheses about the Site effects and the Batch nested in Site effects at the 0.05 level of significance.* ◇

### 6.4.3 Estimation

#### Pairwise Comparisons

Under models I and III we perform pairwise comparisons of the factor $A$ level means if the ANOVA test about the factor $A$ effects rejects $H_o$. For $i = 1, 2, .., a$ and $i' = 1, 2, .., a$, the hypotheses to be tested are:

$$H_o : \mu_{i.} = \mu_{i'.} \; versus \; H_a : \mu_{i.} \neq \mu_{i'.}, \; i \neq i'.$$

The means $\mu_{i.}$ and $\mu_{i'.}$ are estimated by $\bar{Y}_{i..}$ and $\bar{Y}_{i'..}$, respectively.

Under model I the standard error of $\bar{Y}_{i..} - \bar{Y}_{i'..}$ is $\sqrt{2MSE/br}$. Why? The least significant difference for the pairwise comparisons is:

$$LSD = t^{\alpha/2}_{ab(r-1)} \sqrt{2MSE)/br}.$$

Thus, means $\mu_{i.}$ and $\mu_{i'.}$ are significantly different if

$$|\bar{Y}_{i..} - \bar{Y}_{i'..}| > LSD.$$

Under model II the standard error of $\bar{Y}_{i..} - \bar{Y}_{i'..}$ is $\sqrt{2MSB(A))/br}$. Why? The least significant difference for the pairwise comparisons is:

$$LSD = t^{\alpha/2}_{a(b-1)} \sqrt{2MSB(A)/br}.$$

Thus, means $\mu_{i.}$ and $\mu_{i'.}$ are significantly different if

$$|\bar{Y}_{i..} - \bar{Y}_{i'..}| > LSD.$$

**Exercise 6.4** *Under model I what is LSD for pairwise comparisons of the factor B nested in factor A means?* ◇

#### Variance Component Estimation

Under model II we estimate $\sigma^2_\tau$ if the ANOVA test concludes that $\sigma^2_\tau > 0$. We deduce from table 6.3 that:

$$\sigma^2_\tau = \frac{1}{br}(E[MSA] - E[MSB(A)])$$

which suggests that we estimate $\sigma^2_\tau$ with:

$$\hat{\sigma}^2_\tau = \max\{0, \frac{1}{br}(MSA - MSB(A))\}.$$

Under models II and III we estimate $\sigma^2_\beta$ if the ANOVA tests conclude that $\sigma^2_\beta > 0$. Again we deduce from table 6.3 that under both model:

$$\sigma^3_\beta = \frac{1}{r}(E[MSB(A)] - E[MSE])$$

which suggests that we estimate $\sigma^2_\beta$ with:

$$\hat{\sigma}^2_\beta = \max\{0, \frac{1}{r}(MSB(A) - MSE)\}.$$

**Exercise 6.5** *Refer to exercise 6.3. Assume that the Batch nested in Site effects are random. Is it appropriate to estimate $\sigma^2_{batch}$? If so estimate it.* ◇

## 6.5 Exercises

1. An ammunition manufacturer is studying the burning rate of gun powder produced by three processes. Four batches of powder are randomly selected from the output of each process and three independent measurements of burning rate are made on each batch. The results are shown in the table below.

   (a) Identify the design and the model (I, II or II) for the experiment.

   (b) Analyst the data completely and draw conclusions.

| Process | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | 2 | | | | 3 | | | |
| Batch | | | | Batch | | | | Batch | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 25 | 19 | 15 | 15 | 19 | 23 | 18 | 35 | 14 | 35 | 38 | 25 |
| 30 | 28 | 17 | 16 | 17 | 24 | 21 | 27 | 15 | 21 | 54 | 29 |
| 26 | 20 | 14 | 13 | 14 | 21 | 17 | 25 | 20 | 24 | 50 | 33 |

2. A company is studying the number of bad spots on the diskettes produced by four machines. Each machine is run by three different operators who are chosen at random. Two diskettes from each operator are randomly selected and examined for the number of bad spots. The data are shown in the table below.

   (a) If the four machines were chosen at random, analyst the data completely and draw conclusions.

   (a) How would your findings in part (*a*) change if the company was interested in the effects of the four particular machines.

| Machine | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | 2 | | | 3 | | | 4 | | |
| Operator | | | Operator | | | Operator | | | Operator | | |
| 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 79 | 94 | 46 | 92 | 85 | 76 | 88 | 53 | 46 | 36 | 40 | 62 |
| 62 | 74 | 57 | 99 | 79 | 68 | 75 | 56 | 57 | 53 | 56 | 47 |

# Chapter 7
# Split-Plot Designs

## 7.1 What this Chapter is About

We consider a two-factor experiment whereby we wish to study the effects of factor $A$ with $a$ levels and factor $B$ with $b$ levels. Available for experimentation are $a$ experimental units each of which divides into $b$ subunits. For convenience, we will first randomly assign the $a$ experimental units to the $a$ levels of factor $A$. Then we will randomly assign the $b$ subunits of each experimental unit to the $b$ levels of factor $B$. For example, in agriculture factor $A$ may be tillage method and factor $B$ may be the amount of fertilizer applied. In this case the experimental units are the plots on which the tillage methods are used, and within each plot we have subplots upon which the amounts of fertilizer are applied. This design is not a factorial design but a **split-plot design**. In a factorial design the experimental units are randomly assigned to the factor level combinations, which is not the case in a split-plot design. In this chapter we discuss the analysis of the split-plot design.

By the end of this chapter you should be able to

* distinguish between a split-plot design and a factorial experiment;

* analyst a split plot design and draw appropriate conclusions.

## 7.2 Split-Plot Designs

Let us consider a basic two-factor split-plot design. In this design the $a$ "plots" (called **wholeplots**) are randomly assigned to the $a$ levels of factor $A$. Thus, factor $A$ is sometimes called a **wholeplot factor**. The wholeplot is subdivided into $b$ "subplots". Then the $b$ subplots within each wholeplot are randomly assigned to the $b$ levels of factor $B$. Factor $B$ is called the **subplot factor**. In this design the levels of factor $A$ are assigned to more heterogeneous experimental units than the levels of factor $B$ because we expect the wholeplot to wholeplot variation to be larger than the within wholeplot subplot to subplot variation. It is for this reason that the inferences about the factor $A$ effects are usually more precise than the inferences about the factor $B$ effects. (The experimental error associated with wholeplots is usually larger than that associated with the subplots within wholeplots.)

We can replicate the basic split-plot design $r$ times if we have $r$ blocks each with $a$ uniform (identical) wholeplots each with $b$ subplots, or, if we have $ra$ uniform (identical) wholeplots each with $b$ subplots. The layout of a replicated split-plot design is shown in table 7.1.

*Table 7.1 Layout of a replicated basic split-plot design*

| | | Replication (Block) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | | | | 2 | | | |
| | | A | | | | A | | | |
| | | 1 | 2 | ... | $a$ | 1 | 2 | ... | $a$ |
| $B$ | | wholeplot | wholeplot | ... | wholeplot | wholeplot | wholeplot | ... | wholeplot |
| 1 | subplot | $Y_{111}$ | $Y_{211}$ | ... | $Y_{a11}$ | $Y_{112}$ | $Y_{212}$ | ... | $Y_{a12}$ |
| 2 | subplot | $Y_{121}$ | $Y_{221}$ | ... | $Y_{a21}$ | $Y_{122}$ | $Y_{222}$ | ... | $Y_{a22}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| $b$ | subplot | $Y_{1b1}$ | $Y_{2b1}$ | ... | $Y_{ab1}$ | $Y_{1b2}$ | $Y_{2b2}$ | ... | $Y_{ab2}$ |

We present a practical example of a split-plot design. We shall use the example to illustrate the analysis of the design.

**Example 7.1** *Fertilizers (factor A) were applied at random to the wholeplot units within each farm. Soybean varieties (factor B) were randomly allocated to the subplots within each wholeplot. The data on yield (per subplot unit) are shown in table 7.2 below.* ◇

*Table 7.2 Yield of soybeans from a split-plot design*

| | Replication (Farm) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| | Fertilizer | | Fertilizer | Fertilizer | | |
| Variety | 1 | 2 | 1 | 2 | 1 | 2 |
| 1 | 10.6 | 10.9 | 10.8 | 11.6 | 8.1 | 8.2 |
| 2 | 11.4 | 11.7 | 11.5 | 11.9 | 8.7 | 9.3 |
| 3 | 11.8 | 12.4 | 12.1 | 12.6 | 9.5 | 9.8 |

## 7.3 Analysis of Split-Plot Design

### 7.3.1 Model

The ANOVA model for a two-factor split-plot design (data in table 7.1) has the form:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \tau_k + (\alpha\tau)_{ik} + \epsilon_{ijk} \tag{7.1}$$

where:

* $Y_{ijk}$ is the response to the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$ in the $k^{th}$ block (replication);
* $\mu$ is the overall population mean;
* $\alpha_i$ is the effect of the $i^{th}$ level of factor $A$;
* $\beta_j$ is the effect of the $j^{th}$ level of factor $B$;
* $(\alpha\beta)_{ij}$ is the interaction effect of the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$;
* $\tau_k$ is the effect of the $k^{th}$ block (replication);
* $(\alpha\tau)_{ik}$ is the interaction effect of the $i^{th}$ level of factor $A$ and the $k^{th}$ block (replication);
* the $\epsilon_{ijk}$ are the random errors (associated with the subplots) which are assumed to be independent and normally distributed with mean 0 and variance $\sigma^2$;
* $i = 1, 2, ..., a;\ j = 1, 2, ..., b;\ k = 1, 2, ..., r.$

We shall assume that the effects of factors $A$ and $B$ are fixed, and that the block effects are random. Thus, the $\alpha_i's$, $\beta_j's$ and $(\alpha\beta)_{ij}'s$ are fixed real constants satisfying the constraints:

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0.$$

Furthermore, the

* $\tau_k's$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma_\tau^2$;
* $(\alpha\tau)_{ik}'s$ are random variables which are independent and normally distributed with mean 0 and variance $\sigma_{\alpha\tau}^2$;
* $\epsilon_{ij}'s$, $\tau_k's$ and $(\alpha\beta)_{ik}'s$ are independent random variables.

We wish to test the hypotheses that follow.

**Factor A by Factor B Interaction Effects**

$H_o$ : All the $(\alpha\beta)_{ij}$ are equal to zero versus $H_a$ : Some $(\alpha\beta)_{ij}$ are significantly different from zero.

**Factor A Effects**

$H_o$ : All the $\alpha_i's$ are equal to zero versus $H_a$ : Some $\alpha_i's$ are significantly different from zero.

**Factor B Effects**

$H_o$ : All the $\beta_j's$ are equal to zero versus $H_a$ : Some $\beta_j's$ are significantly different from zero.

### 7.3.2 ANOVA

**Sums of Squares**

The total sum of squares are given by the usual formula:

$$SSTO = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(Y_{ijk} - \bar{Y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}Y_{ijk}^2 - \frac{1}{abr}Y_{...}^2 \sim abr - 1\ df.$$

The total sum of squares decompose into

* the factor $A$ sum of squares ($SSA$);
* the factor $B$ sum of squares $SSB$;
* the $A$ by $B$ interaction sum of squares ($SSAB$);
* the block sum of squares ($SSBlk$);
* the block sum of squares ($SSBlk \times A$) which measure the **wholeplot to wholeplot variation.**
* the error sum of the squares ($SSE$) which measure the **within wholeplot subplot to subplot variation** and / or variability due to the random errors.

A two-way ANOVA of the split-plot design ignoring the blocks (replications) obtains the following computing formulae for $SST$, $SSE_b$, $SSA$, $SSB$ (see section 5.3.1 in chapter 5).

$$\begin{aligned}
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{ij.} - \bar{Y}_{...})^2 = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b}Y_{ij.}^2 - \frac{1}{abr}Y_{...}^2 \sim ab - 1\ df \\
SSE_b &= SSTO - SST \sim ab(r-1)\ df \\
SSA &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{i..} - \bar{Y}_{...})^2 = \frac{1}{br}\sum_{i=1}^{a}Y_{i..}^2 - \frac{1}{abr}Y_{...}^2 \sim a - 1\ df \\
SSB &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 = \frac{1}{ar}\sum_{j=1}^{b}Y_{.j.}^2 - \frac{1}{abr}Y_{...}^2 \sim b - 1\ df \\
SSAB &= SST - SSA - SSB \sim (a-1)(b-1)\ df
\end{aligned}$$

When we ignore the blocks as we did above, then $SSE_b$ measures the variability due to the block effects, the block by $A$ interaction effects and the random error effects (errors associated with the subplots). That is,

$$SSE_b = SSBlk + SSBlk \times A + SSE.$$

Next we perform a two-way ANOVA of the split-plot design ignoring factor $B$ (subplot factor). This obtains the computing formulae for "$SST$", $SSE_a$, $SSBlk$ and $SSBlk \times A$ which follow. Here, "$SST$"

denotes the sum of squares due to the combination of the levels of factor $A$ and the blocks, and $SSE_a$ measures the variability due to factor $B$ effects, the $A$ by $B$ interaction effects and the random error effects (subplot errors). That is,

$$SSE_a = SSB + SSAB + SSE.$$

$$\begin{aligned}
"SST" &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{i.k} - \bar{Y}_{...})^2 = \frac{1}{b}\sum_{i=1}^{a}\sum_{k=1}^{r}Y_{i.k}^2 - \frac{1}{abr}Y_{...}^2 \sim ar-1\ df \\
SSE_a &= SSTO - "SST" \sim ar(b-1)\ df \\
SSBlk &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{Y}_{..k} - \bar{Y}_{...})^2 = \frac{1}{ab}\sum_{k=1}^{r}Y_{..k}^2 - \frac{1}{abr}Y_{...}^2 \sim r-1\ df \\
SSBlk \times A &= "SST" - SSA - SSBlk \sim (a-1)(r-1)\ df
\end{aligned}$$

Finally, we calculate $SSE$ using the formulae:

$$SSE = SSE_b - SSBlk - SSBlk \times A \sim a(b-1)(r-1)\ df$$

or

$$SSE = SSE_a - SSB - SSAB\ a(b-1)(r-1) \sim df$$

We summaries the results in the ANOVA table below (table 7.3).

*Table 7.3 ANOVA table for a two-factor split-plot design*

| *Source* | *SS* | *df* | *E[MS]* |
|---|---|---|---|
| Blocks(Reps) | $SSBlk$ | $r-1$ | $\sigma^2 + ab\sigma_\tau^2$ |
| $A$ | $SSA$ | $a-1$ | $\sigma^2 + b\sigma_{\alpha\beta}^2 + \frac{br}{a-1}\sum_{i=1}^{a}\alpha_i^2$ |
| Block $\times A$ | $SSBlk \times A$ | $(a-1)(r-1)$ | $\sigma^2 + b\sigma_{\alpha\beta}^2$ |
| $B$ | $SSB$ | $b-1$ | $\sigma^2 + \frac{ar}{b-1}\sum_{j=1}^{b}\beta_j^2$ |
| $A \times B$ | $SSAB$ | $(a-1)(b-1)$ | $\sigma^2 + \frac{r}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2$ |
| Error | $SSE$ | $a(b-1)(r-1)$ | $\sigma^2$ |

**Remark 7.1** *$MSBlk \times A$ is a measure of the wholeplot error (experimental error associated with the whole plots), and $MSE$ is a measure of the subplot error (experimental error associated with the subplots).* ◇

### Hypotheses Testing

We refer to table 7.3.

#### Factor A Effects

We note that if the null hypothesis is true, then $E[MSA] = E[MSBlk \times A]$, otherwise, $E[MSA] > E[MSBlk \times A]$. This means that the test statistic for the factor $A$ effects is:

$$F = \frac{MSA}{MSBlk \times A} \sim F_{a-1,(a-1)(r-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{a-1,(a-1)(r-1)}$.

### Factor A by Factor B Interactions Effects

Again, we note that if the null hypotheses is true, then $E[MSAB] = E[MSE]$, otherwise, $E[MSAB] > E[MSE]$. Therefore, the test statistic for the $A$ by $B$ interaction effects is:

$$F = \frac{MSAB}{MSE} \sim F_{(a-1)(b-1),a(b-1)(r-1)}$$

if $H_o$ is true and the model error assumptions hold. We reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{(a-1)(b-1),a(b-1)(r-1)}$.

### Factor B Effects

**Exercise 7.1** *Refer to table 7.3. Argue that the test statistic for the factor B effects is:*

$$F = \frac{MSB}{MSE} \sim F_{(b-1),a(b-1)(r-1)}$$

*so that we reject $H_o$ at the $\alpha$ level of significance if $F > F^{\alpha}_{(b-1),a(b-1)(r-1)}$.*◇

**Exercise 7.2** *Refer to table 7.2.*

(a) *Use the data to reproduce the ANOVA table below (table 7.4).*

(b) *Test the significance of the main and the interaction effects at $\alpha = 0.05$.*◇

*Table 7.4 ANOVA table for soybean data*

| *Source* | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| Farm | 29.934 | 2 | 14.967 | ... |
| Fertilizer | 0.841 | 1 | 0.841 | 28.18 |
| Farm × Fertilizer | 0.060 | 2 | 0.030 | ... |
| Variety | 4.955 | 2 | 2.477 | 33.12 |
| Fertilizer × Variety | 0.021 | 2 | 0.010 | 0.14 |
| Error | 0.598 | 8 | 0.075 | ... |
| Total | 35.325 | 17 | | |

## 7.4 Exercise

A paper manufacturer wishes to study the effects of three different pulp preparation methods and four different cooking temperatures (for the pulp) on the tensile strength of the paper. He conducts the experiment over three randomly chosen days. On any day a batch of pulp is produced by one of three methods under study. This batch is divided into four samples, and each sample is cooked at one of four temperatures. Then a second batch is made up using another of the three methods. This second batch is also divided into four samples that are tested at the four temperatures. The process is then repeated using a batch of pulp produced by the third method. The data are shown in the table below.

1. Argue that this is a split-plot experiment.

2. Perform an analysis of the data and draw conclusions.

| | Day | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | | | 2 | | | 3 | | |
| | Method | | | Method | | | Method | | |
| Temperature | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 200 | 30 | 34 | 29 | 28 | 31 | 31 | 31 | 35 | 32 |
| 225 | 35 | 41 | 26 | 32 | 36 | 30 | 37 | 40 | 34 |
| 250 | 37 | 38 | 33 | 40 | 42 | 32 | 41 | 39 | 39 |
| 275 | 36 | 42 | 36 | 41 | 40 | 40 | 40 | 44 | 45 |