1 INTRODUCTION

- In an experiment, the **experimental units** are the objects subjected to specific experimental conditions. If the experimental units are humans, they are called **subjects**. The specific experimental conditions are called **treatments**. An **experimental run** is a set of experimental conditions to be used for collecting an **experimental observation**.

- In any experiment, it is important to distinguish which variables are **independent** and which are **dependent**. The independent variables are those that potentially influence or affect the values of the experimental responses or dependent variables.

- Independent variables are the explanatory variables, i.e., variables with levels that are varied by the experimenter based on the experimental design. The independent variables are called the **factors** of the experiment. In regression analysis, we refer to independent variables as **predictor variables**.

1.1 Five Principles Used In Designing Experiments

<u>Principle 1</u>: **Random selection** is the process of randomly selecting the experimental units that will be included in the experiment.

<u>Principle 2</u>: **Random assignment** is the process of randomly assigning experimental units to treatments to create treatment groups that are similar (except for chance variation) before treatments are applied. This is also known as the **randomization** of treatments.

<u>Principle 3</u>: When observations have to be collected sequentially, the experimental run order should be randomized. **Run order randomization** is the process of randomly assigning experimental units to the order of data collection.

<u>Principle 4</u>: **Control** the effects of factors that are not of interest.

<u>Principle 5</u>: **Replication** is the process of repeating treatments on sets of experimental units.

- When an observed difference in treatment effects is too large to reasonably have occurred purely by chance, we say that the difference is **statistically significant**.

- If significant differences among treatments are found after running a properly designed experiment, we often conclude that the differences are *caused* by the treatments. That is, there is a **cause-and-effect** relationship between the response and the treatments.

- In an **observational study**, treatments are not actively imposed on the cases. Data are collected by observing cases under natural (not controlled) conditions. Conclusions regarding causality are tentative because other variables unaccounted for in the study (*lurking variables)* may be confounded with the explanatory variable.

- Benefits of designed experiments over observational studies:

  (1) Allows for the study of effects of factors that are of particular interest.

  (2) Allows for the control of factors not of interest.

  (3) Allows for the study of combined effects of several factors simultaneously.

1.2 Experimental Objectives

- The first and most important step in an experimental strategy is to clearly state the objectives or goals of the experiment.

- Many researchers discover only after the experiment is run that the data are insufficient to meet their objectives. This is why it is important to stress the importance of the experimental design before running an experiment.

- The **objective** of the experiment can be thought of as a very precise answer to the question **"What do you want to know when the experiment is complete?"** Stating this question will help the researcher decide which experimental design to use.

Identifying the Variables:

- In any experiment, it is important to distinguish which variables are **independent** and which are **dependent**. The independent variables are those deliberately controlled (i.e. set at predetermined levels at each run of the experiment).

- Various measurements may be made from each run of the experiment. The measured variables are the dependent variables with values that are at least partially affected by the levels of the independent variables. In an analysis, often we would like to determine which factors significantly affect the dependent variables.

Recognizing Experimental Error:

- Experimental error (or variability) is a fact of life. By acknowledging this, an experimental design must be developed that will provide usable results in the presence of variability. It becomes necessary to distinguish between two types of experimental error: **bias** and **random** error.

- **Bias** occurs when the responses tend to follow a consistent pattern across the experimental runs. When ignored or overlooked, sources of bias can be attributed to differences in:

  - Batches of raw material, machine operators (manufacturing).
  - Plots of land (agriculture).
  - Litters of animals (biological assay).
  - Gender (psychology).

  Similarly, bias may be due to a regular pattern or cycle depending upon:

  - Hour within a day, day within a week, etc. (manufacturing).
  - Season of the year (agriculture, marketing).
  - Equipment deterioration (manufacturing).
  - Subject age (psychology, medicine).

- **Random errors** occur when the responses change from one run to the next without a consistent pattern. A random error cannot be assigned to a single cause, and it cannot be predicted.

- Random error is the inherent "noise" in a measured response. Random errors can mask the effects of factors on the response ("signals"), that is, when a given response is "noisy", it is difficult to reliably quantify the relationship between the response and the treatments.

- A principle statistical tool for dealing with random error through design is **replication** which involves repeating experimental runs. The benefit of replication is that the average result for the replicated runs is generally closer to the true value than a single observation.

- Good experimental strategies require limited replication because the efficiency of replication decays for a large number of replicate runs.