# Review of Experimental and Statistical Objectives.

Dr. Mutua Kilai

Department of Pure and Applied Sciences

Jan-April 2024

Innovative technology for a dynamic world
Kirinyaga University

# Introduction

- **Experiment** is a test or series of runs in which purposeful changes are made to the input variables of a process or system so that we may observe and identify reasons for changes that may be observed in the output response.
- The models of such phenomena that follow directly from the physical mechanism are usually called *mechanistic models.*
- A simple example is the familiar equation for current flow in an electrical circuit, Ohm's law, $E = IR$
- Well-designed experiments can often lead to a model of system performance; such experimentally determined models are called *empirical models.*

- In general experiments are used to study the performance of processes and systems
- The process or system can be represented by the model shown in Figure 1.

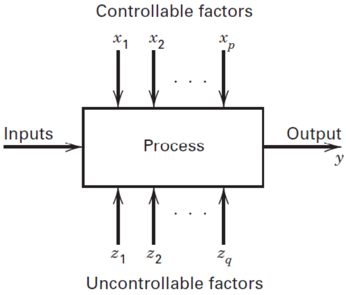


Figure 1: General Model of a Process

# Objectives of an Experiment

i. Determining which variables are most influential on the response $y$

ii. Determining where to set the influential $x's$ so that $y$ is almost always near the desired value

iii. Determining where to set the influential $x's$ so that variability in $y$ in small.

iv. Determining where to set the influential $x's$ so that the effects of the uncontrollable variables $z_1, z_2, ..., z_q$ are minimized

The general approach to planning and conducting the experiment is called the **strategy of experimentation.**

# Typical Applications of Experimental Designs

- The application of experimental design techniques early in the process development can result in:
    i. Improved process yields
    ii. Reduced variability and closer conformance to nominal or target requirements
    iii. Reduced development time
    iv. Reduced overall costs.

# Basic Principles

- Statistical design of experiments refers to the process of planning the experiment so that appropriate data will be collected and analyzed by statistical methods, resulting in valid and objective conclusions.
- The three basic principles of experimental design are:
    i. Randomization
    ii. Replication
    iii. Blocking
    iv. Factorial principle

# Randomization

- Randomization is the cornerstone underlying the use of statistical methods in experimental design.
- *randomization means that both the allocation of the experimental material and the order in which the individual runs of the experiment are to be performed are randomly determined.*
- Statistical methods require that the observations (or errors) be independently distributed random variables.
- Randomization usually makes this assumption valid.

# Replication

- By replication we mean an independent repeat run of each factor combination.
- Replication has two important properties.
  i. First, it allows the experimenter to obtain an estimate of the experimental error.
  ii. Replication permits the experimenter to obtain a more precise estimate of the sample mean.

# Blocking

- Blocking is the arranging of experimental units in groups (blocks) that are similar to one another.
- Blocking is a design technique used to improve the precision with which comparisons among the factors of interest are made.
- Often blocking is used to reduce or eliminate the variability transmitted from nuisance factors—that is, factors that may influence the experimental response but in which we are not directly interested.

# Guidelines for Designing Experiments

- Recognition of and statement of the problem
- Selection of the response variable
- Choice of factors, levels and ranges
- Choice of experimental design
- Performing the experiment
- Statistical analysis of the data
- Conclusions and recommendations

# General Incomplete Block Design

- One of the basic principles in experimental design is that of reduction of experimental error.
- This can be achieved through the device of blocking leading to designs such as:
    i. Randomized Complete Block Design
    ii. Latin square designs
- The problem we shall be discussing is the comparison of a number of treatments using blocks the size of which is less than the number of treatments.
- Designs of this type are called incomplete block designs

# Analysis of Incomplete Block Designs

- The analysis of incomplete block designs is different from the analysis of complete block designs in that comparisons among treatment effects and comparisons among block effects are no longer orthogonal to each other.
- This nonorthogonality leads to an analysis analogous to that of the two-way classification with unequal subclass numbers.
- However, this is only partly true and applies only to the analysis that has come to be known as the intrablock analysis.
- The name of the analysis is derived from the fact that contrasts in the treatment effects are estimated as linear combinations of comparisons of observations in the same block.

# The Intrablock Analysis

Notation and Model

- Suppose we have $t$ treatments replicated $r_1, r_2, ..., r_t$ times respectively and $b$ blocks with $k_1, k_2, ..., k_b$ units respectively then:

$$\sum_{i=1}^{t} r_i = \sum_{j=1}^{b} k_j = n$$

  where $n$ is the total number of observations.

- An appropriate linear model for the observations from an incomplete block design is:

$$y_{ijl} = \mu + \tau_i + \beta_j + e_{ijl} \tag{1}$$

# Cont'd

- $i = 1, 2..., t; j = 1, 2, ...b; l = 0, 1, ..., n_{ij}$ where $\tau_i$ is the effect of the $i$th treatment, $\beta_j$ the effect of the $j^{th}$ block and $e_{ijl}$ the error associated with the observation $y_{ijl}$
- The $e_{ijl}$ contain both experimental and observational error.

$$e_{ijl} = \epsilon_{ijl} + \eta_{ijl}$$

- Model 1 can be written in matrix notation as

$$y = \mu J + X_\tau \tau + X_\beta \beta + e \quad (2)$$

where J is a column vector consisting of n unity elements, $X_\beta$ is the observation block incidence matrix

# Normal and Reduced Normal Equations

- The normal equations for $\mu, \tau_i$ and $\beta_j$ are then given by:

$$n\hat{\mu} + \sum_{i=1}^{t} r_i \hat{\tau}_i + \sum_{j=1}^{b} k_j \hat{\beta}_j = G \tag{3}$$

$$r_i \hat{\mu} + r_i \hat{\tau} i + \sum_{j=1}^{b} n_{ij} \hat{\beta}_j = T_i \ \ (i = 1, 2, ..., t) \tag{4}$$

$$k_j \hat{\mu} + r_i \hat{\tau}_i + \sum_{j=1}^{t} n_{ij} \hat{\tau}_i + k_j \hat{\beta}_j = B_j \ \ (j = 1, 2, ..., b) \tag{5}$$

# Cont'd

- $$T_i = \sum_{jl} y_{ijl} = ith \ \ treatment \ \ total$$

- $$B_j = \sum_{il} y_{ijl} = jth \ \ block \ \ total$$

- $$G = \sum_{i} T_i = \sum_{j} B_j = overall \ \ total$$

# Analyses of Variance

- The treatment-after-block ANOVA or T|B-ANOVA as it is associated with the ordered model

$$y = \mu J + X_\beta \beta + X_\tau \tau + e \quad (6)$$

| Source | df | SS | E(MS) |
|---|---|---|---|
| $X_\beta \mid J$ | b-1 | $\sum_{j=1}^{b} \frac{B_j^2}{k_j} - \frac{G^2}{n}$ | |
| $X_\tau \mid J, X_\beta$ | t-1 | $\sum_{i=1}^{t} \hat{\tau}_i Q_i$ | $\sigma_e^2 + \frac{\tau' C_\tau}{t-1}$ |
| $I \mid J, X_\beta, X_\tau$ | n-b-t+1 | Difference | $\sigma_e^2$ |
| Total | n-1 | $\sum_{ijl} y_{ijl}^2 - \frac{G^2}{n}$ | |

# Cont'd

- The ANOVA is associated with the ordered model

$$y = \mu J + X_\tau \tau + X_\beta \beta + e \qquad (7)$$

and hence shall be referred to as the block-after-treatment ANOVA or B|T ANOVA.

| Source | df | SS |
|---|---|---|
| $X_\tau \mid J$ | t-1 | $\sum_{i=1}^{t} \frac{T_i^2}{r_i} - \frac{G^2}{n}$ |
| $X_\beta \mid J, X_\tau$ | t-1 | Difference |
| $I \mid J, X_\beta, X_\tau$ | n-b-t+1 | From Table 1 |
| Total | n-1 | $\sum_{ijl} y_{ijl}^2 - \frac{G^2}{n}$ |

# Cont'd

- The T|B ANOVA is the appropriate ANOVA for the intrablock analysis as it allows to test the hypothesis:

$$H_0 : \tau_1 = \tau_2 = ... = \tau_t$$

by means of the appropriate test:

$$F = \frac{SS(X_\tau | J, X_\beta)/(t-1)}{SS(I | J, X_\beta, X_\tau)/(n-b-t+1)}$$

# Thank You!