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						
Treatment	1	2	.	.	r	Mean	Variance
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 σ^2 (unknown). The σ^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., σ^2 can not be estimated. However, when $r > 1$, then the estimate of σ^2 (S_p^2) is well defined. This confirms that replication of the basic experiment allows the estimation the experimental error σ^2. If r is large then S_p^2 will be very close to σ^2.

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

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

The variance $\sigma_{\bar{1}\bar{2}}^2$ is estimated by $\hat{\sigma}_{\bar{1}\bar{2}}^2 = \frac{2}{r}S_p^2$. The $\sigma_{\bar{1}\bar{2}}^2$ 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 $\sigma_{\bar{1}\bar{2}}^2$ is small. What happens to $\sigma_{\bar{1}\bar{2}}^2$ 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 $\sigma_{\bar{1}\bar{2}}^2 \to 0$ 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 σ^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?* $\diamond$

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 σ^2 is known, then the appropriate test is the **z-test**. The test statistic for the $z - test$ is:

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

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 $\sigma_{\bar{1}\bar{2}} \to 0$ 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.

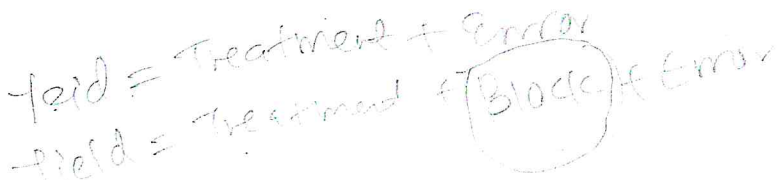
A measure of the sensitivity of a test for comparing treatment means is the **power** of the test to detect differences between treatment means. The power of a test is the probability of rejecting H_o (treatment means are not different) when H_a is true (treatment means are different). In other words, 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 α **level of significance**. Let δ 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.



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 Afrikandèr 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 (σ^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 (σ^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
```

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 (σ^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
```