STA6246

Design and Analysis of Experiments

Instructor: Achraf Cohen, PhD,
acohen@uwf.edu

Department of Mathematics and Statistics
University of West Florida

Definitions

What is Statistics?

- The numerical facts or data in the news items ($100,000, 4.9%) commonly are referred to as statistics. In common, everyday usage, the term statistics refers to numerical facts or data.

Definitions

What is Statistics?

- The numerical facts or data in the news items ($\$100,000$, 4.9%) commonly are referred to as statistics. In common, everyday usage, the term statistics refers to numerical facts or data.
- The field of statistics involves much more than simply the computation and presentation of numerical data. In a broad sense the subject of statistics involves the study of how data are **collected**, how they are **analyzed**, and how they're **interpreted**. A major reason for collecting data, analyzing, and interpreting data is to provide engineers, managers, public, other researchers, with the information needed to make effective decisions.

Definitions

What is Statistics?

- The numerical facts or data in the news items (\$100, 000, 4.9%) commonly are referred to as statistics. In common, everyday usage, the term statistics refers to numerical facts or data.
- The field of statistics involves much more than simply the computation and presentation of numerical data. In a broad sense the subject of statistics involves the study of how data are **collected**, how they are **analyzed**, and how they're **interpreted**. A major reason for collecting data, analyzing, and interpreting data is to provide engineers, managers, public, other researchers, with the information needed to make effective decisions.

This course covers the methods that help collect the data and analyze them.

Definitions

Design of experiments

Statistical **design of experiments** refers to the process of planning the experiment so that the appropriate data will be collected and analyzed by statistical methods, resulting in valid and objective conclusions.

Introduction: Example 1

- As an example of an experiment, suppose that an engineer is interested in studying the effect of two different hardening processes, oil quenching and saltwater quenching, on an aluminum alloy.
- The objective of the experimenter is to determine which quenching solution (oil or saltwater) produces the maximum hardness for this particular alloy.
- The engineer decides to subject a number of alloy specimens to each quenching solutions and measure the hardness of the specimens after quenching. The average hardness of the specimens treated in each quenching solution will be used to determine which solution is best.

Introduction: Example 1

As we consider this simple experiment, a number of important questions come to mind:

- Are these two solutions the only quenching media of potential interest?
- Are there any other factors that might affect hardness that should be investigated or controlled in this experiment (such as the temperature of the quenching media)?
- How many coupons of alloy should be tested in each quenching solution?
- What method of data analysis should be used?
- What difference in average observed hardness between the two quenching media will be considered important?

All of these questions, and perhaps many others, will have to be answered satisfactorily before the experiment is performed.

Introduction

Well-designed experiments?

A well-designed experiment is crucial because the results and conclusions that can be drawn from the experiment depend to a large extent on the manner in which the data were collected.

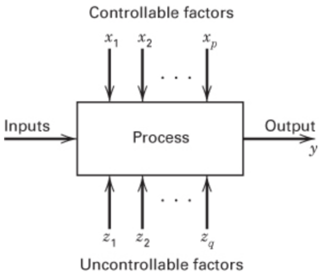
To illustrate this point, suppose that the engineer in the above experiment used specimens from one heat in the oil quench and specimens from a second heat in the saltwater quench. Now, when the mean hardness is compared, the engineer is unable to say how much of the observed difference is the result of the quenching media and how much is the result of inherent differences between the heats.

Thus, the method of data collection has adversely affected the conclusions that can be drawn from the experiment.

Introduction

The objectives of the experiment may include the following:

- Determining which variables are most influential on the response y
- Determining where to set the influential x's so that y is almost always near the desired nominal value
- Determining where to set the influential x's so that variability in y is small
- Determining where to set the influential x's so that the effects of the uncontrollable variables $z_1, z_2, \ldots, z_q$



Example 2: the golf experiment

Consider the golf game, some of the factors that may be important and can influence the golf score are:

1. The type of driver (oversized or regular sized)
2. The type of ball used (balata or three piece)
3. Walking and carrying the golf curbs or riding in a golf cart
4. Drinking water or drinking something else while playing
5. Playing in the morning or in the afternoon
6. Other factors

Engineers, scientists, and business analysts often decide that some factors are not important because of their effects that are small or have no practical value.

Strategy of Experimentation

The best-guess approach

The best-guess approach consists of selecting an arbitrary combination of these factors, test them and see what happens. For example, the following factors are selected in the first round:

- Oversized driver
- Balata ball
- Golf cart,
- Water

Strategy of Experimentation

The best-guess approach

The best-guess approach consists of selecting an arbitrary combination of these factors, test them and see what happens. For example, the following factors are selected in the first round:

- Oversized driver
- Balata ball
- Golf cart,
- Water

Next, the second round:

- Regular driver
- Balata ball
- Golf cart,
- Water

This approach could be continued almost indefinitely!

Strategy of Experimentation

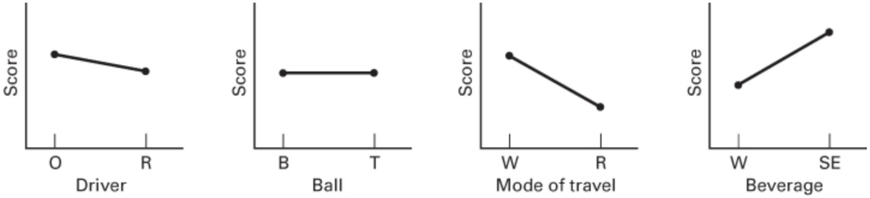
The best-guess approach

This approach is often used by engineers and scientists and its works reasonably well because the experimenters generally have a great technical knowledge of the process they are studying.
However, there are some advantages:

- suppose the initial best-guess does not produce the desired results then the experimenter should take another guess and this can continue for a long time
- suppose the initial best-guess does produce the desired results. Now the experimenter is tempted to stop testing although there is no guarantee that the best solution has been found.

One-factor-at-a-time (OFAT) Approach

The OFAT method consists of selecting a starting point, or baseline set of levels, for each factor, and then successively varying each factor over its range with the other factors held constant at the baseline level.

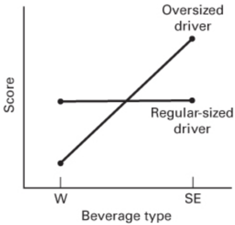


Strategy of Experimentation

One-factor-at-a-time (OFAT) Approach

The major disadvantage of the OFAT strategy is that it fails to consider any possible interaction between the factors.

An interaction is the failure of one factor to produce the same effect on the response at different levels of another factor



Strategy of Experimentation

The correct approach to dealing with several factors is to conduct **a factorial experiment**. This is an experimental strategy in which factors are varied together, instead of one at a time.

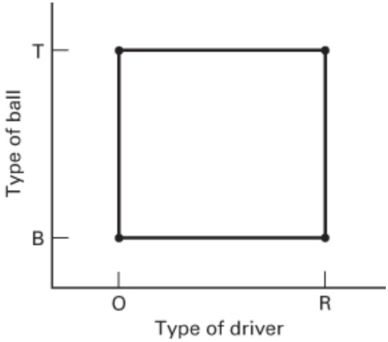


Figure: A two-factor factorial experiment

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.

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 **randomization**, **replication**, and **blocking**.

Basic Principles

Randomization

- 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.

Basic Principles

Randomization

- 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.

Basic Principles

Randomization

- 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.
- Also assist in averaging out the effects of extraneous factors that may be present.

For example, suppose that the specimens in the hardness experiment are of slightly different thicknesses. If all the specimens subjected to the oil quench are thicker than those subjected to the saltwater quench, we may be introducing systematic bias into the experimental results. This bias handicaps one of the quenching media and consequently invalidates our results. Randomly assigning the specimens to the quenching media alleviates this problem.

Basic Principles

Replication:

is an independent repeat run of each factor combination. In the metallurgical experiment discussed, replication would consist of treating a specimen by oil quenching and treating a specimen by saltwater quenching. Thus, if five specimens are treated in each quenching medium, we say that five replicates have been obtained.

- Each of the 10 observations should be run in random order.
- It allows the experimenter to obtain an estimate of the experimental error.
- If the sample mean is used to estimate the true mean response for one of the factor levels in the experiment, replication permits the experimenter to obtain a more precise estimate of this parameter.

Basic Principles

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**.

For example, an experiment in a chemical process may require two batches of raw material to make all the required runs. However, there could be differences between the batches due to supplier-to-supplier variability, and if we are not specifically interested in this effect, we would think of the batches of raw material as a nuisance factor. Generally, a block is a set of relatively homogeneous experimental conditions. In the chemical process example, **each batch of raw material would form a block**, because the variability within a batch would be expected to be smaller than the variability between batches

Guidelines for Designing an Experiment

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

Basic concepts of Statistics

We briefly review the following concepts:

- Hypothesis testing
- Probability distributions
- Sampling distributions (Normal, t, χ^2,F)
- Expected values and their properties
- Comparing two groups (t-tests)

See lecture notes!

An Example: Etching process

- An engineer is interested in investigating the relationship between the *Radio Frequency (RF) power* setting and the *etch rate* for a plasma etching tool.
- The objective of an experiment like this is to model the relationship between etch rate and RF power, and to specify the power setting that will give a desired target etch rate.
- The engineer wants to test four levels of RF power: 160, 180, 200, and 220 W. She decided to test five wafers at each level of RF power.

This is an example of a **single-factor experiment** with $a = 4$ levels of the factor (RF Power) and $n = 5$ replicates – this gives us 20 observations or runs.

An Example: Etching process

Randomization

These 20 runs should be made in random order.

Suppose we use a statistical software to randomize. What this means is that we have the 20 runs we want

160	160	160	160	160	180	180	180	180	180
200	200	200	200	200	220	220	220	220	220

and then use the software to reorder them so that we are not performing tests in order.
As an example of randomization,

200	220	220	160	160	180	200	160	180	200
220	220	160	160	220	180	180	180	200	200

An Example: Etching process

Why is this important?

Randomization helps account for unforeseen circumstances.
Suppose the tool has a warming up period – by randomizing, the warm up period won't consist of an entire level (i.e., all 5 runs of 160).
If there was a warm up period and we did not randomize, the data (or any inferences made with it) would not be valid as it would not control for the warm up period.

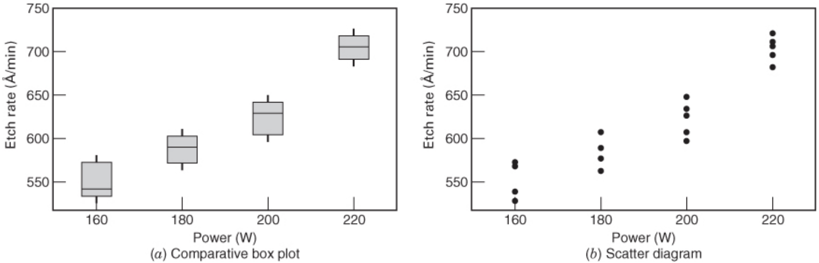
An Example: Etching process

The following data is resulting from the engineer's experiment

	Observations						
Power (W)	**1**	**2**	**3**	**4**	**5**	**Total**	**Average**
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

An Example: Etching process

We should first visualize teh data:



- Both graphs indicate that etch rate increases as the power setting increases.
- There is little evidence to suggest that the variability in etch rate around the average depends on the power setting.
- Based on the graphs, we believe that (1) RF power setting affects the etch rate and (2) higher power settings result in increased etch rate.

An Example: Etching process

- While graphs are nice at helping us visualize the data, **they do not provide any concrete evidence** that the trend we are seeing is "legitimate."
- We will need to perform a hypothesis test to determine if there's a difference between the levels of RF power. That is, we want to **test the equality of four means**.

One possibility is to perform multiple t-tests for all (4 choose 2) six possible pairs of means. However, this is not the best solution to this problem.

BECAUSE

1) performing all six pairwise t-tests is inefficient. 2) conducting all these pairwise comparisons inflates the type I error.

Experiments with a Single Factor

The **an**alysis **o**f **va**riance (ANOVA) allows us to compare more than two means. We can actually use it to compare two means (and will get the same result as a t-test!).

This course focuses on different ways to construct the ANOVA to account for the different factors in a variety of designs.

Experiments with a Single Factor

Suppose we have a treatments (or different levels) of a single factor that we wish to compare. The observed response from each of the a treatments is a random variable. The data would appear as in the table below.

Treatment (Level)	Observations				Total	Average
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a.}$	$\bar{y}_{a.}$

- i represents the factor level (or treatment)
- j represents the observation (or subject) number
- y_{ij} represents the j^{th} observation taken under factor level or treatment i.

We have n observations under the i^{th} treatment. Consider a model for the data.

The means model

We can write as follows:

$$y_{ij} = \mu_i + \varepsilon_{ij} \tag{1}$$

where

- y_{ij} is the ij^{th} observation,
- μ_i is the mean of the i^{th} factor level (or treatment),
- ε_{ij} is a random error component,
- $i = 1, 2, \ldots, a$, and $j = 1, 2, \ldots, n$.

Experiments with a Single Factor

The random error component represents error from “other sources" like measurement, variability arising from uncontrolled (or unmeasured) factors, differences between the experimental units to which the treatments are applied, and the background noise in the process.

Random Error ε_{ij}

We have:

$$E[\varepsilon_{ij}] = 0 \tag{2}$$

which implies the $E[y_{ij}] = \mu_i$

$$E[y_{ij}] = E[\mu_i + \varepsilon_{ij}] = \mu_i \tag{3}$$

The effects model

Now, if we rewrite the means as follows:

$$\mu_i = \mu + \tau_i \tag{4}$$

then

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{5}$$

- y_{ij} is the ij^{th} observation,
- μ is the overall mean, a parameter common to all treatments
- τ_i is the i^{th} treatment effect

Experiments with a Single Factor

Remarks

- This model is also called the one-way or single-factor analysis of variance (ANOVA) model because only one factor is investigated.
- We require that the experiment should be performed in **random order** so that the environment in which the treatments are applied (often called the experimental units) is as uniform as possible.
- The experimental design is a **completely randomized design**.
- The objective is **to estimate and then test appropriate hypotheses about the treatment means**.

Experiments with a Single Factor

Hypothesis testing

- The model errors are assumed to be normally and independently distributed random variables with mean zero and variance σ^2. $\varepsilon_{ij} \sim IIDN(0, \sigma^2)$
- The variance is assumed to be constant for all levels of the factor, implying that the observations

$$y_{ij} \sim N(\mu + \tau_i, \sigma^2) \tag{6}$$

and that the observations are mutually independent.

Experiments with a Single Factor

The effects model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ can fall under two situations:

1. Fixed Effects

- The a treatments could have been specifically chosen by the experimenter.
- Test hypotheses about the treatment means.
- Conclusions will **only apply to the factor levels considered** in the analysis. We cannot extend to similar treatments that were not explicitly considered.

2. Random Effects

- The a treatments could be a **random sample** from a larger population of treatments.
- Extend conclusions to **all** treatments in the population, even if they were not explicitly considered in our analysis.
- The τ_i are random variables.
- Test hypotheses about **the variability** of the τ_i and want to estimate this variability.

Notations

- Recall that $y_{i.}$ represents the total of the observations under the i^{th} treatment, that is $y_{i.} = \sum_{j=1}^{n} y_{ij}$
- $\bar{y}_{i.}$ represents the average of the observations under the i^{th} treatment, that is
$$\bar{y}_{i.} = \frac{y_{i.}}{n} \tag{7}$$
- $y_{..}$ represents the grand total of all the observations, that is
$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \tag{8}$$
- $\bar{y}_{..}$ represents the grand average of all the observations, that is
$$\bar{y}_{..} = \frac{y_{..}}{N} \tag{9}$$

ANOVA - Fixed Effects Model

Hypothesis testing

Now, we are interested in testing the equality of the treatment means. We write the hypotheses as follows:

$$\begin{aligned} H_0 :& \mu_1 = \mu_2 = \ldots = \mu_a \\ H_1 :& \mu_i \neq \mu_j \text{ for at least one pair } (i, j) \end{aligned} \tag{10}$$

In the effects model, $\mu_i = \mu + \tau_i$. The overall mean $\mu = \frac{\sum_{i=1}^{a}}{a}$ implies that $\sum_{i=1}^{a} \tau_i = \sum_{i=1}^{a} (\mu_i - \mu) = a\mu - a\mu = 0$.

Equivalent Hypothesis testing

Testing the all effects are zeros. We write the hypotheses as follows:

$$\begin{aligned} H_0 :& \tau_1 = \tau_2 = \ldots = \tau_a = 0 \\ H_1 :& \tau_i \neq 0 \text{ for at least one } i \end{aligned} \tag{11}$$

ANOVA - Fixed Effects Model

The reason we call this an analysis of variance is because we are **partitioning total variability into different components**.

Decomposition of the Total Sum of Squares

The corrected total sum of squares is given by:

$$\mathrm{SS_T} = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 \tag{12}$$

The $\mathrm{SS_T}$ measures the overall variability in the data, looks like the sample variance if divided by the appropriate degrees of freedom.

Let's decompose the $\mathrm{SS_T}$. See lecture notes.

ANOVA - Fixed Effects Model

This decomposition is **the fundamental ANOVA identity**:

$$SS_{\text{T}} = SS_{\text{Trt}} + SS_{\text{E}} \tag{13}$$

- We are saying that the total variability observed in the data can be partitioned into a *sum of squares due to treatments* (i.e., this is "between" treatments) and a *sum of squares due to error* (i.e., this is "within" treatments).
- There are $an = N$ observations, so SS_T has $\text{df}_\text{T} = N - 1$.
- There are a levels of the factor (a means), so SS_Trt has $\text{df}_\text{Trt} = a - 1$.
- Finally, there are n replicates within each treatment, and a treatments, so SS_E has $\text{df}_\text{E} = a(n-1) = an - a = N - a$

ANOVA - Fixed Effects Model

SS_E and SS_{Trt}

- SS_E is a pooled estimate of the common variance σ^2 within each of the a treatments.
- If the a treatment means are equal, then SS_{Trt} is also an estimate of σ^2.
- The ANOVA identity provides us with **two estimates of σ^2**.
- The intuition: If there are no difference between the treatment means then the SS_E and SS_{Trt} should give similar results, otherwise we can conclude that the observed difference is caused by the differences in the treatment means.

ANOVA - Fixed Effects Model

SS_E and SS_{Trt}

Let's define the mean squares as follows:

$$MS_E = \frac{SS_E}{N-a} \qquad \text{and} \qquad MS_{Trt} = \frac{SS_{Trt}}{a-1} \tag{14}$$

and

$$E[MS_E] = \sigma^2 \qquad \text{and} \qquad E[MS_{Trt}] = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1} \tag{15}$$

Remark

- Note that if the treatment means do differ, the expected value of the treatment mean square is greater than σ^2.
- It seems reasonable to test the hypothesis of no difference in Trt means by comparing MS_{Trt} and MS_E

ANOVA - Fixed Effects Model

Statistical Analysis

The hypothesis test is as follows:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0 \tag{16}$$
$$H_1 : \tau_i \neq 0 \text{ for at least one } i$$

We assumed that $\varepsilon_{ij} \sim IIDN(0, \sigma^2)$ and under the null hypothesis we have:

- $\mathrm{SS}_{\mathrm{Trt}}/\sigma^2 \sim \chi^2_{a-1}$
- $\mathrm{SS}_{\mathrm{E}}/\sigma^2 \sim \chi^2_{N-a}$
- $\mathrm{SS}_{\mathrm{T}}/\sigma^2 \sim \chi^2_{N-1}$

All the three sums of squares may not necessarily be independent as $SS_{\mathrm{T}} = SS_{\mathrm{Trt}} + SS_{\mathrm{E}}$!

ANOVA - Fixed Effects Model

Cochran's Theorem - results

Let Z_i be independently distributed as $N(0,1)$ for $i = 1, 2, \ldots, v$ and

$$\sum_{i=1}^{v} Z_i^2 = Q_1 + Q_2 + \cdots + Q_s,$$

where $s \leq v$ and Q_i has v_i degrees of freedom $(i = 1, 2, \ldots, s)$. Then $Q_1, Q_2, \cdots, Q_s$ are independent χ^2 random variables with $v_1, v_2, \cdots v_s$ degrees of freedom, respectively, if and only if

$$v = v_1 + v_2 + \cdots + v_s.$$

Cochran's Theorem - results

Because $\text{df}_{\text{Trt}} + \text{df}_{\text{E}} = \text{df}_{\text{T}}$, the theorem implies that $\text{SS}_{\text{Trt}}/\sigma^2$ and $\text{SS}_{\text{E}}/\sigma^2$ are independent χ^2 random variables.

Thus, the ratio,

$$F_0 = \frac{\text{SS}_{\text{Trt}}/(a-1)}{\text{SS}_{\text{E}}/(N-a)} = \frac{\text{MS}_{\text{Trt}}}{\text{MS}_{\text{E}}}$$

is distributed as an $F_{a-1,N-a}$.

This F_0 is the test statistic for the hypothesis of no differences in treatment means.

When testing this hypothesis, we have an upper-tail (i.e., one-tailed) rejection region and we reject if $F_0 > F_{\alpha,a-1,N-a}$.

ANOVA TABLE - Fixed Effects Model

Let's put all of this into what we call an ANOVA table.

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**Test Statis**
Between Trt.	$SS_{Trt} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$df_{Trt} = a - 1$	MS_{Trt}	F_0
Error (within Trt.)	$SS_E = SS_T - SS_{Trt}$	$df_E = N - a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$df_T = N - 1$		

Etch Rate Data in R.

ANOVA - Fixed Effects Model

Estimation of the model parameters

Recall that the single-factor model is given by

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{17}$$

Further, we have "reasonable" estimates of the overall mean and the treatment effects are given by

$$\hat{\mu} = \bar{y}_{..} \qquad \text{and} \qquad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..},$$

and from those two, we know that the estimated treatment mean is given by

$$\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.},$$

where $i = 1, 2, \ldots, a$.

ANOVA - Fixed Effects Model

Estimation of the model parameters

Let's now discuss confidence intervals. Definition: $100(1-\alpha)\%$ confidence interval for the i^{th} treatment mean, μ_i

$$\left(\bar{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{\text{MS}_\text{E}}{n}}, \bar{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{\text{MS}_\text{E}}{n}}\right)$$

Definition: $100(1-\alpha)\%$ confidence interval for the difference between two treatment means, $\mu_i - \mu_j$

$$\left((\bar{y}_{i.} - \bar{y}_{j.}) - t_{\alpha/2,N-a}\sqrt{\frac{2\text{MS}_\text{E}}{n}}, (\bar{y}_{i.} - \bar{y}_{j.}) + t_{\alpha/2,N-a}\sqrt{\frac{2\text{MS}_\text{E}}{n}}\right)$$

Note that the confidence intervals defined above are considered one-at-a-time confidence intervals.

ANOVA - Fixed Effects Model

Simultaneous Confidence Intervals

- The $1 - \alpha$ confidence level only applies to one particular estimate.
- If we have r $100(1 - \alpha)\%$ intervals, the probability that the r intervals will simultaneously be correct is at least $1 - r\alpha$.
- Thus, we see that as the number of confidence intervals increases, the probability that all intervals will be correct begins decreasing (multiple testing)
- If we want to calculate several confidence intervals, we should apply a Bonferroni correction to the α to ensure we do not inflate the experiment-wise error rate.
- We do this by replacing the $\alpha/2$ we use in the critical value by $\alpha/2r$.
- By doing this, we will construct r confidence intervals with an overall confidence level of at least $100(1 - \alpha)\%$.

Unbalanced design

- In some experiments, the number of observations taken within each treatment may be different – this is called an unbalanced design.
- We must then used the modified versions of the sum of squares,

$$\mathrm{SS_T} = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}, \qquad \text{and} \qquad \mathrm{SS_{Trt}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

- Note that although there are methods available for the unbalanced design, we still prefer the balanced design if we can get it because:
  - First, the test statistic is relatively robust to small departures from the assumption of equal variances for the a treatments if the sample sizes are equal.
  - Second, the power of the test is maximized when we have equal sample sizes.

Model Adequacy

Our model is $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ and recall that ε_{ij} is the term for random error.

- We assume that $\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$.
- If this assumption holds with the data we're analyzing, then the ANOVA is an exact test of the hypothesis of no difference in treatment means.

Definition: residual for observation j in treatment i as follows

$$e_{ij} = y_{ij} - \hat{y}_{ij},$$

where $\hat{y}_{ij}$ is the estimate of the corresponding observation y_{ij},

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) = \bar{y}_{i.}$$

Thus, the residual tells us how far away an observation is from its treatment mean.

ANOVA - Fixed Effects Model

Model Adequacy

- **Normality Assumption**
  - Look at Q-Q plot or histogram
  - Test normality: Shapiro-Wilk test
  - The appearance of a moderate departure from normality does not necessarily imply a serious violation of the assumptions.
  - Large deviations from normality are potentially serious and require further analysis.
- **Outliers:** Examine the standardized residuals $\sim N(0,1)$
- **Residuals vs. fitted values:** If the model is correct and the assumptions are satisfied, the residuals should be "structureless."
- **Constant variance:** examine Residuals vs. fitted values plots and also can use test for homogeneity of variances.

Shapiro-Wilk Test for Normality

Data The data consist of a random sample $X_1, X_2, X_3, ..., X_n$.

Hypothesis

$$H_0 : F(x) \text{ is normal with unspecified mean and variance}$$
$$H_1 : F(x) \text{ is nonnormal}$$

Test Statistic The order statistic is given as $X^{(1)}, X^{(2)}, X^{(3)}, ..., X^{(n)}$ from the smallest to the largest observation in the sample.

$$W = \frac{\left(\sum_{i=1}^{k} a_i (X^{(n-i+1)} - X^{(i)})\right)^2}{\sum_{i=1}^{n} (X_i - \overline{X})} \tag{18}$$

The quantiles of W can be found in tables of the Test or using R.

Decision Reject H_0 if $W > W_{1-\alpha}$

Bartlett Test for Equal variances

Hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_a^2$$
$$H_1 : \text{at least one is different}$$

Test Statistic

$$\chi_0^2 = 2.3026 \frac{q}{c} \quad (19)$$

where $q = (N-a)\log_{10}(s_p^2) - \sum_{i=1}^{a}(n_i - 1)\log_{10}(s_i^2)$; $c = 1 + \frac{\sum_{i=1}^{a}(n_i-1)^{-1} - (N-a)^{-1}}{3(a-1)}$; and $s_p^2 = \frac{\sum_{i=1}^{a}(n_i-1)s_i^2}{N-a}$ and s_i^2 is the sample variance of the i^{th} population.

Decision Reject H_0 if $\chi_0^2 > \chi_{\alpha,a-1}^2$

Bartlett Test assumes normality! Use Levene Test (robust) or Fligner-Killeen Test(nonparametric).

ANOVA- Model Adequacy

Variance-stabilizing transformation

A general class of variance-stabilizing transformations is given by Cox-Box transformation:

$$f_\lambda(X) = \begin{cases} \frac{X^\lambda - 1}{\lambda} & \text{if} \quad \lambda \geq 0 \\ \log(X) & \text{if} \lambda = 0, X > 0 \end{cases}$$

In practice λ is often 0 or 0.5.

R: We can use *powerTransform{car}* to estimate λ.

Multiple Comparisons Among Trt. Means

In ANOVA, we detect when there are differences between the treatment means.

- However, from ANOVA alone, we can't determine exactly which means differ.
- There may be times where further comparisons among groups of treatment means may be useful.
- We will discuss multiple comparisons where the goal is to compare pairs of treatment means.
- When using any procedure for pairwise testing of means, we occasionally find the overall F-test from the ANOVA is significant, but the pairwise comparison of means fails to reveal any significant differences. This is because F-test is simultaneously testing all possible contrasts.

Multiple Comparisons Among Trt. Means

Suppose that we are interested in comparing all pairs of a treatment means and that the null hypotheses that we wish to test are $H_0: \ \mu_i = \mu_j \ \forall \ i \neq j$.

Tukey's test

For equal sample sizes, Tukey's test declares two means significantly different if the absolute value of their sample differences exceeds

$$T_\alpha = q_\alpha(a, \mathrm{df_E})\sqrt{\frac{\mathrm{MS_E}}{n}}, \tag{20}$$

where a is the number of sample means. When **sample sizes are not equal**, we compare the absolute value of the sample differences to

$$T_\alpha = \frac{q_\alpha(a, \mathrm{df_E})}{\sqrt{2}}\sqrt{\mathrm{MS_E}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{21}$$

Multiple Comparisons Among Trt. Means

The Fisher Least Significant Difference (LSD) test

When we have equal group size $(n_1 = n_2 = \ldots = n_a = n)$,

$$\mathrm{LSD} = t_{\alpha/2,\mathrm{df}_\mathbf{E}} \sqrt{\frac{2\mathrm{MS_E}}{n}} \tag{22}$$

and when we do not have equal group size,

$$\mathrm{LSD} = t_{\alpha/2,\mathrm{df}_\mathbf{E}} \sqrt{\mathrm{MS_E}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \tag{23}$$

To use this procedure, we compare the observed difference between each pair of averages to the LSD.

If $|\bar{y}_{i.} - \bar{y}_{j.}| > \mathrm{LSD}$, we conclude that the population means differ.

Multiple Comparisons Among Trt. Means

Control for experiment-wise error rate

- For Tukey's test, the overall significance level is **exactly** α when the sample sizes are equal **at most** α when the sample sizes are not equal.
- However, The Fisher LSD method for comparing all pairs of means controls the error rate for each individual pairwise comparison but **does not control the experiment-wise or family error rate**.
- **How do we know which pairwise comparison method to use?**

Multiple Comparisons Among Trt. Means

Control for experiment-wise error rate

- For Tukey's test, the overall significance level is **exactly** α when the sample sizes are equal **at most** α when the sample sizes are not equal.
- However, The Fisher LSD method for comparing all pairs of means controls the error rate for each individual pairwise comparison but **does not control the experiment-wise or family error rate**.
- **How do we know which pairwise comparison method to use?** There is no clear answer to this – and everyone will answer it differently.

Comparing Treatment Means with a Control

In many experiments, we have a control group. Sometimes we are not interested in all pairwise comparisons, but only those that are comparing to the control group. We will be making $a-1$ comparisons. We can use **Dunnett's method** here, again using the differences between the sample means.

We reject $H_0 : \mu_i = \mu_c$ when

$$|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a-1, \mathrm{df_E})\sqrt{\mathrm{MS_E}\left(\frac{1}{n_i} + \frac{1}{n_c}\right)}, \qquad (24)$$

where the constant $d_\alpha(a-1, \mathrm{df_E})$ is given by Table VII.

We note that α is the joint significance level associated with all $a-1$ tests.

ANOVA - Random Effects

- We are often interested in a factor that has a large number of possible levels.
- If we randomly select a of the levels from the population of factor levels, then we will say that the factor is random.
- Because the levels were chosen randomly, we can make inference about the entire population of factor levels.

Our model is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \tag{25}$$

where $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, n$, and both the treatment effects (τ_i) and (ε_{ij}) are random variables.

- We assume that $\tau_i \sim N(0, \sigma_\tau^2)$ and,
- $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- Also, we assume that τ_i and ε_{ij} are independent.

ANOVA - Random Effects Model

Because τ_i is independent of ε_{ij}, the variance of any observation is

$$\text{Var}[y_{ij}] = \sigma_\tau^2 + \sigma^2 \tag{26}$$

- We call σ_τ^2 and σ^2 **variance components** and our model is called **the random effects model**.
- In the fixed effects model, *all* y_{ij} are independent.
- In the random effects model, y_{ij} are only independent if they are from different factor levels.

We can show that the covariance of any two observations is

$$\text{Cov}[y_{ij}, y_{ij'}] = \sigma_\tau^2, \; j \neq j'$$
$$\text{Cov}[y_{ij}, y_{i'j'}] = 0, \; i \neq i'$$

Observations that do not have the same factor level have covariance 0. Consider Example with $a = 3$ and $n = 2$ replicates.

ANOVA - Random Model

Our basic ANOVA sum of squares identity,

$$SS_{Tot} = SS_{Trt} + SS_E \tag{27}$$

is still valid. Testing hypotheses about individual treatment effects is no longer meaningful because they were selected randomly.

The variance component

We are interested in testing hypotheses about the variance component, σ_τ^2.

$$\begin{aligned} H_0 &: \sigma_\tau^2 = 0 \\ H_1 &: \sigma_\tau^2 > 0 \end{aligned} \tag{28}$$

- If $\sigma_\tau^2 = 0$, then all treatments are identical.
- However, if $\sigma_\tau^2 > 0$, we know that variability exists between treatments.

ANOVA - Random Model - Variance Component

Here, we want to estimate the σ^2 and σ_τ^2. We can show that the expected mean squares are as follows

$$E[\mathrm{MS_{Trt}}] = \sigma^2 + n\sigma_\tau^2 \tag{29}$$

$$E[\mathrm{MS_E}] = \sigma^2 \tag{30}$$

So, we estimate as follows:

$$\hat{\sigma}^2 = \mathrm{MS_E} \tag{31}$$

$$\hat{\sigma}_\tau^2 = \frac{\mathrm{MS_{Trt}} - \mathrm{MS_E}}{n} \tag{32}$$

Note that if we have unequal sample sizes, we replace n by

$$n_0 = \frac{1}{a-1}\left[\sum_{i=1}^{a} n_i - \frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}\right]$$

This is called a **method of moments procedure** to estimate σ^2 and σ_τ^2. We can also estimate using maximum likelihood.

ANOVA - Random Model - Variance Component

Confidence Intervals

Let us now discuss confidence intervals for our variance components.
$100(1-\alpha)\%$ CI for σ^2:

$$\frac{(N-a)\mathrm{MS_E}}{\chi^2_{\alpha/2,N-a}} \leq \sigma^2 \leq \frac{(N-a)\mathrm{MS_E}}{\chi^2_{1-\alpha/2,N-a}} \tag{33}$$

Note that we cannot compute an exact confidence interval for σ^2_τ – we do not have a closed-form expression for the appropriate distribution.

Instead, we can find an exact expression for a CI on the ratio

$$\frac{\sigma^2_\tau}{\sigma^2_\tau + \sigma^2}$$

this ratio is called the intraclass correlation coefficient (ICC) and reflects the proportion of the variance that is the result of differences between treatments.

ANOVA - Random Model - Variance Component

Confidence Intervals

$100(1-\alpha)\%$ CI for ICC:

$$\frac{L}{1+L} \leq \text{ICC} \leq \frac{U}{1+U}, \tag{34}$$

where

$$L = \frac{1}{n}\left(\frac{\text{MS}_{\text{Trt}}}{\text{MS}_{\text{E}}}\frac{1}{F_{\alpha/2,a-1,N-a}} - 1\right),$$

and

$$U = \frac{1}{n}\left(\frac{\text{MS}_{\text{Trt}}}{\text{MS}_{\text{E}}}\frac{1}{F_{1-\alpha/2,a-1,N-a}} - 1\right)$$

ANOVA - Random Model

Confidence Intervals of the Overall Mean, μ

In many random effects experiments, we are interested in estimating the overall mean μ. An unbiased estimator of the overall mean is

$$\hat{\mu} = \bar{y}_{..} \tag{35}$$

100(1 − α)% CI for μ

$$\bar{y}_{..} - t_{\alpha/2,a(n-1)}\sqrt{\frac{\mathrm{MS_{Trt}}}{an}} \leq \mu \leq \bar{y}_{..} + t_{\alpha/2,a(n-1)}\sqrt{\frac{\mathrm{MS_{Trt}}}{an}} \tag{36}$$

Estimating variance component using MLE

The method of moments to estimate the variance component has some disadvantages:

- It is a method of moments estimator – generally we do not prefer method of moments estimators (the parameter estimates do not have good properties).
- We also note that it does not lend itself to easy confidence interval construction (see: lack of CI for σ_τ^2, and we would **really** like to have a CI for that).
- In most software programs MLE is default.

ANOVA - Random Model

The method of maximum likelihood

Suppose x is a random variable with probability distribution $f(x, \theta)$, where θ is an unknown parameter. Let $x_1, x_2, \ldots, x_n$ be a random sample of n observations. The joint probability distribution of the sample is given by $\prod_{i=1}^{n} f(x_i, \theta)$. We can write the likelihood function as

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i, \theta) \tag{37}$$

The maximum likelihood estimator (MLE) of θ is the value of θ that maximizes the likelihood function $L(x_1, x_2, \ldots, x_n; \theta)$. MLE's have useful properties:

- For large samples, they are unbiased and have a normal distribution.
- The inverse of the matrix of second derivatives of the likelihood function (multiplied by -1) is the covariance matrix of the MLE's.
- This is important because it allows us to obtain CI's on the MLE's.

The residual maximum likelihood (REML)

REML is a variant of the MLE, known as the residual maximum likelihood method.

- **It is popular because it produces unbiased estimators and like MLE's, it allows us to easily find CI's.**
- If $\hat{\theta}$ is the MLE of θ and $\hat{\sigma}(\hat{\theta})$ is its estimated standard error, then the approximate $100(1-\alpha)\%$ CI on θ is as follows

$$\hat{\theta} - z_{\alpha/2}\hat{\sigma}(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2}\hat{\sigma}(\hat{\theta}) \tag{38}$$

Note that we can use this approach to find the CI for σ_τ^2.

Examples with R.

ANOVA - Single Factor

Examples with R/RStudio

- Etching Experiment
- Peak Discharge Data
- Rental Car
- Cardiovascular health and Chocolate
- Fabric strength and looms
- Vascular Graft Experiment

Randomized Complete Block Design - RCBD

When analyzing data from an experiment, variability caused by a nuisance factor can affect the results.

nuisance factor

A design factor that may have an effect on the response, but we are not interested in that effect.

- Sometimes a nuisance factor is unknown and uncontrolled. Randomization is the design technique used to guard against such a “lurking" nuisance factor. In other cases, the nuisance factor is known but uncontrollable.
- If we can at least observe the value that the nuisance factor takes on at each run of the experiment, we can then adjust for it.

Example - Hardness

We wish to determine whether or not four different tips produce different readings on a hardness testing machine.

- The machine operates by pressing the tip into a metal test coupon, and from the depth of the resulting depression, the hardness of the coupon can be determined.
- The experimenter has decided to obtain four observations on the hardness for each tip.

We would like to make the experimental error as small as possible; that is, we would like to remove the variability between coupons from the experimental error.

Randomized Complete Block Design - RCBD

Example - Hardness

A design that would accomplish this requires the experimenter to test each tip once on each of four coupons.

Test Coupon (Block)			
1	**2**	**3**	**4**
Tip 3	Tip 3	Tip 2	Tip 1
Tip 1	Tip 4	Tip 1	Tip 4
Tip 4	Tip 2	Tip 3	Tip 2
Tip 2	Tip 1	Tip 4	Tip 3

Randomized Complete Block Design - RCBD

Example - Hardness

This design is called a randomized complete block design (RCBD).

- The word “complete" indicates that each block (coupon) contains all the treatments (tips).
- By using this design, the blocks, or coupons, form a more homogeneous experimental unit on which to compare the tips.
- This design strategy improves the accuracy of the comparisons among tips by eliminating the variability among the coupons.

Within a block, the order in which the four tips are tested is randomly determined.

Randomized Complete Block Design - RCBD

Statistical Analysis

Suppose we have *a* treatments and *b* blocks.
The data resulting from the experiment can be shown as follows

Block 1	Block 2	$\cdots$	Block b
y_{11}	y_{12}	$\cdots$	y_{1b}
y_{21}	y_{22}	$\cdots$	y_{2b}
y_{31}	y_{32}	$\cdots$	y_{3b}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{a1}	y_{a2}	$\cdots$	y_{ab}

There is one observation per treatment in each block, and the order in which the treatments are run within each block is determined randomly. Randomization is done within the block, and is not an overall randomization. **We can't randomize to blocks**.

Randomized Complete Block Design - RCBD

Statistical Analysis - Model

We have an effects model,

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \tag{39}$$

where

- $i = 1, 2, \ldots, a$ represents the treatments,
- $j = 1, 2, \ldots, b$ represents the blocks,
- μ is the overall mean,
- τ_i is the treatment effect for treatment i,
- β_j is the block effect for block j, and
- ε_{ij} is our usual $N(0, \sigma^2)$ random error term.

In general, we think of the treatment and block effects as deviations from the overall mean so that $\sum_{i=1}^{a} \tau_i =$ and $\sum_{j=1}^{b} \beta_j = 0$.

Randomized Complete Block Design - RCBD

Statistical Analysis - ANOVA identity

The total corrected sum of squares:

$$SS_T = SS_{Trt} + SS_{Blocks} + SS_E$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2$$

Nice problem!

Randomized Complete Block Design - RCBD

Statistical Analysis - ANOVA identity

The ANOVA table is then:

Source	SS	df	MS	F
Treatments	SS_{Trt}	df_{Trt}	MS_{Trt}	F_0
Blocks	SS_{Blocks}	df_{Blocks}	MS_{Blocks}	
Error	SS_E	df_E	MS_E	
Total	SS_T	df_T		

Example: Vascular Grafts.

Randomized Complete Block Design - RCBD

Example - Vascular Grafts

- A medical device manufacturer produces vascular grafts (artificial veins).
- These grafts are produced by extruding billets of polytetrafluoroethylene (PTFE) resin combined with a lubricant into tubes.
- Some of the tubes in a production run contain defects are known as “flicks."
- The product developer responsible for the vascular grafts suspects that the extrusion pressure affects the occurrence of flicks and therefore intends to conduct an experiment to investigate this hypothesis.
- The resin is manufactured by an external supplier and is delivered to the medical device manufacturer in batches.
- The engineer also suspects that there may be significant batch-to-batch variation.
- Therefore, the product developer decides to investigate the effect of **four different levels of extrusion pressure** on flicks using a randomized complete block design considering **batches of resin as blocks**.

Randomized Complete Block Design - RCBD

Example - Vascular Grafts

The RCBD is shown in the table below.

Extrusion Pressure (PSI)	Batch of Resin (Block) 1	2	3	4	5	6	Treatment Total
8500	90.3	89.2	98.2	93.9	87.4	97.9	556.9
8700	92.5	89.5	90.6	94.7	87.0	95.8	550.1
8900	85.5	90.8	89.6	86.2	88.0	93.4	533.5
9100	82.5	89.5	85.6	97.4	78.9	90.7	514.6
Block Totals	350.8	359.0	364.0	362.2	341.3	377.8	2155.1

Good exercise to find the ANOVA table!

Randomized Complete Block Design - RCBD

Additivity

The linear model we have used for the randomized block design is additive:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \tag{40}$$

- Suppose $\tau_1 = 5$ (treatment effect) and $\beta_1 = 2$ (block effect), Then the expected increase in both treatment 1 and block 1 (together) is

$$\mathrm{E}[y_{11}] = \mu + \tau_1 + \beta_1 = \mu + 5 + 2 = \mu + 7$$

- Although this model is useful, there are times where it's inadequate. Suppose we are looking at 4 formulations of a product in 6 batches of raw material (and we consider the batches as blocks). Suppose further that we have batch 2 affect formulation 2 such that it gives an unusually low yield, however, batch 2 does not affect other formulations. **This is an interaction**. An interaction is where the level of one factor affects the relationship between another factor and the outcome. We should use a factorial design – we will see this later.

Randomized Complete Block Design - RCBD

Random Treatment and Blocks

There are situations where either treatments or blocks (or both) are random factors. It's common for blocks to be random. Recall that if the blocks are random, our conclusions will be valid across all populations of blocks – not only the ones used in our experiment. Our model is still

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

however, we now assume

$$\beta_j \sim N(0, \sigma^2_\beta)$$

where $j = 1, 2, \ldots, b$. That is, our β_j are now random variables. This model is called a mixed model (because it contains both fixed and random factors). If our blocks are random and the treatments are fixed, we can show that

$$\begin{aligned} \mathrm{E}[y_{ij}] &= \mu + \tau_i, i = 1, 2, \ldots, a \text{ and } j = 1, 2, \ldots, b \\ \mathrm{Var}[y_{ij}] &= \sigma^2_\beta + \sigma^2 \\ \mathrm{Cov}[y_{ij}, y_{i'j'}] &= 0, i \neq i' \text{ and } j \neq j' \\ \mathrm{Cov}[y_{ij}, y_{i'j}] &= \sigma^2_\beta, i \neq i' \end{aligned}$$

Latin Square Design

We first discussed the randomized complete block design as a design to reduce the residual error in an experiment by removing variability due to a known and controllable nuisance variable. **There are several other types of designs that utilize the blocking principle.**

Example Rocket Propellant in R

- Suppose that an experimenter is studying the effects of five different formulations of a rocket propellant used in aircrew escape systems on the observed burning rate.
- Each formulation is mixed from a batch of raw material that is only large enough for five formulations to be tested.
- Furthermore, the formulations are prepared by several operators, and there may be substantial differences in the skills and experience of the operators.
- Thus, it would seem that there are two nuisance factors to be "averaged out" in the design: batches of raw material and operators.

Latin Square Design

The appropriate design for this problem consists of testing each formulation exactly once in each batch of raw material and for each formulation to be prepared exactly once by each of five operators. The resulting design, shown below, is called a Latin square design.

Raw Material	Operators				
	1	2	3	4	5
1	A = 24	B = 20	C = 19	D = 24	E = 24
2	B = 17	C = 24	D = 30	E = 27	A = 36
3	C = 18	D = 38	E = 26	A = 27	B = 21
4	D = 26	E = 31	A = 26	B = 23	C = 22
5	E = 22	A = 30	B = 20	C = 29	D = 31

Latin Square Design

- The Latin square design is used to eliminate two nuisance sources of variability.
- That is, it systematically allows blocking in two directions.
- Thus, the rows and columns actually represent two restrictions on randomization.
- In general, a Latin square for p factors, or a $p \times p$ Latin square, is a square containing p rows and p columns.
- Each of the resulting p^2 cells contains one of the p letters that corresponds to the treatments, and each letter occurs once (and only once) in each row and column.

Latin Square Design

The statistical (effects) model

$$y_{ijk} = \mu + \alpha_i + \tau_j + \beta_k + \varepsilon_{ijk},$$

where $i = 1, 2, \ldots, p$ corresponds to the row,
$j = 1, 2, \ldots, p$ corresponds to the treatment,
$k = 1, 2, \ldots, p$ corresponds to the column,
y_{ijk} is the observation in the i^{th} row and k^{th} column for the j^{th} treatment,
μ is the overall mean, α_i is the i^{th} row effect, τ_j is the j^{th} treatment effect,
β_k is the k^{th} column effect, and ε_{ijk} is the random error.
The ANOVA table for the Latin Square design,

Source	**SS**	**df**	**MS**	F
Treatment	SS_{Trt}	df_{Trt}	MS_{Trt}	F_0
Rows	SS_{Rows}	df_{Rows}	MS_{Rows}	
Columns	SS_{Cols}	df_{Cols}	MS_{Cols}	
Error	SS_{E}	df_{E}	MS_{E}	
Total	SS_{T}	df_{T}		

Latin Square Design with Replication

Case 1 - same levels of of the row and column blocking factors are used in each replicate

The sum of squares are found as follows:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{N}; \qquad df_T = np^2 - 1$$

$$SS_{Trt} = \frac{1}{np}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}; \qquad df_{Trt} = p - 1$$

$$SS_{Rows} = \frac{1}{np}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{N}; \qquad df_{Rows} = p - 1$$

$$SS_{Cols} = \frac{1}{np}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}; \qquad df_{Cols} = p - 1$$

$$SS_{Reps} = \frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}; \qquad df_{Reps} = n - 1$$

$$SS_E = SS_T - SS_{Trt} - SS_{Rows} - SS_{Cols} - SS_{Reps}; \qquad df_T = np^2 - 1$$

Latin Square Design with Replication

Case 2 - New batches of raw material but the same operators are used in each replicate

The sum of squares are found as follows:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{N}; \qquad df_T = np^2 - 1$$

$$SS_{Trt} = \frac{1}{np}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}; \qquad df_{Trt} = p - 1$$

$$SS_{Rows} = \frac{1}{p}\sum_{l=1}^{n}\sum_{i=1}^{p} y_{i..l}^2 - \sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}; \qquad df_{Rows} = n(p-1)$$

$$SS_{Cols} = \frac{1}{np}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}; \qquad df_{Cols} = p - 1$$

$$SS_{Reps} = \frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}; \qquad df_{Reps} = n - 1$$

$$SS_E = SS_T - SS_{Trt} - SS_{Rows} - SS_{Cols} - SS_{Reps}; \qquad df_E = (p-1)(np-1)$$

Latin Square Design with Replication

Case 2 - New batches of raw material and new operators are used in each replicate

The sum of squares are found as follows:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{N}; \qquad df_T = np^2 - 1$$

$$SS_{Trt} = \frac{1}{np}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}; \qquad df_{Trt} = p - 1$$

$$SS_{Rows} = \frac{1}{p}\sum_{l=1}^{n}\sum_{i=1}^{p} y_{i..l}^2 - \sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}; \qquad df_{Rows} = n(p-1)$$

$$SS_{Cols} = \frac{1}{p}\sum_{l=1}^{n}\sum_{k=1}^{p} y_{..kl}^2 - \sum_{l=1}^{n}\frac{y_{...l}^2}{p^2}; \qquad df_{Cols} = n(p-1)$$

$$SS_{Reps} = \frac{1}{p^2}\sum_{l=1}^{n} y_{...l}^2 - \frac{y_{....}^2}{N}; \qquad df_{Reps} = n - 1$$

$$SS_E = SS_T - SS_{Trt} - SS_{Rows} - SS_{Cols} - SS_{Reps}; \qquad df_E = (p-1)[n(p-1) - 1]$$

The Graeco-Latin Square Design

This is an extension of the Latin Square Design.

How?

Suppose we have a $p \times p$ Latin square, and we are going to superimpose another $p \times p$ Latin square in which we denote the treatments by Greek letters. When superimposed properly, each Greek letter appears once (and only once) with each Latin letter, and this is called the Graeco-Latin square. In table form,

	Column			
Row	**1**	**2**	**3**	**4**
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

The Graeco-Latin square design can be used to control three sources of extraneous variability (i.e., to block in three directions).

The Graeco-Latin Square Design

Statistical Model

The statistical model for the Graeco-Latin square design is

$$y_{ijkl} = \mu + \theta_i + \tau_j + \omega_k + \Psi_l + \varepsilon_{ijkl} \tag{41}$$

where $i = 1, 2, \ldots, p$ corresponds to the row,
$j = 1, 2, \ldots, p$ corresponds to the Latin letter,
$k = 1, 2, \ldots, p$ corresponds to the Greek letter,
$l = 1, 2, \ldots, p$ corresponds to the column,
y_{ijkl} is the observation in row i and column l for Latin letter j and Greek letter k,
θ_i is the effect of the i^{th} row,
τ_j is the effect of Latin letter treatment j,
ω_k is the effect of Greek letter treatment k,
Ψ_l is the effect of column l, and
ε_{ijkl} is the $N(0, \sigma^2)$ random error component.

We have the following ANOVA table

Source	SS	df	MS	F
Treatment (Greek)	SS_{Greek}	df_{Greek}	MS_{Greek}	$F_{0_{\text{Greek}}}$
Treatment (Latin)	SS_{Latin}	df_{Latin}	MS_{Latin}	$F_{0_{\text{Latin}}}$
Rows	SS_{Rows}	df_{Rows}	MS_{Rows}	
Columns	SS_{Cols}	df_{Cols}	MS_{Cols}	
Error	SS_{E}	df_{E}	MS_{E}	
Total	SS_{T}	df_{T}		

We note that we now compute an F for each the Greek and Latin treatment factors, if they are of interest.
The critical value is $F_{\alpha,p-1,(p-3)(p-1)}$.

The Graeco-Latin Square Design

We compute the sum of squares as follows:

$$SS_T = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} y_{ijkl}^2 - \frac{y_{....}^2}{N}; \qquad df_T = p^2 - 1$$

$$SS_{Latin} = \frac{1}{p}\sum_{j=1}^{p} y_{.j..}^2 - \frac{y_{....}^2}{N}; \qquad df_{Latin} = p - 1$$

$$SS_{Greek} = \frac{1}{p}\sum_{k=1}^{p} y_{..k.}^2 - \frac{y_{....}^2}{N}; \qquad df_{Greek} = p - 1$$

$$SS_{Rows} = \frac{1}{p}\sum_{i=1}^{p} y_{i...}^2 - \frac{y_{....}^2}{p^2}; \qquad df_{Rows} = p - 1$$

$$SS_{Cols} = \frac{1}{p}\sum_{l=1}^{p} y_{...l}^2 - \frac{y_{....}^2}{N}; \qquad df_{Cols} = p - 1$$

$$SS_E = SS_T - SS_{Latin} - SS_{Greek} - SS_{Rows} - SS_{Cols}; \qquad df_E = (p-3)(p-1)$$

Balanced Incomplete Block Design

- In certain experiments using randomized block designs, we may not be able to run all the treatment combinations in each block.
- Situations like this usually occur because of shortages of experimental apparatus or facilities or the physical size of the block.
- For example, in the vascular graft experiment (Example 4.1), suppose that each batch of material is only large enough to accommodate testing three extrusion pressures. Therefore, each pressure cannot be tested in each batch.
- For this type of problem it is possible to use randomized block designs in which every treatment is not present in every block.

These designs are known as randomized incomplete block designs.

Balanced Incomplete Block Design

- When all treatment comparisons are equally important, the treatment combinations used in each block should be selected in a balanced manner, so that any pair of treatments occur together the same number of times as any other pair.

Thus, a balanced incomplete block design (BIBD) is an incomplete block design in which any two treatments appear together an equal number of times.

See Example in R: Reaction Time Experiment with 4 catalysts.

Balanced Incomplete Block Design

Statistical Model

- Consider a treatments and b blocks
- Each block contains k treatments (different)
- Each treatment occurs r times in the design
- Each pair of treatments appears together in $\lambda = \frac{r(k-1)}{a-1}$ blocks
- Example: a=3, b=3, k=2,r=2,$\lambda = 1$

The statistical model for the BIBD is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Extensive list of BIBDs can be found in Fisher and Yates (1963) and Cochran and Cox (1957).

Balanced Incomplete Block Design

ANOVA Table

We can still partition the total variability:

$$SS_{\text{Tot}} = SS_{\text{Trt(adj)}} + SS_{\text{Blocks}} + SS_E$$

Note that the SS_{Trt} is adjusted to separate the treatment and the block effects.

- $SS_{\text{Tot}} = \sum_{ij} y_{ij}^2 - \frac{y_{..}^2}{N}$
- $SS_{\text{Blocks}} = \frac{1}{k} \sum_j y_{.j}^2 - \frac{y_{..}^2}{N}$
- $SS_{\text{Trt(adj)}} = \frac{k \sum_i Q_i^2}{\lambda a}$; where Q_i is the adjusted total for the i^{th} treatment and $Q_i = y_{i.} - \frac{1}{k} \sum_j n_{ij} y_{.j}$ where $n_{ij} = 1$ if treatment i appears in block j and $n_{ij} = 0$ otherwise.

ANOVA table is put together in the usual way:

Source	SS	df	MS	F
Treatment (adjusted)	$SS_{\text{Trt(adj)}}$	$a-1$	$MS_{\text{Trt(adj)}}$	$F_{0_{\text{Trt(adj)}}}$
Blocks	SS_{Blocks}	$b-1$	MS_{Blocks}	
Error	SS_{Error}	$N-a-b+1$	MS_{Error}	
Total	SS_{Tot}	$N-1$		

Factorial Design

Many experiments require the study of the effects of two or more factors.

Definitions

- **Factorial design**: In each complete trial or replicate of the experiment all possible combinations of the levels of the factors are investigated.
- **Main Effect**: The change in response produced by a change in the level of the factor.
- **Interaction**: The difference in response between the levels of one factor is not the same at all levels of the other factors.

Factorial Design

The Two-Factor Factorial Design

The simplest types of factorial designs involve only two factors or sets of treatments.

- There are a levels of factor A and b levels of factor B, and these are arranged in a factorial design
- That is, each replicate of the experiment contains all ab treatment combinations.
- In general, there are n replicates.

Factorial Design

Example: Design a battery

- An engineer is designing a battery for use in a device that will be subjected to some extreme variations in temperature.
- The only design parameter that he can select at this point is the plate material for the battery, and he has three possible choices.
- When the device is manufactured and is shipped to the field, the engineer has no control over the temperature extremes that the device will encounter, and he knows from experience that temperature will probably affect the effective battery life.

Factorial Design

Example: Design a battery

- The engineer decides to test all three plate materials at three temperature levels – 15°F, 70°F, and 125°F – because these temperature levels are consistent with the product end-use environment.
- Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order.

Factorial Design

The experiment and the resulting observed battery life data are given in the table below.

Example: Design a battery

Material Type	Temperature					
	15°F		70°F		125°F	
1	130	155	34	40	20	70
	74	180	80	85	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

Factorial Design

The experiment and the resulting observed battery life data are given in the table below.

Example: Design a battery

In this problem the engineer wants to answer the following questions:

1. What effects do material type and temperature have on the life of the battery?
2. Is there a choice of material that would give uniformly long life regardless of temperature?

Some remarks here:

- It may be possible to find a material alternative that is not greatly affected by temperature.
- If this is so, the engineer can make the battery robust to temperature variation in the field.
- This is an example of using statistical experimental design for robust product design, a very important engineering problem.

Factorial Design

In general, a two-factor factorial experiment will appear as in the table below.

		Factor B			
		1	2	...	b
Factor A	1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$
	2	$y_{211}, y_{212}, \ldots, y_{21n}$	$y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{2b2}, \ldots, y_{2bn}$
	$\vdots$				
	a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$

Where y_{ijk} is the the k^{th} replicate for the i^{th} level of factor A and j^{th} level of factor B. The order in which the abn observations are taken is selected at random so that this design is a **completely randomized design**.

Factorial Design: Statistical Model

The effects model is written as

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \qquad (42)$$

- $i = 1, 2, \ldots, a$,
- $j = 1, 2, \ldots, b$,
- $k = 1, 2, \ldots, n$;
- μ is the overall mean effect,
- τ_i is the effect of the i^{th} level of the row factor A,
- β_j is the effect of the j^{th} level of column factor B,
- $(\tau\beta)_{ij}$ is the effect of the interaction between τ_i β_j,
- ε_{ijk} is the random error component.
- There are *abn* observations.

Factorial Design: Hypotheses

We are looking at two factors, A and B, and they are of equal interest.
If we are looking at row treatment effects,

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ H_1 &: \text{at least one } \tau_i \neq 0 \end{aligned} \tag{43}$$

If we are looking at column treatment effects,

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \cdots = \beta_b = 0 \\ H_1 &: \text{at least one } \beta_i \neq 0 \end{aligned} \tag{44}$$

We are also interested in the interaction,

$$\begin{aligned} H_0 &: (\tau\beta)_{ij} = 0 \ \forall i, j \\ H_1 &: \text{at least one } (\tau\beta)_{ij} \neq 0 \end{aligned} \tag{45}$$

Factorial Design: Notations

- Let $y_{i..}$ denote the total of all observations under the i^{th} level of factor A
- $y_{.j.}$ denote the total of all observations under the j^{th} level of all observations under the j^{th} level of factor B,
- $y_{ij.}$ denote the total of all observations in the ij^{th} cell,
- $y_{...}$ denote the grand total of all observations.

We define the following,

$$y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{i..} = \frac{y_{i..}}{bn} \qquad i = 1, 2, \ldots, a$$

$$y_{.j.} = \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{.j.} = \frac{y_{.j.}}{an} \qquad j = 1, 2, \ldots, b$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij.} = \frac{y_{ij.}}{n} \qquad \begin{array}{l} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{array}$$

$$y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{...} = \frac{y_{...}}{abn}$$

Factorial Design: Sums

Now, we write our total corrected sum of squares as follows,

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left[(\bar{y}_{i..}-\bar{y}_{...})+(\bar{y}_{.j.}-\bar{y}_{...})+(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})+(y_{ijk}-\bar{y}_{ij.})\right]^2$$

$$= bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.}-\bar{y}_{...})^2$$

$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...})^2$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij.})^2$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

Factorial Design: ANOVA Table

$$
\begin{array}{llll}
SS_A & = bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 & = & \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{abn} \\
SS_B & = an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 & = & \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{abn} \\
SS_{AB} & = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 & = & \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B \\
SS_E & = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2 & = & SS_T - \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} \\
\hline
SS_T & = & = & \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}
\end{array}
$$

The degrees of freedom are given as follows,

A	=	$a-1$
B	=	$b-1$
A×B	=	$(a-1)(b-1)$
Error	=	$ab(n-1)$
Total	=	$abn-1$

Factorial Design: ANOVA Table

Of course we create the mean squares by dividing the sum of squares by the appropriate degrees of freedom. This leads to the following ANOVA table,

Source	**SS**	**df**	**MS**	**F**
A treatments	SS_A	df_A	MS_A	$F_0 = \frac{MS_A}{MS_E}$
B treatments	SS_B	df_B	MS_B	$F_0 = \frac{MS_B}{MS_E}$
A×B (interaction)	SS_{AB}	df_{AB}	MS_{AB}	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	df_E	MS_E	
Total	SS_T	df_T		

Factorial Design: Example

Life (in hours) observed in the battery design

The table below presents the life (in hours) observed in the battery design example described earlier. The row and column totals are shown in the margins of the table; the circled numbers are the cell totals.

	Temperature (°F)									
Material	**15**			**70**			**125**			$y_{i..}$
1	130 74	155 180	**539**	24 80	40 75	**229**	20 82	70 58	**230**	998
2	150 159	188 126	**623**	136 106	122 115	**479**	25 58	70 45	**198**	1300
3	138 168	110 160	**576**	174 150	120 139	**583**	96 82	104 60	**342**	1501
$y_{.j.}$	1738			1291			770			$y_{...} = 3799$

Factorial Design: Example

ANOVA Table

Source	SS	df	MS	F	p-value
Material	10,683.72	2	5,341.86	7.91	0.0020
Temperature	39,118.72	2	19,559.36	28.97	< 0.0001
Interaction	9,614.78	4	2,403.44	3.56	0.0186
Error	18,230.75	27	675.21		
Total	77,646.97	35			

Factorial Design with one replicate

Sometimes we have a two-factor experiment with only one replicate. Our effects model becomes

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ij}$$

where $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$. Note that we just dropped the subscript k.

- Note that unless we can assume the interaction effect to be zero, we cannot test main effects. When we can assume that there is not an interaction effect (i.e., $(\tau\beta)_{ij} = 0 \forall\ i$ and j), our model becomes

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

- Note that Tukey developed an additivity test for an interaction in the case of one replicate.

The General Factorial Design

- The results for the two-factor factorial design may be extended to the general case where there are a levels of factor A, b levels of factor B, c levels of factor C, and so on, arranged in a factorial experiment.
- There will be $abc \dots n$ total observations if there are n replicates of the complete experiment.

The General Factorial Design: Example

A soft drink bottle

- A soft drink bottler is interested in obtaining more uniform fill heights in the bottles produced by his manufacturing process. The filling machine theoretically fills each bottle to the correct target height, but in practice, there is variation around this target, and the bottler would like to understand the sources of this variability better and eventually reduce it.
- The process engineer can control three variables during the filling process: the percent carbonation (A), the operating pressure in the filler (B), and the bottles produced per minute or the line speed (C). The pressure and speed are easy to control, but the percent carbonation is more difficult to control during actual manufacturing because it varies with product temperature. However, for purposes of an experiment, the engineer can control carbonation at three levels: 10, 12, and 14 percent. She chooses two levels for pressure (25 and 30 psi) and two levels for line speed (200 and 250 bpm).
- She decides to run two replicates of a factorial design in these three factors, with all 24 runs taken in random order. The response variable observed is the average deviation from the target fill height observed in a production run of bottles at each set of conditions. The data that resulted from this experiment are shown in the table below. Note: Positive deviations are fill heights above the target, whereas negative deviations are fill heights below the target.

The General Factorial Design: Example

Data of the example:

	Operating Pressure (B)				
	25 psi		30 psi		
	Line Speed (C)		Line Speed (C)		
Carbonation (A)	200	250	200	250	$y_{i\ldots}$
10	-3, -1 (-4)	-1, 0 (-1)	-1, 0 (-1)	1, 1 (2)	-4
12	0, 1 (1)	2, 1 (3)	2, 3 (5)	6, 5 (11)	20
14	5, 4 (9)	7, 6 (13)	7, 9 (16)	10, 11 (21)	59
B × C Totals, $y_{.jk.}$	6	15	20	34	$y_{\ldots.} = 75$
$y_{.j..}$	21		54		

The General Factorial Design: Example

Which results in the following ANOVA table

Source	SS	df	MS	F	p-value
Percentage of carbonation (A)	252.750	2	126.375	178.412	< 0.0001
Operating Pressure (B)	45.375	1	45.375	64.059	< 0.0001
Line speed (C)	22.042	1	22.042	31.118	0.0001
AB	5.250	2	2.625	3.706	0.0558
AC	0.583	2	0.292	0.412	0.6713
BC	1.042	1	1.042	1.471	0.2485
ABC	1.083	2	0.542	0.765	0.4867
Error	8.500	12	0.708		
Total	336.625	23			

Blocking in a Factorial Design

- We have discussed factorial designs in the context of a *completely randomized experiment*.
- Sometimes, it is not feasible or practical to completely randomize all of the runs in a factorial.
- Consider a factorial experiment with two factors (A and B) and n replicates. the linear statistical model for this design is

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

- Suppose that we are adding a blocking factor – **a single replicate of a complete factorial experiment is run within each block**.
- Now, our model becomes

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \boldsymbol{\delta_k} + \varepsilon_{ijk},$$

Blocking in a Factorial Design

Then, we have the following ANOVA table,

Source	SS	df	MS	F
Blocks	SS_{Blocks}	df_{Blocks}	MS_{Blocks}	
A	SS_A	df_A	MS_A	F_A
B	SS_B	df_B	MS_B	F_B
AB	SS_{AB}	df_{AB}	MS_{AB}	F_{AB}
Error	SS_E	df_E	MS_E	
Total	SS_T	df_T	MS_T	

Blocking in a Factorial Design

The sums of squares and degrees of freedom are as follows,

Source	SS	df
Blocks	$\frac{1}{ab}\sum_k y_{..k}^2 - \frac{y_{...}^2}{abn}$	$n-1$
A	$\frac{1}{bn}\sum_i y_{i..}^2 - \frac{y_{...}^2}{abn}$	$a-1$
B	$\frac{1}{an}\sum_j y_{.j.}^2 - \frac{y_{...}^2}{abn}$	$b-1$
AB	$\frac{1}{n}\sum_i \sum_j y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B$	$(a-1)(b-1)$
Error	SS_T - SS_{Blocks} - SS_A - SS_B - SS_{AB}	$(ab-1)(n-1)$
Total	$\sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{abn}$	$abn-1$

Blocking in a Factorial Design: Example

Detect targets on a radar scope

- An engineer is studying methods for improving the ability to detect targets on a radar scope. Two factors she considers to be important are the amount of background noise, or "ground clutter," on the scope and the type of filter placed over the screen.
- An experiment is designed using three levels of ground clutter and two filter types. We will consider these as fixed-type factors. The experiment is performed by randomly selecting a treatment combination (ground clutter level and filter type) and then introducing a signal representing the target into the scope. The intensity of this target is increased until the operator observes it. The intensity level at detection is then measured as the response variable.
- Because of operator availability, it is convenient to select an operator and keep him or her at the scope until all the necessary runs have been made. Furthermore, operators differ in their skill and ability to use the scope. Consequently, it seems logical to use the operators as blocks. Four operators are randomly selected. Once an operator is chosen, the order in which the six treatment combinations are run is randomly determined.

Blocking in a Factorial Design: Example

Detect targets on a radar scope

We have a 3×2 factorial experiment run in a randomized complete block. The data are shown in the table below.

Operators (blocks):	**1**		**2**		**3**		**4**	
Filter Type:	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
Ground Clutter:								
Low	90	86	96	84	100	92	92	81
Medium	102	87	106	90	105	97	96	80
High	114	93	112	91	108	95	98	83

Blocking in a Factorial Design: Example

Detect targets on a radar scope

Which leads us to the ANOVA table

Source	SS	df	MS	F	p-value
Ground clutter (G)	335.58	2	167.79	15.13	0.0003
Filter type (F)	1066.67	1	1066.67	96.19	< 0.0001
GF	77.08	2	38.54	3.48	0.0573
Blocks	402.17	3	134.06		
Error	166.33	15	11.09		
Total	2047.83	23			

The 2^k Factorial Design

- Factorial designs are widely used in experiments involving several factors where it is necessary to study the joint effect of the factors on a response.
- One of the most important of the factorial design cases is that of k factors, each at only **two levels**.
- These levels may be **quantitative**, such as two values of temperature, pressure, or time or they may be **qualitative**, such as two machines, two operators, the "high" and "low" levels of a factor, or perhaps the presence and absence of a factor.

The 2^k Factorial Design

- A complete replicate of such a design requires $2 \times 2 \times \ldots \times 2 = 2^k$ observations and is called a 2^k factorial design.
- This section focuses on this class of designs. Throughout this chapter, we assume that
  - the factors are fixed
  - the designs are completely randomized
  - the usual normality assumptions are satisfied

The 2^k Factorial Design

Why use 2^k?

- The 2^k design is particularly useful in the early stages of experimental work when many factors are likely to be investigated.
- It provides the smallest number of runs with which k factors can be studied in a complete factorial design
- Consequently, these designs are widely used in factor screening experiments.

The 2^2 Factorial Design

The 2^2 Design

- The first design in the 2^k series is one with only two factors, say A and B, each run at two levels.
- This design is called a 2^2 factorial design.
- The levels of the factors may be arbitrarily called “low” and “high.”

The 2^2 Factorial Design

Example: Catalyst in a Chemical Process

- Consider an investigation into the effect of the concentration of the reactant and the amount of the catalyst on the conversion (yield) in a chemical process.
- The objective of the experiment was to determine if adjustments to either of these two factors would increase the yield.
- Let the **reactant concentration be factor A** and let the two levels of interest be 15 and 25 percent.
- **The catalyst is factor B**, with the high level denoting the use of 2 pounds of the catalyst and the low level denoting the use of only 1 pound.
- The experiment is replicated three times, so there are 12 runs. The order in which the runs are made is random, so this is a completely randomized experiment.

The 2^2 Factorial Design

The data obtained are as follows:

Factor		Treatment	Replicate			
A	B	Combination	I	II	III	Total
$-$	$-$	A low, B low	28	25	27	$(1) = 80$
$+$	$-$	A high, B low	36	32	32	$a = 100$
$-$	$+$	A low, B high	18	19	23	$b = 60$
$+$	$+$	A high, B high	31	30	29	$ab = 90$

The 2^k Factorial Design

- In experiments involving 2^k designs, it is always important to examine **the magnitude and direction of the factor effects** to determine which variables are likely to be important.
- The ANOVA can generally be used to confirm this interpretation.
- Effect magnitude and direction should always be considered along with the ANOVA, because **the ANOVA alone does not convey this information**.

The 2^k Factorial Design

We now have a different way of writing down the treatment combinations.

Effects

- We will use the order (1), a, b, ab – this is called standard order.
- Note that (1), a, b, ab represent **the total of the response observation** at all n replicates taken under that specific treatment combination.

Effects

Let's look at the main effect of A: The effect of A at the low level of B is

$$\frac{a-(1)}{n}$$

while the effect of A at the high level of B is

$$\frac{ab-b}{n}$$

and averaging these two gives the main effect of A

$$A=\frac{[ab-b]+[a-(1)]}{2n}=\frac{ab+a-b-(1)}{2n}$$

Effects

Now let's look at the main effect of B: The effect of B at the low level of A is

$$\frac{b-(1)}{n}$$

while the effect of B at the high level of A is

$$\frac{ab-a}{n}$$

and averaging these two gives the main effect of B

$$B=\frac{[ab-a]+[b-(1)]}{2n}=\frac{ab-a+b-(1)}{2n}$$

The 2^k Factorial Design

The interaction effect, AB, is the average difference between the effect of A at the high level of B and the effect of A at the low level of B,

Effects

The effect of A at the low level of B is

$$\frac{a - (1)}{n}$$

while the effect of A at the high level of B is

$$\frac{ab - b}{n}$$

and averaging the difference of two gives the interaction AB,

$$AB = \frac{[ab - b] - [a - (1)]}{2n} = \frac{ab - a - b + (1)}{2n}$$

We could have looked at AB as the average difference between the effect of B at the high level of A and the effect of B at the low level of A – we would come up with the same equation.

The 2^k Factorial Design

Example: Catalyst in a Chemical Process

Back to our example, we estimate the average effects as

$$A = \frac{ab + a - b - (1)}{2n} = \frac{90 + 100 - 60 - 80}{2(3)} = 8.33$$

$$B = \frac{ab - a + b - (1)}{2n} = \frac{90 - 100 + 60 - 80}{2(3)} = -5.00$$

$$AB = \frac{ab - a - b + (1)}{2n} = \frac{90 - 100 - 60 + 80}{2(3)} = 1.67$$

The 2^k Factorial Design

Brief interpretations

The effect of A is positive – increasing A from the low level to the high level increases the yield.

The effect of B is negative – increasing B from the low level to the high level decreases the yield.

The interaction effect is small compared to the main effects.

When looking at 2^k designs, we should examine the magnitude and direction of factor effects. This helps us determine which variables are important.

This should also be considered along with ANOVA – note that ANOVA doesn't give us this information directly!

The 2^k Factorial Design

We use contrasts (and their coefficients) to estimate the effects A, B, and AB.

The coefficients we use for the contrasts in the 2^2 design,

Treatment	**Factorial Effect**			
Combination	**I**	**A**	**B**	**AB**
(1)	+	−	−	+
a	+	+	−	−
b	+	−	+	−
ab	+	+	+	+

We call this orthogonal/effects coding.

So, from the table above,

$$\begin{array}{ll} A & -(1) + a - b + ab \\ B & -(1) - a + b + ab \\ AB & +(1) - a - b + ab \end{array}$$

The 2^k Factorial Design

Applying this to our example,

$$\begin{array}{ll} A & -80 + 100 - 60 + 90 = 50 \\ B & -80 - 100 + 60 + 90 = -30 \\ AB & +80 - 100 - 60 + 90 = 10 \end{array}$$

Now, we can use the contrasts to compute the sums of squares.

Source	**SS**	**df**
A	$\frac{[ab+a-b-(1)]^2}{4n}$	1
B	$\frac{[ab+b-a-(1)]^2}{4n}$	1
AB	$\frac{[ab+(1)-a-b]^2}{4n}$	1
E	SS_T - SS_A - SS_B - SS_{AB}	$4(n-1)$
T	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{4n}$	$4n-1$

where a, b, and ab represent the total of the response observations at all n replicates taken at the treatment combination.

The 2^k Factorial Design

The Regression Model

In a 2^k factorial design, we can express the results of our experiment in terms of a regression model.

Note that we could use either an effects or a means model, but we prefer regression models.

For our example, the model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

where x_1 is a coded variable that represents the reactant concentration, x_2 is a coded variable that represents the amount of catalyst, and the β's are regression coefficients.

The 2^k Factorial Design

Note that how we calculate β depends on how x is coded!

- The textbook uses effect coding. This means that one level will be coded as –1 and the other level will be coded as +1. The estimates of β as one-half the effect estimate.
- There is also reference coding where x is coded 0, 1.

$$y = overallMean + EffectA/2x_1 + EffectB/2x_2 + EffectAB/2x_1x_2$$

Thus, in our example,

$$\hat{y} = 27.5 + \frac{8.33}{2}x_1 - \frac{5}{2}x_2$$

The 2^k Factorial Design

The 2^3 Design

Suppose now we have three factors, A, B, and C, each at two levels.

Run	A	B	C	Labels	A	B	C
1	−	−	−	(1)	0	0	0
2	+	−	−	a	1	0	0
3	−	+	−	b	0	1	0
4	+	+	−	ab	1	1	0
5	−	−	+	c	0	0	1
6	+	−	+	ac	1	0	1
7	−	+	+	bc	0	1	1
8	+	+	+	abc	1	1	1

The 2^k Factorial Design

The average effect of A is as follows

$$A = \frac{[a - (1) + ab - b + ac - c + abc - bc]}{4n}$$

The average effect of B,

$$B = \frac{[b + ab + bc + abc - (1) - a - c - ac]}{4n}$$

and the average effect of C,

$$C = \frac{[c + ac + bc + abc - (1) - a - b - ab]}{4n}$$

Then, the interactions are given as follows

$$AB = \frac{[abc - bc + ab - b - ac + c - a + (1)]}{4n}$$

$$AC = \frac{[(1) - a + b - ab - c + ac - bc + abc]}{4n}$$

$$BC = \frac{[(1) + a - b - ab - c - ac + bc + abc]}{4n}$$

Finally, the three way interaction,

$$ABC = \frac{[abc - bc - ac + c - ab + b + a - (1)]}{4n}$$

The 2^k Factorial Design

We can actually represent these things with a table of + and – signs.

Treatment	**Factorial Effect**							
Combination	**I**	**A**	**B**	**AB**	**C**	**AC**	**BC**	**ABC**
(1)	+	−	−	+	−	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	+	−	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	−	+	+	−	−
bc	+	−	+	−	+	−	+	−
abc	+	+	+	+	+	+	+	+

We can use the contrasts to compute the SS. In the 2^3 design with n replicates, the SS for any effect is

$$\text{SS} = \frac{\text{contrast}^2}{8n}$$

We compute the total SS using the formula

$$\text{SS}_{\text{Tot}} = \sum_{ijkl} y_{ijkl}^2 - \frac{y_{....}^2}{4n}$$

and the error SS by subtraction

$$\text{SS}_{\text{E}} = \text{SS}_{\text{T}} - \text{SS}_{\text{A}} - \text{SS}_{\text{B}} - \text{SS}_{\text{C}} - \text{SS}_{\text{AB}} - \text{SS}_{\text{BC}} - \text{SS}_{\text{AC}} - \text{SS}_{\text{ABC}}$$

Following the format of the last section,

$$\text{df}_{\text{E}} = 8(n-1)$$

and

$$\text{df}_{\text{Tot}} = 8n - 1$$

The 2^k Factorial Design

Example: Etch process

A 2^3 factorial design was used to develop a nitride etch process on a single-wafer plasma etching tool. The design factors are the gap between the electrodes, the gas flow, and the RF power applied to the cathode. Each factor is run at two levels, and the design is replicated twice. The response variable is the etch rate for silicon nitride (Å/m). The etch rate data are shown in the table below.

	Coded Factors			Etch Rate		
Run	**A**	**B**	**C**	**Rep. 1**	**Rep. 2**	**Total**
1	−1	−1	−1	550	604	(1) = 1154
2	1	−1	−1	669	650	a = 1319
3	−1	1	−1	633	601	b = 1234
4	1	1	−1	642	635	ab = 1277
5	−1	−1	1	1037	1052	c = 2089
6	1	−1	1	749	868	ac = 1617
7	−1	1	1	1075	1063	bc = 2138
8	1	1	1	729	860	abc = 1589

The 2^k Factorial Design

We begin by constructing the contrasts.

$$
\begin{array}{llr}
A & -1154 + 1319 - 1234 + 1277 - 2089 + 1617 - 2138 + 1589 = & -813 \\
B & -1154 - 1319 + 1234 + 1277 - 2089 - 1617 + 2138 + 1589 = & 59 \\
C & -1154 - 1319 - 1234 - 1277 + 2089 + 1617 + 2138 + 1589 = & 2449 \\
AB & +1154 - 1319 - 1234 + 1277 + 2089 - 1617 - 2138 + 1589 = & -199 \\
AC & +1154 - 1319 + 1234 - 1277 - 2089 + 1617 - 2138 + 1589 = & -1229 \\
BC & +1154 + 1319 - 1234 - 1277 - 2089 - 1617 + 2138 + 1589 = & -17 \\
ABC & -1154 + 1319 + 1234 - 1277 + 2089 - 1617 - 2138 + 1589 = & 45
\end{array}
$$

We will first use the contrasts to find the average effects by dividing by 8 (note that the denominator is $4n$ and $n = 2$).

Then we will create our ANOVA table by converting the contrasts to sums of squares (we will divide by $8n = 16$).

The 2^k Factorial Design

First, the average effects.

$$A = \frac{-813}{8} = -101.625$$
$$B = \frac{59}{8} = 73.75$$
$$C = \frac{2449}{8} = 306.125$$
$$AB = \frac{-199}{8} = -24.875$$
$$AC = \frac{-1229}{8} = -153.625$$
$$BC = \frac{-17}{8} = -2.125$$
$$ABC = \frac{45}{8} = 5.625$$

We see that the largest effects are for power (C = 306.125), gap (A = –101.625), and the power-gap interaction (AC = –153.625).

The 2^k Factorial Design

Moving to the ANOVA table – recall that we divide the contrasts by $8n = 8 \times 2 = 16$ to get the sums of squares.

Source	SS	df	MS	F	p
Gap (A)	41,310.56	1	41,310.56	18.34	0.0027
Gas Flow (B)	217.56	1	217.56	0.10	0.7639
Power (C)	374,850.06	1	374,850.06	166.41	0.0001
AB	2475.06	1	2475.06	1.10	0.3252
AC	94,402.56	1	94,402.56	41.91	0.0002
BC	18.06	1	18.06	0.01	0.9308
ABC	126.56	1	126.56	0.06	0.8186
Error	18,020.50	8	2252.56		
Total	531,420.94	15			

We note that there is no three-way interaction, but there is a two-way interaction between gap and power.

The 2^k Factorial Design

We can estimate the regression model.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \hat{\beta}_{13} x_1 x_3$$
$$= 776.0625 + \left(\frac{-101.625}{2}\right) x_1 + \left(\frac{306.125}{2}\right) x_3 + \left(\frac{-153.625}{2}\right) x_1 x_3$$

The General 2^k Factorial Design

We will now generalize the method we have been discussing to the 2^k factorial design – we now have k factors each at two levels.

The statistical model will contain the following

- k main effects
- $\binom{k}{2}$ two-factor interactions
- $\binom{k}{3}$ three-factor interactions
- etc.
- and one k-factor interaction

Thus, the model will contain $2^k - 1$ effects for a 2^k design.

Standard order is always introducing the factors one a time, with each new factor combining with those that came before it. e.g., standard order for a 2^4 design,

(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, and abcd

The 2^k Factorial Design

The general approach for the 2^k design is as follows

1. Estimate factor effects.
2. Form initial model.
    1. If the design is replicated, fit the full model.
    2. If there is no replication, form the model using a normal probability plot of the effects.
3. Perform statistical testing.
4. Refine model.
5. Analyze residuals.
6. Interpret results.

We start with the full model, then work our way backwards. All sums of squares will have 1 degree of freedom other than the error term, which will have $2^k(n-1)$ degrees of freedom, and the total sums of squares, which will have $n2^k - 1$ degrees of freedom.

Blocking in the 2^k Factorial Design

- In many situations it is impossible to perform all of the runs in a 2^k factorial experiment under the same conditions.
- As an example, we've discussed using batches of raw materials. What if a single batch of raw material is not be large enough to make all of the required runs?
- As another example, a chemical engineer may run a pilot plant experiment with several batches of raw material because he knows that different raw material batches of different quality grades are likely to be used in the actual full-scale process.

The design technique used in these situations is blocking.

Blocking in the 2^k Factorial Design

Blocking a Replicated 2^k Factorial Design

- Suppose that the 2^k factorial design has been replicated n times.
- If there are n replicates, then each set of non-homogeneous conditions defines a block, and **each replicate is run in one of the blocks**.
- The runs in each block (or replicate) would be made in random order.

The analysis of the design is similar to that of any blocked factorial experiment.
See Example of the chemical process and catalysts.

Blocking in the 2^k Factorial Design

Example of the chemical process and catalysts

Factor A	Factor B	Treatment Combination	Replicate I	Replicate II	Replicate III	Total
−	−	A low, B low	28	25	27	(1) = 80
+	−	A high, B low	36	32	32	a = 100
−	+	A low, B high	18	19	23	b = 60
+	+	A high, B high	31	30	29	ab = 90

The table below shows the design, where each batch of raw material corresponds to a block.

	Block 1	Block 2	Block 3
	(1) = 28	(1) = 25	(1) = 27
	a = 36	a = 32	a = 32
	b = 18	b = 19	b = 23
	ab = 31	ab = 30	ab = 29
Block Totals:	$B_1 = 113$	$B_2 = 106$	$B_3 = 111$

Blocking in the 2^k Factorial Design

Example of the chemical process and catalysts

The blocking ANOVA table

Source	SS	df	MS	F	p-value
Blocks	6.50	2	3.25		
Concentration (A)	208.33	1	208.33	50.32	< 0.0001
Catalyst (B)	75.00	1	75.00	18.12	0.0053
Interaction (AB)	8.33	1	8.33	2.01	0.2060
Error	24.84	6	4.14		
Total	323.00	11			

And the ignoring-the-blocking ANOVA table

Source	SS	df	MS	F	p-value
Concentration (A)	208.33	1	208.33	53.19	0.0004
Catalyst (B)	75.00	1	75.00	19.15	0.0024
Interaction (AB)	8.33	1	8.33	2.13	0.1828
Error	31.33	8	3.92		
Total	323.00	11			

Confounding in the 2^k Factorial Design

In many problems it is impossible to perform a complete replicate of a factorial design in one block.

Definition

Confounding is a design technique for arranging a complete factorial experiment in blocks, where the block size is smaller than the number of treatment combinations in one replicate.
The technique causes information about certain treatment effects (usually high-order interactions) to be indistinguishable from, or confounded with, blocks.

Note that even though the designs presented are incomplete block designs because each block does not contain all the treatments or treatment combinations, the special structure of the 2^k factorial system allows a simplified method of analysis

Confounding in the 2^k Factorial Design

- Suppose that we wish to run a single replicate of the 2^2 design.
- Each of the $2^2 = 4$ treatment combinations requires a quantity of raw material, for example, and each batch of raw material is only large enough for two treatment combinations to be tested.
- Thus, two batches of raw material are required.
- If batches of raw material are considered as blocks, then we must assign two of the four treatment combinations to each block.

For example:

- block 1 contains (1) and ab
- block 2 contains a and b

The order in which the treatment combinations are run within a block is randomly determined. We would also randomly decide which block to run first.

Confounding in the 2^k Factorial Design

Suppose we estimate the main effects of A and B just as if no blocking had occurred.

$$A = \frac{ab + a - b - (1)}{2}$$
$$B = \frac{ab + b - a - (1)}{2}$$

Note that both A and B are unaffected by blocking because in each estimate there is one plus and one minus treatment combination from each block. **That is, any difference between block 1 and block 2 will cancel out.**
If we consider the AB interaction,

$$AB = \frac{ab + (1) - a - b}{2}$$

Because the two treatment combinations with the plus sign [ab and (1)] are in block 1 and the two with the minus sign (a and b) are in block 2, the block effect and the AB interaction are identical.

That is, **AB is confounded with blocks**.

Confounding in the 2^k Factorial Design

Remarks

- When the number of variables is small, say $k = 2$ or 3, it is usually necessary to replicate the experiment to obtain an estimate of error.
- If k is moderately large, say $k \geq 4$, we can frequently afford only a single replicate. The experimenter usually assumes higher order interactions to be negligible and combines their sums of squares as error.
- Unless experimenters have a prior estimate of error or are willing to assume certain interactions to be negligible, they must replicate the design to obtain an estimate of error.
- In 2^3 design, if the ABC interaction is confounded in each replicate then it cannot be retrieved. **This design is said to be completely confounded.**

Partial Confounding in the 2^k Factorial Design

Consider the alternative. Once again, there are four replicates of the 2^3 design, but a different interaction has been confounded in each replicate.

- ABC is confounded in replicate I,
- AB is confounded in replicate II,
- BC is confounded in replicate III,
- and AC is confounded in replicate IV.

Then,

- The information on ABC can be obtained from the data in replicates II, III, and IV;
- The information on AB can be obtained from replicates I, III, and IV;
- The information on AC can be obtained from replicates I, II, and III;
- The information on BC can be obtained from replicates I, II, and IV.

Partial Confounding in the 2^k Factorial Design

- We say that three-quarters information can be obtained on the interactions because they are unconfounded in only three replicates.
- This design is said to be *partially confounded*.
- When analyzing the partially confounded data, sums of squares are calculated using only data from the replicates where an interaction is unconfounded.

See Example in R.

References

CRAN Task View: Design of experiments (DoE) & Analysis of Experimental.
`https://cran.r-project.org/web/views/ExperimentalDesign.html`.
Accessed: 2021-8-6.

Douglas C Montgomery.
Design and analysis of experiments.
John wiley & sons, 2017.