Unit 1: Basic Research Design
Principles and Completely Randomized Designs

STA 643: Advanced Experimental Design

Derek S. Young
Learning Objectives

▶ Become familiar with experimental design terminology
▶ Know the difference between experimental units and measurement units as well as the respective errors
▶ Understand the basic strategy for planning, conducting, and analyzing results from an experiment
▶ Grasp the general experimental design philosophy, especially as it pertains to controlling for variability
▶ Understand how to characterize variability among EUs
Outline of Topics

1. Research Design Principles
2. Towards Understanding Variability
3. Completely Randomized Designs
Outline of Topics

1. Research Design Principles

2. Towards Understanding Variability

3. Completely Randomized Designs
Motivating Discussion

▶ In most of your statistics courses, you have primarily analyzed data where the design of the data collection and any data cleaning have already been done.
▶ However, most practitioners quickly learn that many problems that arise during the analysis stage could have been avoided had an experiment been designed and conducted properly from the outset.
▶ The study of experimental design (also called design of experiments or DOE) provides us with the tools, strategy, and knowledge to avoid such shortcomings.
▶ Sir Ronald A. Fisher is the father of modern statistics, especially as it concerns experimental design. He also has numerous insightful quotes, two of which pertain directly to the importance of experimental design:

“The best time to plan an experiment is after you’ve done it.”
“To consult the statistician after an experiment is finished is often merely to ask him to conduct a postmortem examination. He can perhaps say what the experiment died of.”

▶ The study of experimental design also has some specific terminology, more so than other areas of statistics.
▶ Thus, this first lecture will introduce some of the frequent terminology used in this course.
Planning for Research

- A research program is an organized effort on the part of a researcher to acquire knowledge about a natural or manufactured process.
- The entire program may require many individual studies, each with specific objectives.
  - These are sometimes called individual research projects.
- The individual studies usually answer related questions and provide related pieces of information, all with the end goal of the program being kept in mind.
- Good planning helps the researcher organize the required tasks in order to perform an effective research study.
Documented Plans

- By documenting plans, a researcher can avoid serious oversights. Some items to consider include:
  1. the specific objectives of the experiment
  2. identification of important factors and which should be varied and which should be held constant
  3. the characteristics to be measured
  4. specific procedures for conducting tests
  5. the number of repetitions of the basic experiment to conduct
  6. available resources, materials, and cost constraints

- Simple questions can help aid the design process. Such simple questions might include:
  - What is my objective?
  - What do I want to know?
  - Why do I want to know a certain result?
  - Can any improvements be made?
Terminology

- An **experiment** shall be confined to investigations that establish a particular set of circumstances under a specified protocol to observe and evaluate implications of the resulting observations.
- The investigator establishes and controls the protocols in an experiment.
- The intent is to evaluate and test something that, for the most part, is unknown up to that time.
- A **comparative experiment** is where there is more than one set of circumstances in the experiment and that the results from the differing circumstances will be compared with one another.
  - This is the type of experiment familiar to those in experimental sciences, like biology, chemistry, and agriculture.
Terminology

- **Treatments** are the set of circumstances created for the experiment in response to research hypotheses and are often the focus of the investigation.

- Examples of treatments include:
  - temperatures
  - soil type
  - concentration of a certain chemical in a chemical mixture
  - specific diet
  - quantity of a certain type of drug
  - percentage of copper in a proposed type of tubing
Terminology

- The **experimental unit** (or EU) is the smallest unit or group to which a treatment is applied.
- The **measurement unit** (or MU) is the unit on which the measurement is actually made. This is also called the observational unit or sampling unit.
- In many cases, EU=MU, which is called a **simple experiment**. When EU≠MU, we have a **complex experiment**.
- Here are two simple examples:
  1. An experiment is conducted at a hospital to test the effectiveness of three back braces – say, A, B, and C – on children with scoliosis. Five children were randomly assigned to wear each brace for six months, at which time the curvature in the spine was recorded. What are the EU, MU, and treatment?
  2. In a nutrition experiment, two horses were randomly assigned to each of four diets. After four months, the growth of each hoof was measured. What are the EU, MU, and treatment?
Experimental error is the variability among independent and identically distributed (iid) EUs. Various origins of experimental error include:
- natural variation among EUs;
- variability in measurement of the response;
- inability to reproduce the treatment conditions exactly from one unit to another;
- interaction of treatments and EUs; and
- any other extraneous factors that influence the measured response.

Measurement error is the variability among measurements made on the same EU, including:
- instrument error (absorbed into experimental error when EU=MU) and
  - error of MU within an EU (nonexistent when EU=MU).

Differences among treatment means must be tested against experimental error.
The research hypothesis establishes a set of circumstances and the consequences that follow from those circumstances.

The treatments are a creation of the circumstances for the experiment.

Thus, it is important to make a decision as to which treatments, say, $T_1, T_2, \ldots, T_t$, are to be included in the experiment.

This process of making decisions is called the treatment design.

Each treatment may be a combination of levels of two or more factors.

Properly chosen treatments can lead to better understanding of underlying mechanisms of research hypotheses.
Example: Break Strength

- Suppose we are planning a study of the breaking strength (ft-lb) of a certain product and the effect that two alloying elements – Nickel and Manganese – have on that strength.
- Consider the two possible designs below. Which, intuitively, seems “better”?

<table>
<thead>
<tr>
<th></th>
<th>Design 1</th>
<th></th>
<th></th>
<th>Design 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Nickel</td>
<td>% Manganese</td>
<td></td>
<td>% Nickel</td>
<td>% Manganese</td>
</tr>
<tr>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.21</td>
<td>2.50</td>
<td>3.12</td>
</tr>
<tr>
<td>3.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.60</td>
<td>3.21</td>
<td>3.42</td>
</tr>
<tr>
<td>0.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.44</td>
<td>2.50</td>
<td>0.45</td>
</tr>
<tr>
<td>3.00</td>
<td>2.00</td>
<td>2.00</td>
<td>3.21</td>
<td>1.60</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Table: Two different designs for the break strength study.
Example: Break Strength

- To the right is a plot of the design points for the two different designs.
- Clearly, Design 1 has a rectangular shape in the design space.
- Design 1 is orthogonal (to be discussed later) and would allow us to properly estimate each factor’s effect on the break strength.
- Design 2 is more like what we would record in an observational study, where the levels of the factors are recorded with the response and not specified beforehand.
Experimental Designs

- Experimental designs are the framework for decision-making about how EUs are to be allocated to treatments and how many EUs are to be used in the experiment.
  - This is the primary focus of this course.
- The attainment of maximum information, precision, and accuracy in the results, along with the most efficient use of existing resources, is a guiding principle in choosing an appropriate experimental design.
- Some of the basic experimental designs we eventually discuss:
  - completely randomized designs
  - randomized block designs
  - Latin square designs
  - incomplete block designs
  - split-plot designs
Basic Principles in Experimental Design

Generally, we design our studies so that the following are satisfied:

1. We maximize our power of detecting the differences of interest.
2. We maximize the precision of our estimates.
3. Our results are valid within the study population, which is called internal validity.
4. Our results can be extended to the general population of interest, which is called external validity.
5. We achieve our results within our research constraints, which includes monetary expenditures, time to conduct the entire study, and ethical considerations of EUs (e.g., animals or humans).
Basic Principles in Experimental Design

- Power and precision are functions of sample size, variance of the study population, and the design points.
- The variance of the study population can be reduced by various techniques (e.g., blocking and controlling extraneous factors).
- Internal validity is achieved by randomization and control.
- External validity depends on the nature of the relationship between the sample and the population, where the former should be highly-representative of the latter.
- Optimizing with respect to research constraints will require the investigator to be vigilant in adhering to such constraints when considering how to obtain the end results of the study.
Control Treatments

- The **control treatment** is a necessary benchmark treatment to evaluate the effectiveness of experimental treatments.
- Conditions of the experiment may disallow the effectiveness of the experimental treatments that are known generally to be effective, thus, a control of *no treatment* will reveal the conditions under which the experiment was conducted.
- For example, when examining test tubes for catalytic reactions of enzymes when added to a specific substrate, the control test tube would be identical to all other test tubes with the exception of lacking the enzyme.
- Sometimes treatments require manipulating the EUs or subjects where the manipulation alone can produce a response.
- **Placebos** establish a basis for treatment effectiveness such that processing a placebo unit or subject is just like treatment units, but the active treatment is not included or disclosed.
- For example, when testing a drug for the common cold, half of the subjects could receive the actual drug while the other half of the subjects receive a sugar pill (placebo) in its place.
Blinding

- Subjects positive response to a placebo is called the **placebo effect**.
- If subjects in the control group know that they are receiving a placebo, the placebo effect will be reduced or eliminated and the placebo will not serve its intended control purpose.
- **Blinding** is the practice of not telling subjects whether they are receiving a placebo.
- By blinding, subjects in the control and treatment groups experience the placebo effect equally.
- Oftentimes, knowledge of which groups receive placebos is also kept from people who administer or evaluate the experiment, which is called **double blinding**.
- Double blinding prevents the experimenter from influencing (intentionally or unintentionally) subjects and it assures that the analyst’s evaluation is not contaminated by awareness of the true treatment conditions.
Multiple-Factor Treatments

- A factor is a controlled independent variable whose levels are set by the experimenter.
  - The levels of a factor are, appropriately enough, called factor levels.

- A factorial treatment design (sometimes simply called a factorial design) is when one set of treatments is tested over all of the other sets of treatments.

- For example, suppose we are testing the effectiveness of different diet treatments on cattle. One factor is protein intake, which has two levels: high protein or low protein. A second factor is the amount of a vitamin mix, which has three levels: 25 mg, 50 mg, or 75 mg. How many treatments are studied in this experiment?
Factors

- We usually talk about treatment factors, which are the factors of primary interest to the investigator.
- In addition to treatment factors, there are nuisance factors which are not your primary focus, but still need to be dealt with.
  - Sometimes these are called blocking factors, mainly because we will try to block on these factors to prevent them from influencing the results.
- Note: Sometimes the usage of *treatment* and *factor* is not always transparent in the literature, but we will be consistent with how we present these terms.
- There are other ways to categorize factors, which we briefly discuss.
Experimental vs. Classification Factors

- **Experimental factors:** These are factors that you can specify (and set the levels) and then assign at random as the treatment to the EUs. Examples are temperature, level of an additive fertilizer amount per acre, and the amount of resin mixed in a substrate compound.

- **Classification factors:** These cannot be changed or assigned, but rather these come as labels on the EUs. Examples are age, height, and sex of the study participants. These are all classification factors which cannot be changed or randomly assigned. However, you can select individuals from these groups randomly.
Quantitative vs. Qualitative Factors

- **Quantitative factors**: These factors can be assigned any practically-specified level. Examples are percent or pH level of a chemical, amount of water added to a mixture, and dosage of medicine administered to a subject.

- **Qualitative factors**: These factors have categories which are different types. Examples are species of a plant or animal, gender, and brands in the marketing field (e.g., Coke, Pepsi, and Dr. Pepper in a taste test of different sodas). Qualitative factors are not ordered or continuous, but are arranged, perhaps, in sets.
Example: Herbal Supplement Study

A nutritionist wants to test the effectiveness of a performance-enhancing herbal supplement on people who exercise rigorously 4-5 times a week. To create experimental groups that are similar at the beginning of the study, the subjects are assigned into two groups at random. Subjects in both groups are given a pill to take everyday, but they do not know whether the pill is a placebo (sugar pill) or the herbal supplement. The nutritionist gives Group A the herbal supplement and Group B the placebo. The subjects’ fitness levels are compared before and after six weeks of consuming the supplement or the sugar pill. No differences in performance ability were found between the two groups suggesting that the herbal supplement was not effective.

1. What kind of study is this?
2. What is the factor and what is its type?
3. What are the levels of this factor?
4. What are the EUs and the MUs?
5. What is the response?
Example: Washing Machine Study

An experiment is conducted comparing the effectiveness of two brands of washing machines (Whirlpool and Samsung) and three different models (space-saver, standard capacity, and extra large capacity). Separate, but equally soiled, pieces of cloth were washed in each machine. Each cloth was then cut into 50 swatches and the amount of remaining dirt was measured using a microscope.

1. What kind of study is this?
2. What is the factors and what are each factor’s type?
3. What are the EUs and the MUs?
4. What is the response?
Outline of Topics

1. Research Design Principles
2. Towards Understanding Variability
3. Completely Randomized Designs
Replication

- **Replication** implies an independent repetition of the basic experiment.
- The scientific community regards replication of experiments to practically be a requirement for obtaining valid experimental results.
- Recall that the standard error of the mean is $\sqrt{s^2/n}$, where $s^2$ is the sample variance and $n$ is the sample size. The width of the corresponding confidence interval for the mean is determined by the standard error. Thus, our estimate of the mean becomes less variable as the sample size increases.
- Replication is the bedrock of every method we will use to get a handle on how precise our estimates are.
- Replication also allows us to estimate or control the uncertainty in our results.
How Many Replications?

- The number of replications (often denoted by $r$) in a research study affects
  - precision of estimates for treatment means, and
  - power of statistical tests to detect differences between the means of treatment groups.
- Cost also practically constrains the number of replications.
- Thus, replication numbers are also a function of practical constraints that are germane to the study.
- Also directly impacts the degrees of freedom (df) that are available for the testing.
The method for determining the number of replications is often based on a test of a hypothesis about differences among treatment group means.

We will discuss an elementary method for experiments with two independent samples to illustrate a few attributes of the determining the number of replicates.

The method is based on a hypothesis test about the difference between two treatment group means: \( d = m_1 - m_2 \), with known experimental error variance \( \sigma^2 \), using the standard normal distribution test statistic.

The method determines the minimum number of replications required to test the difference between two independent sample means with specified Type I and Type II errors.
Replication Numbers for Testing Hypotheses

The required number of replications is affected by four primary factors that are required for the calculation:

1. the variance, $\sigma^2$
2. the size of the difference between the two means, $\delta$
3. the significance level (or the probability of a Type I error) of the test, $\alpha$
4. the power of the test, $1 - \beta$, or the probability of detecting $\delta$, where $\beta$ is the probability of a Type II error

Note that $\sigma^2$ will often be posited based on previous studies or known information, $\delta$ will be a difference determined to be of practical importance, and $\alpha$ and $1 - \beta$ will usually be chosen according to traditional rule-of-thumb values (e.g., typical values are $\alpha \in \{0.01, 0.05, 0.10\}$ and $\beta \in \{0.05, 0.10, 0.20\}$). The required replication number for each treatment group for a two-sided alternative is estimated with

$$r \geq 2 \left( \frac{\sigma \left[ z_{\alpha/2} + z_{\beta} \right]}{\delta} \right)^2,$$

where $z_P$ is the standard normal variate exceeded with probability $P$. 
Using the CV

- The replication number can also be estimated with knowledge of the coefficient of variation, or CV, which is given by

\[ CV = \frac{\sigma}{\mu} \]

- The CV is a unitless, standardized measure of the amount of dispersion of a probability distribution.
- We can also use the percent CV, which is given by

\[ \%CV = 100 \times CV \]

- For the replication formula on the previous slide, we can substitute \( \%CV \) for \( \sigma \) and the percent difference \( \%\delta = 100 \times (\delta/\mu) \) for \( \delta \), giving us

\[ r \geq 2 \left( \frac{\%CV \left[ \frac{z_{\alpha/2} + z_\beta}{\%\delta} \right]}{\%\delta} \right)^2 \]
We have thus far only discussed replication numbers in the context of two independent samples; however, the general influence of different quantities are similar in more complex experiments.

Required replication numbers generally increase if

- \( \sigma^2 \) or \( \%CV \) increases
- the size of the difference to detect, \( \delta \) or \( \%\delta \), decreases
- the significance level of the test, \( \alpha \), decreases
- the power of the test, \( 1 - \beta \), increases (or, equivalently, the Type II error rate \( \beta \) decreases)

Again, usually variance estimates from previous studies help guide us to set \( \sigma^2 \) for the presented formulas.
Example: Horse Blood Pressure

Suppose we want to test the effect of two different diets (the treatments) on systolic blood pressure of thoroughbred horses (the EUs). The measurements are in mmHg, or millimeters of mercury. A difference of 10 mmHg is desired to be detected. From a similar study, a variance of $\sigma^2 = 16$ is assumed. The Type I error rate is controlled at $\alpha = 0.05$ (i.e., the 95% significance level) and the power of the test is to be controlled at 0.90 (or 90%). Therefore, the minimum number of replicates necessary for this design is

$$r = \left\lceil 2 \left( \frac{4(1.96 + 1.28)}{10} \right)^2 \right\rceil = 4,$$

where $z_{0.05/2} \approx 1.96$ and $z_{0.10} \approx 1.28$. Note that the notation $\lceil \cdot \rceil$ is the ceiling function, which means that you round up to the next highest integer.
Variation

- Controlling for variation in an experiment is crucial in order to obtain accurate and meaningful results.
- In experimental design, we often use a model to characterize and estimate the different sources of variation.
- Subsequent statistical tests (mostly $F$-tests), utilize the estimated variability in order to determine the statistical significance of a given factor or treatment.
- A statistically significant result does not mean the result is practically significant, so further measures, such as relying on a content-expert’s opinion and using visualization approaches, can help in determining the practical significance of a result.
- We will explore variability among EUs and the effect when controlling for factors.
Example: Cigarette Data

We are interested in determining if the tar content (in milligrams) for three different brands of cigarettes differ. A lab took 6 cigarettes (i.e., the EUs) from each of the three brands (i.e., the treatments). The data are provided in the following table:

<table>
<thead>
<tr>
<th>Brand A</th>
<th>Brand B</th>
<th>Brand C</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.21</td>
<td>11.32</td>
<td>11.60</td>
</tr>
<tr>
<td>10.25</td>
<td>11.20</td>
<td>11.90</td>
</tr>
<tr>
<td>10.24</td>
<td>11.40</td>
<td>11.80</td>
</tr>
<tr>
<td>9.80</td>
<td>10.50</td>
<td>12.30</td>
</tr>
<tr>
<td>9.77</td>
<td>10.68</td>
<td>12.20</td>
</tr>
<tr>
<td>9.73</td>
<td>10.90</td>
<td>12.20</td>
</tr>
</tbody>
</table>

Table: Tar content of cigarettes.
Example: Cigarette Data

Suppose that we treat the data as univariate data. Then, $\bar{x} = 11.00$ and $s^2 = 0.79$. The data can be plotted using a dotplot as we’ve done below. The dotplot allows us to visualize the location (mean) and spread (standard deviation) of the data.
Example: Cigarette Data

► When we summarize the data by the brand, we obtain the dotplot to the right.

► The estimated means and variances for the three brands are:
  - Brand A: $\bar{x} = 10.00$ and $s^2 = 0.07$
  - Brand B: $\bar{x} = 11.00$ and $s^2 = 0.13$
  - Brand C: $\bar{x} = 12.00$ and $s^2 = 0.08$

► The variability within each group is comparable across the brands and clearly the variabilities are smaller than the overall variability.

► There is a clear, visual separation between the groups. Without yet performing a formal ANOVA, we would expect to find that there is a significant difference between the group means.
We can also use side-by-side boxplots for comparing.

If the boxplots each looked similar to one another (in terms of location and spread), then we would expect to not find any significant difference between the treatments (brands).

Recall that the bottom and top of the box are the first and third quartiles (Q1 and Q3), respectively; the horizontal line within the box is the median; the solid diamond is the mean; the interquartile range (IQR) is Q3-Q1 such that the whiskers extend $1.5 \times \text{IQR}$ below Q1 or above Q3, or to the minimum and maximum, respectively. This last criterion is sometimes called an inner fence.

Suspected outliers (if present) will be marked by asterisks outside of the inner fence threshold.

Sometimes an outer fence criterion is used, which would identify any values $3.0 \times \text{IQR}$ below Q1 or above Q3.
Blocking to Reduce Experimental Error Variation

- Sir Ronald Fisher argued that no advantage had to be given up to have a valid estimate of error, but two things were necessary:
  1. that a sharp distinction should be drawn between those components of error which are to be eliminated in the field and those which are not to be eliminated, and
  2. the statistical process of the estimation of error shall be modified so as to take account of the field arrangement, and so that the components of error actually eliminated in the field shall equally be eliminated in the statistical laboratory.

- **Blocking** is a technique to include other factors in our experiment which contribute to undesirable variation.
  - Blocking provides local control of the environment to reduce experimental error.

- When blocking, the EUs are grouped such that the variability of units within the groups is less than that among all units prior to grouping.

- We want the unknown error variance at the end of the experiment to be as small as possible, so blocking is one way to achieve that goal.
Criteria for Blocking

- Four major criteria frequently used to block EUs are:
  1. proximity (e.g., neighboring field plots, housing units)
  2. physical characteristics (e.g., age, weight)
  3. time (e.g., hours, days)
  4. management of tasks in the experiment (e.g., blocking on individual technicians, assigning one person to each replicate when many people are available)

- Some classic blocking examples include:
  - placing contiguous plots into one group for agricultural field experiments, where each of the treatments was assigned to a plot in that group;
  - using the animal litter as a blocking unit; and
  - using a single batch of raw materials for industrial experiments such that the batch is sufficiently large for one replication of all treatments of interest.
Example: Tree Heights

- Suppose we have the following data on $n = 8$ tree heights (in inches): 83, 86, 82, 80, 67, 71, 74, 72.
- The sample mean and sample variance of these trees are $\bar{x} = 76.9$ and $s^2 = 45.8$, respectively.
- However, we know that the first four trees were planted 5 years earlier than the last four trees.
- If we block on when the trees were planted, then we have
  - Block 1: 83, 86, 82, 80  $\bar{x} = 82.8$ and $s^2 = 6.3$
  - Block 2: 67, 71, 74, 72  $\bar{x} = 71.0$ and $s^2 = 8.7$
- Thus, the total variance has been drastically reduced when considering the variances within the separate blocks.
Confounding

- **Confounding** is when an extraneous variable correlates with both the response variable and one or more of the treatment factors.
  - Thus, when we have variables (or treatments) that are confounded, we are not able to properly estimate the effect of either variable on the response.

- For example, a drug manufacturer tests a new cold medicine with 200 volunteer subjects - 100 men and 100 women. The men receive the drug and the women do not. At the end of the test period, the men report fewer colds. This experiment implements no controls and many variables are confounded. Thus, it is impossible to say whether the drug was effective. For example, gender is confounded with drug use.
The primary focus of this course will be on which experimental design is most appropriate for the researcher’s scientific questions of interest.

Of course, it will also be helpful to know how to analyze the resulting data!

The primary approach used when analyzing results from an experiment is the analysis of variance or ANOVA.

We will present the models used for the experimental designs of interest, how to set-up the design matrix for those models, and how to construct the ANOVA table.

You already have experience with ANOVA tables, so you already have the basic understanding (and hopefully, intuition) about the quantities in these tables.
Planning, Conducting, and Analyzing an Experiment

Below are the general steps for the scientific method:

1. Recognition and statement of the problem
2. Choice of factors, levels, and ranges
3. Selection of the response variable(s)
4. Choice of design (the primary focus of this course)
5. Conducting the experiment
6. Performing statistical analysis
7. Drawing conclusions and making recommendations
Outline of Topics

1. Research Design Principles
2. Towards Understanding Variability
3. Completely Randomized Designs
Randomization

- **Randomization** is the random assignment of treatments to EUs.
- Randomization is an essential component of any experiment that is going to have validity.
- If you are doing a comparative experiment where you have two treatments, a treatment and a control for instance, you need to include in your experimental process the assignment of those treatments by some random process.
- You need to have a deliberate process to eliminate potential biases from the conclusions, and random assignment is a critical step.
Randomization Tests

- Independent observations are critical for estimation and tests of hypotheses because they provide valid estimates of experimental error variance.
- Randomization provides appropriate reference populations for statistical inference free of any assumptions about the distribution of the observations.
- Significance tests can be based on the distribution created by randomization and normal theory tests provide reasonable approximations to these results.
  - Thus, the random allocation of treatments to the EUs simulates the effect of independence and permits us to proceed as if the observations are iid normal.
- A **randomization test** (also called a **permutation test** or an **exact test**) is a significance test in which the distribution of the test statistic under the null hypothesis ($H_0$) is obtained by calculating all possible values of the test statistic under rearrangements (i.e., permutations) of the labels on the observed data points.
- In other words, the method by which treatments are allocated to subjects in an experimental design is mirrored in the analysis of that design.
- If the labels are exchangeable under $H_0$, then the resulting tests yield exact significance levels.
  - A finite sequence of random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is exchangeable such that any permutation of the indices, $\mathbf{X}_\pi = (X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$ has the same joint probability distribution as the original sequence. Moreover, this definition can be extended to an infinite sequence.
More on the Utility of Randomization

▶ The utility of randomization can be demonstrated with a randomization test that makes no assumptions about the form of the probability distribution for the observations.

▶ Randomization creates a population of experiments that could have been performed.
  ▶ Although only one arrangement has been chosen at random for the actual experiment.

▶ The randomization test evaluates the test statistic for all possible arrangements of treatments on the EUs.

▶ The **randomization distribution** is the distribution of those values that would be obtained under $H_0$ of no treatment effects.
Consider an experiment in which two treatments (A and B) are randomly assigned to four EUs. Thus, two EUs receive treatment A and two EUs receive treatment B. The results are given in the table at the bottom. If there is no difference between the effects of these two treatments, then they are merely labels on the EUs and do not affect the results. For example, if $H_0$ were true, then the result for Unit 1 would be 12, regardless if treatment A or treatment B were applied. The labels may be allocated to the four EUs in \( \binom{4}{2} = \frac{4!}{(2!2!)} = 6 \) possible arrangements. These are the 6 possible experiments if the treatments were randomly assigned to the EUs. All 6 possible arrangements, along with the difference between group means \( \bar{y}_A - \bar{y}_B \) based on the assignments, are given on the next slide.

<table>
<thead>
<tr>
<th>Unit</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Response</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>9</td>
</tr>
</tbody>
</table>

Table: Measurements from hypothetical experiment.
Example: Illustrating Randomization Test

<table>
<thead>
<tr>
<th>Arrangement</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>$(\bar{y}_A - \bar{y}_B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>-5</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>-5</td>
</tr>
</tbody>
</table>

Table: Arrangements of treatments from hypothetical experiment.

Consider testing $H_0 : \mu_A - \mu_B = 0$ versus the alternative $H_A : \mu_A - \mu_B \neq 0$. The arrangement corresponding to the actual experiment on the previous slide is Arrangement 2, which has a difference of -5. There are 2 arrangements with an absolute value of 5 or higher, which yields a frequency of $2/6 \approx 0.33$, which would be the exact $p$-value. Therefore, we would fail to reject $H_0$. 
Relative Efficiency of Experimental Designs

- **Efficiency** is the measure of optimality of an estimator, hypothesis testing, or experimental design, with our focus being on the latter.
  - Efficiency relates to the ability of an experimental design to achieve the objective of the study with minimal expenditure of resources, such as time and money.

- **Relative efficiency** measures the effectiveness of blocking (or some other mechanism) in experimental designs to reduce experimental error variance.
  - In simple cases, the relative efficiency of experimental designs can be expressed as the ratio of the sample sizes required to achieve an objective.

- Relative efficiency is measured to determine the efficiency of the design actually used relative to another simpler design that *could have been* used.

- The variance of a treatment mean \( \sigma^2_{\bar{y}} = \sigma^2/r \) is a measure of the precision of the estimated treatment means in an experiment.

- The use of \( \sigma^2_{\bar{y}} \) as a measure of precision provides a means for comparing the relative precision of two experimental designs.
Example: Relative Efficiency

Suppose that one design has a true experimental error variance of $\sigma_1^2 = 1$ and a second design has a true experimental error variance of $\sigma_2^2 = 5$. Therefore, $\sigma_2^2 = 5\sigma_1^2$. The variance of the treatment mean in each design is

Design 1: $\sigma_{\bar{y}_1}^2 = \frac{\sigma_1^2}{r_1}$

Design 2: $\sigma_{\bar{y}_2}^2 = \frac{\sigma_2^2}{r_2} = \frac{5\sigma_1^2}{r_2}$

The variances $\sigma_{\bar{y}_1}^2$ and $\sigma_{\bar{y}_2}^2$ will be the same only if $r_2 = 5r_1$, or rather if Design 2 has five times as many replicates as Design 1. Therefore, Design 1 is more efficient than Design 2 with respect to the number of replications required to have the same precision for an estimate of the treatment mean.
Unknown (Again!)

- As we have emphasized, $\sigma^2$ will be unknown in practice for each of the designs and must be estimated from the data.
- Also, the df for the estimate of the variance changes with designs.
- Under these circumstances, the relative precision of the two designs is determined based on the (Fisher) information, which is the amount of information that the estimated difference between two means provides about the true difference between the population means.
- The calculation for information under the above paradigm is

$$I = \frac{f + 1}{f + 3} \frac{1}{s^2},$$

where $s^2$ is the estimated experimental error variance with $f$ df.
- If $\sigma^2$ is assumed known, then $I = \sigma^{-2}$ and the coefficient $(f + 1)/(f + 3) \equiv 1.$
Relative Efficiency of Experimental Designs

- The relative efficiency \( (RE) \) of two experimental designs is defined as the ratio of their respective informations:

\[
I_1 = \frac{(f_1 + 1)}{(f_1 + 3)} \frac{1}{s_1^2} \quad \text{and} \quad I_2 = \frac{(f_2 + 1)}{(f_2 + 3)} \frac{1}{s_2^2}
\]

\[
\Rightarrow RE = \frac{I_1}{I_2} = \frac{(f_1 + 1)(f_2 + 3)}{(f_1 + 3)(f_2 + 1)} \frac{s_2^2}{s_1^2}.
\]

- If \( RE = 1 \), then the information in the two designs is equal and the designs each require the same number of replications to have the same variance of the treatment means.

- If \( RE > 1 \), then Design 1 is more efficient than Design 2.
  - For example, if \( RE=4 \), then Design 2 requires four times the number of replicates than Design 1.

- If \( RE < 1 \), then Design 1 is less efficient than Design 2.
  - For example, if \( RE=0.5 \), then Design 2 requires half the number of replicates than Design 1.
Completely Randomized Design (CRD)

▶ Suppose we are designing an experiment for a study with a treatment design consisting of \( t \) treatments: \( T_1, T_2, \ldots, T_t \).

▶ A **completely randomized design** (or CRD) involves allocating \( n_1, n_2, \ldots, n_t \) EUs to treatments \( T_1, T_2, \ldots, T_t \), respectively.

▶ The units are randomly allocated such that every possible combination of \( n_1, n_2, \ldots, n_t \) units are equally likely to occur.
  
  ▶ Recall how this was an underlying concept of the exact \( p \)-value calculation for the randomization test.

▶ The total number of units is denoted by \( N = \sum_{i=1}^{t} n_i \).
  
  ▶ Note that \( N \) is often used to denote the unknown population size; however, we will likely not make reference to a *known* population size, so it will be understood that \( N \) refers to the total number of EUs in a study.

▶ When \( n_i \equiv n \) for all \( i \) (and, thus, \( N = tn \)), then we have a **balanced design**.
How to Randomize

- Assuming that the EUs are (relatively) homogeneous, a CRD can be used to avoid any subjective assignment of treatments to the EUs.

- The proper randomization procedure for a CRD is given below:

1. For $N$ EUs, assign the sequence of numbers $1, \ldots, N$ to the EUs.
2. For an unbalanced design, determine the number of EUs that are to be allocated to each of the $t$ treatments. For a balanced design, this step is unnecessary as $n$ EUs will be allocated to each treatment.
3. Obtain a random permutation of the numbers of $1, \ldots, N$. Then, allocate the first $n_1$ EUs to $T_1$, the next $n_2$ EUs to $T_2$, and so on until you allocate the last $n_t$ EUs to $T_t$. Again, for a balanced design, you would just replace $n_1, n_2, \ldots, n_t$ by $n$ in each step of the allocations.

- Any statistical software can be used to provide a simple permutation of a sequence of numbers.
Example: Blood Coagulation

Below is a table giving the coagulation time for blood samples drawn from $N = 24$ animals (EUs) receiving four different diets (treatments): A, B, C, and D. The reported data have been rounded to whole numbers. The animals were randomly allocated to the diets and the order in which the blood samples were drawn are indicated by the superscripts in parentheses. For example, the third animal in Treatment C was the first animal measured and the sixth animal in Treatment A was the last animal measured.

<table>
<thead>
<tr>
<th>Diet (Treatment)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>62(20)</td>
<td>63(12)</td>
<td>68(16)</td>
<td>56(23)</td>
</tr>
<tr>
<td></td>
<td>60(2)</td>
<td>67(9)</td>
<td>66(7)</td>
<td>62(3)</td>
</tr>
<tr>
<td></td>
<td>63(11)</td>
<td>71(15)</td>
<td>71(1)</td>
<td>60(6)</td>
</tr>
<tr>
<td></td>
<td>59(10)</td>
<td>64(14)</td>
<td>67(17)</td>
<td>61(18)</td>
</tr>
<tr>
<td></td>
<td>63(5)</td>
<td>65(4)</td>
<td>68(13)</td>
<td>63(22)</td>
</tr>
<tr>
<td></td>
<td>59(24)</td>
<td>66(8)</td>
<td>68(21)</td>
<td>64(19)</td>
</tr>
</tbody>
</table>

Table: Blood coagulation times of animals.
Example: Blood Coagulation

*Research Hypothesis:* The researcher is interested in determining the effects of different diets (each with different levels of nutrients, calories, and fat) on the coagulation times of animal’s blood samples.

*Treatment Design:* The treatment design consists of $t = 4$ diets (treatments), each having a certain balance of nutrients, fat levels, and calories. Note that these elements of the diets are *not* controlled as individual factors in the design.

*Experimental Design:* Since the animals were randomly allocated to the diets and the order in which the blood samples were drawn, this is a balanced CRD with $n = 6$. 
Example: Blood Coagulation

As we noted, the blood coagulation study is a balanced CRD. So $n = 6$ animals (EUs) were allocated to each of the $k = 4$ diets (treatments). The superscripted numbers indicate the permutation results of the randomization used by the researchers. For the sake of completeness, we provide a table below showing another possible permutation that could have resulted in the assignment of treatments to the EUs.

<table>
<thead>
<tr>
<th>Diet (Treatment)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>19</td>
<td>20</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>18</td>
<td>21</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>4</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>14</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>12</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Table: Another possible permutation.
A statistical analysis is based on an underlying formal statistical model. Proper interpretation of the analysis requires an understanding of the model. In comparative studies, the measurement taken on the units is the response variable, denoted by $Y$. The statistical model for comparative studies assumes that there is a reference population of the EUs. Each individual unit in the population has a value for the response variable $Y$, which has mean $\mu$ and variance $\sigma^2$. A reference population is assumed for each treatment condition in the study and the EUs are assumed to be realizations of draws from the reference population as a result of randomization. $\sigma^2$ is assumed to be the same for each of the populations and unaffected by the treatments.
Observations are expressed as a sum of the treatment population means and the experimental errors with what we call the cell means model:

\[ Y_{ij} = \mu_i + \epsilon_{ij}, \quad (1) \]

where

- \( i = 1, \ldots, t \) is the index over the treatment levels;
- \( j = 1, \ldots, n_i \) is the index over the EUs from treatment \( i \);
- \( \mu_i \) is the \( i^{th} \) treatment mean; and
- \( \epsilon_{ij} \) is the experimental error.

The above is a linear statistical model for the one-way factor classification in a CRD.

Note that since we are only considering one factor, the factor dually serves as the treatment.
Cell Means Model

- The cell means model allows for variation among the observations from a given treatment group.
- Because of experimental error, each observation deviates from the population mean \( \mu_i \) by some amount \( \epsilon_{ij} \), or error.
- The experimental error variance \( \sigma^2 \) is the variance of the \( \epsilon_{ij} \)s, and thus is assumed to be the same for all treatment populations.
- Note that the model is written in terms of the random variable \( Y_{ij} \) on the left-hand side, but when we obtain a realization of that random variable from our study, then it will be written as \( y_{ij} \).
Hypothesis Test - Cell Means Model

- We want to test the hypothesis that the means are equal versus at least one is different; i.e.,

\[ H_0 : \mu_1 = \mu_2 = \ldots = \mu_t \]
\[ H_A : \mu_i \neq \mu_k \quad \text{for some } i \neq k \]

- The null (reduced) model represents no differences among the treatment means.
- The alternative (full) model represents some differences among the treatment means.
To the right is a hypothetical visualization if faced with one population (top) versus several populations (bottom).

The top figure would correspond to

\[ y_{ij} = \mu + \epsilon_{ij} \]

The bottom figure would correspond to

\[ y_{ij} = \mu_i + \epsilon_{ij} \]
An equivalent way to formulate the cell means model, which is often done for the mathematical convenience, is to use the **factor effects model**:

\[
Y_{ij} = \mu + \tau_i + \epsilon_{ij},
\]

(2)

where

- \(i = 1, \ldots, t\) is the index over the treatment levels;
- \(j = 1, \ldots, n_i\) is the index over the EUs from treatment \(i\);
- \(\mu\) is the grand mean;
- \(\tau_i\) is the deviation from the grand mean due to the \(i^{th}\) treatment; and
- \(\epsilon_{ij}\) is the experimental error.
Hypothesis Test - Factor Effects Model

- We want to test the hypothesis that the means are equal versus at least one is different; i.e.,

\[ H_0 : \tau_1 = \tau_2 = \ldots = \tau_t = 0 \]

\[ H_A : \text{not all } \tau_i \text{ equal } 0 \]

- The above is equivalent to the test for the cell means model.

- The statement of the null hypothesis follows from the fact that the grand mean \( \mu \) is common to all factor levels.
Notes About the Two ANOVA Models

▶ In the cell means model $E(Y_{ij}) = \mu_i$, while in the factor effects model $E(Y_{ij}) = \mu + \tau_i$.

▶ Regardless of the formulation, we assume that the $\epsilon_{ij}$ are iid with mean 0 and variance $\sigma^2$ (i.e., the experimental error).

▶ For the factor effects model, note that

$$\tau_i = \mu_i - \mu$$

$$\mu = t^{-1} \sum_{i=1}^{t} \mu_i.$$ 

▶ Splitting up the factor level mean $\mu_i$ into two components – a grand mean $\mu$ and a factor (treatment) effect $\tau_i$ – depends on the definition of $\mu$, which can be defined in various ways.
(1) Unweighted Mean

- An unweighted average of all the factor level means $\mu_i$ is

$$\mu = t^{-1} \sum_{i=1}^{t} \mu_i.$$ 

- The above definition implies that

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

because we have

$$\sum_{i=1}^{t} \tau_i = \sum_{i=1}^{t} (\mu_i - \mu) = \sum_{i=1}^{t} \mu_i - t\mu.$$

- Because $\sum_{i=1}^{t} \mu_i = t\mu$, this implies a restriction on the $\tau_i$ that they must sum to 0.
When performing least squares for estimating the factor effect model, it is sometimes more convenient to use a parameterization that utilizes a weighted average.

A weighted average of the factor level means $\mu_i$ is

$$
\mu = \sum_{i=1}^{t} w_i \mu_i, \quad \text{where} \quad \sum_{i=1}^{t} w_i = 1.
$$

The restriction on the $\tau_i$ implied by the above definition is

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

If the $w_i$ are unknown, one can use the sample sizes by setting $w_i = n_i / N$:

$$
\mu = t^{-1} \sum_{i=1}^{t} n_i \mu_i.
$$

Regardless if using the above or the unweighted mean, the null hypothesis of equality of the true means is equivalent to the null hypothesis that the treatment effects are all identically 0.
Before proceeding, we should clarify some notation.

Suppose we have $Y_{ij}$, where $i = 1, \ldots, t$ and $j = 1, \ldots, n_i$.

When we include a “·” in the subscript, that means we are summing over that particular index. Specifically:

\[
Y_i. = \sum_{j=1}^{n_i} Y_{ij}
\]

\[
Y_.j = \sum_{i=1}^{t} Y_{ij}
\]

\[
Y_{..} = \sum_{i=1}^{t} \sum_{j=1}^{n_i} Y_{ij}
\]
Example: Car Rental Firm

Let us first consider an example where it makes sense to posit weights for a study. A car rental firm wants to estimate the average fuel consumption (in miles per gallon) for its fleet of cars, which consists of 50% compacts, 20% station wagons, 20% sedans, and 10% luxury cars. Here, a meaningful measure of $\mu$ might be in terms of overall mean fuel consumption. Letting $\mu_1$, $\mu_2$, $\mu_3$, and $\mu_4$ be the mean fuel consumption for each of the four types of cars in the fleet, how would we define our value of interest, $\mu$? Then, how would we estimate $\mu$?
The cell means model is a special case of the **general linear model** (or **GLM**), which is the basis for regression modeling. (You should all be familiar with the GLM!)

This relationship allows us to utilize ordinary least squares (OLS) for estimation.

The response variable $Y$ is considered to be dependent on a set of $p$ fixed design variables or measured covariates: $X_1, X_2, \ldots, X_p$.

The GLM is the statistical model relating $Y$ to the $X_i$s through the $(p + 1)$ parameters $\beta_0, \beta_1, \ldots, \beta_p$ is

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon.$$  \hfill (3)

For a sample of size $n$, the above is then written in matrix notation as

$$Y = X\beta + \epsilon.$$  \hfill (4)
Example: Blood Coagulation

Suppose that the blood coagulation study had not been well-designed from the beginning. The experimenter might simply measure the animal’s daily intake of vitamin K (which is known to increase coagulation) and the coagulation times. The investigator may hypothesize that this relationship is linear, in which case they use the simple linear regression (SLR) model

\[ Y = \beta_0 + \beta_1 X_1 + \epsilon, \]

where \( x_1 \) is the amount of vitamin K intake by the animal. If the investigator thinks there is a quadratic effect, then they can use

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon, \]

where \( X_2 = X_1^2 \).
Regression Approach: Cell Means Model

- Recall that an **indicator variable** is a variable that equals 1 if a specific event occurs and 0 otherwise.
- Indicator variables are used when we have categorical variables.
- Consider a single-factor study with \( t = 3 \) and two EUs per treatment. The matrices are

\[
Y = \begin{pmatrix}
Y_{11} \\
Y_{12} \\
Y_{21} \\
Y_{22} \\
Y_{31} \\
Y_{32}
\end{pmatrix}, \quad X = \begin{pmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{pmatrix}, \quad \beta = \begin{pmatrix}
\mu_1 \\
\mu_2 \\
\mu_3
\end{pmatrix}, \quad \text{and} \quad \epsilon = \begin{pmatrix}
\epsilon_{11} \\
\epsilon_{12} \\
\epsilon_{21} \\
\epsilon_{22} \\
\epsilon_{31} \\
\epsilon_{32}
\end{pmatrix}
\]

- Because of the structure of \( X \), we must estimate \( \beta \) by not fitting an intercept term.
Regression Approach: Factor Effect Model

- Since $\sum_{i=1}^{t} \tau_i = 0$, it follows that
  $$\tau_t = -\tau_1 - \tau_2 - \cdots - \tau_{t-1}.$$  

- Thus, we only use the parameter $\mu, \tau_1, \tau_2, \ldots, \tau_{t-1}$ for the GLM.

- To define the factor effects model as a regression model, let $X_{ijk}$ equal 1 if case $j$ is from treatment $k$, equal -1 if case $j$ is from treatment $t$, and equal 0 otherwise. Note that $k$ goes from 1, $\ldots$, $t-1$.

- Consider a single-factor study with $t = 3$ and two EUs per treatment. The matrices are

$$Y = \begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{pmatrix}, \quad \beta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix}, \quad \text{and} \quad \epsilon = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{pmatrix}.$$
Regression Approach: Factor Effect Model

- Using the matrices defined on the previous slide, we see that

\[
E(Y) = \begin{pmatrix}
E(Y_{11}) \\
E(Y_{12}) \\
E(Y_{21}) \\
E(Y_{22}) \\
E(Y_{31}) \\
E(Y_{32})
\end{pmatrix} = X\beta = \begin{pmatrix}
\mu + \tau_1 \\
\mu + \tau_1 \\
\mu + \tau_2 \\
\mu + \tau_2 \\
\mu - \tau_1 - \tau_2 \\
\mu - \tau_1 - \tau_2
\end{pmatrix}.
\]

- Therefore, the factor effect model can be estimated using OLS using the formulation we provided above.
Example: Blood Coagulation

The cell means model for the blood coagulation study is

\[ Y_{ij} = \mu_i + \epsilon_{ij}, \]

where \( i = 1, \ldots, 4 \) and \( j = 1, \ldots, 6 \). We find that

\[ \hat{\mu} = 64 \quad \hat{\mu}_1 = 61 \quad \hat{\mu}_2 = 66 \quad \hat{\mu}_3 = 68 \quad \hat{\mu}_4 = 61. \]

We then get \( \hat{\beta} = (61, 66, 68, 61)^T \) as the OLS estimate. Thus,

\[ \bar{y}_1. = \hat{\mu}_1 = 61 \]
\[ \bar{y}_2. = \hat{\mu}_2 = 66 \]
\[ \bar{y}_3. = \hat{\mu}_3 = 68 \]
\[ \bar{y}_4. = \hat{\mu}_4 = 61. \]
Example: Blood Coagulation

The one-way factor effect model for the blood coagulation study is

\[ Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \]

where \( i = 1, \ldots, 4 \) and \( j = 1, \ldots, 6 \). We find that

\[ \hat{\mu} = 64 \quad \hat{\mu}_1 = 61 \quad \hat{\mu}_2 = 66 \quad \hat{\mu}_3 = 68 \quad \hat{\mu}_4 = 61 \]
\[ \hat{\tau}_1 = -3 \quad \hat{\tau}_2 = 2 \quad \hat{\tau}_3 = 4 \quad \hat{\tau}_4 = -3. \]

We then get \( \hat{\beta} = (64, -3, 2, 4)^T \) as the OLS estimate. Thus,

\[ \bar{y}_1 = \]
\[ \bar{y}_2 = \]
\[ \bar{y}_3 = \]
\[ \bar{y}_4 = \]
This is the end of Unit 1.