

STATISTICS WORKSHOP VI

United States Department of Agriculture

*Experimental
Design II:
Two Sizes of
Experimental Units*

*Split Plot
&
Repeated Measures*

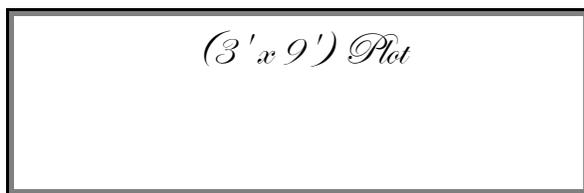
presented by
Vicki Lancaster
NEPTUNE AND CO., INC.
vlancast@neptuneandco.com

Goals

- C To *discuss* the elements of a good experimental design, *replication, randomization*, and the necessity for *homogeneous experimental units*.
- C To *review* the taxonomy of an experimental design, the *design, treatment, and error structures*.
- C To *review* the *analysis of variance table* and the concept of *degrees of freedom*.
- C To *introduce and define* the concept of *random and fixed effects*.
- C To *introduce* the taxonomy of the *split plot* and *repeated measures* designs and the concept of more than one size of experimental unit.

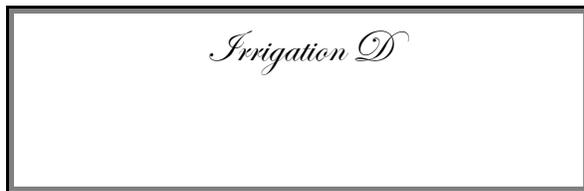
Motivation . . .

How do we analyze data from an experiment when one treatment level is randomly assigned to the entire experimental unit and other treatment levels are randomly assigned to partitions of the large EU?



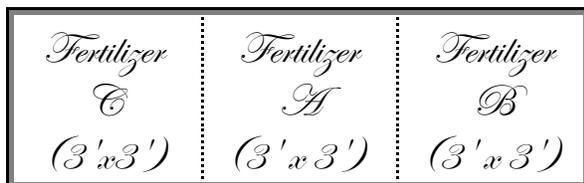
for each large EU

2



randomly assign a treatment level to the entire large EU

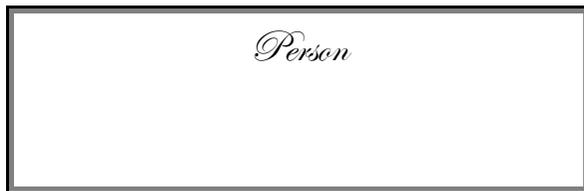
2



within each large EU randomly assign all levels of a second treatment

Motivation . . .

How do we analyze data from an experiment when the measurements on each experimental unit are taken at numerous points over time?



for each large EU

+



randomly assign a treatment level to the large EU

2



within each large EU take measurements at different points in time

Let start with a description of a SRRC investigation . . .

Suppose a SRRC researcher is interested in how the type of soil a melon is grown in and how the melon is processed influences the flavor and texture of the melon over time. In order to conduct the investigation, the researcher randomly selects four panelists to evaluate the sensory attributes of the melons from the various treatment combinations.

The researcher wants to evaluate two soil types and two preparation methods across four different storage times.

Two soil types:

1. sandy soil
2. clay soil

with two preparation methods:

1. no preparation
2. acid wash

and four storage times:

1. Day 0
2. Day 4
3. Day 7
4. Day 10.

The melons were randomly selected from each soil type location and sent to the SRRC facility. At the facility, the melons were culled for defects, washed, peeled, and cut into 2 to 3 cm x 2.5 cm cubes. Ten cubes from a particular soil type were randomly assigned to a juice catcher container and a container randomly assigned to one of the two preparation methods. After preparation, the juice catcher containers were randomly assigned to the shelves within a refrigerator and stored for 24 hours.

The researcher made the assumption that the conditions in the refrigerator were homogeneous, that is, the EUs would not be affected by where they were placed in the refrigerator.

When it came time for tasting, the containers were randomly assigned to the panelist such that each received the four treatment combinations. Two cubes were randomly removed from the container, and the order of presentation was randomly determined.

<i>Panelist 1</i>	<i>none / Sandy</i>	<i>acid / Clay</i>	<i>acid / Sandy</i>	<i>none / Clay</i>
<i>Panelist 2</i>	<i>acid / Sandy</i>	<i>none / Clay</i>	<i>acid / Clay</i>	<i>none / Sandy</i>
<i>Panelist 3</i>	<i>acid / Clay</i>	<i>none / Clay</i>	<i>acid / Sandy</i>	<i>none / Sandy</i>
<i>Panelist 4</i>	<i>acid / Clay</i>	<i>none / Clay</i>	<i>none / Sandy</i>	<i>acid / Sandy</i>



Each panelist is a Block. Each received all four of the treatment combinations. The treatment combinations were randomly assigned to each panelist.

For this investigation, the data matrix looks like:

EU	<i>Explanatory or Independent Variables</i>				<i>Response or Dependent Variables</i>			
	<i>Panelist</i>	<i>Soil</i>	<i>Prep.</i>	<i>Day</i>	<i>Sweetness</i>	<i>Wetness</i>	<i>...</i>	<i>Hardness</i>
1	Joe	clay	none	0				
2	Joe	sandy	none	0				
3	Joe	clay	acid	0				
4	Joe	sandy	acid	0				
1	Joe	clay	none	4				
2	Joe	sandy	none	4				
3	Joe	clay	acid	4				
4	Joe	sandy	acid	4				
1	Joe	clay	none	7				
2	Joe	sandy	none	7				
3	Joe	clay	acid	7				
4	Joe	sandy	acid	7				
⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮
16	Mary	sandy	acid	10				

This investigation will be used to help explain *repeated measures designs*.

Definition . . .

- C A *repeated measures design* is a combination of two complete experimental designs. Each experimental design has an its own experimental unit and error term, therefore when you combine them, there are two EUs and two error terms. A unique feature of a repeated measures design is that there is at least one treatment that cannot be randomly assigned to its experimental unit. In most cases this is time.

Before we start on experimental designs with more than one size of experimental unit, let's briefly review:

C *experimental design*, and

C *analysis of variance* (ANOVA).

An *experiment* imposes a *treatment* on an *experimental unit* in order to observe the *response variable*. The purpose of an experiment is to study whether the *treatment* causes a change in the *response variable*.

- C For example, in the SRRC investigation, the *treatments* (independent or explanatory variables) are the soil type, preparation method, and length of storage.
- C The *response variables* (dependent or response variables) are the nine flavor and five texture attributes.
- C The *experimental unit* (EU), the smallest unit to which a treatment level or treatment combination can be applied, is a juice catcher container.

An *experimental design* (ED) is concerned with

- C providing unambiguous information on the primary objectives of an experiment; and
- C providing the maximum amount of information with respect to the primary objectives per minimal amount of experimental effort.

How is this accomplished?

Through the three components of experimental design the *treatment*, the *design* and the *error* structures.

Definition . . .



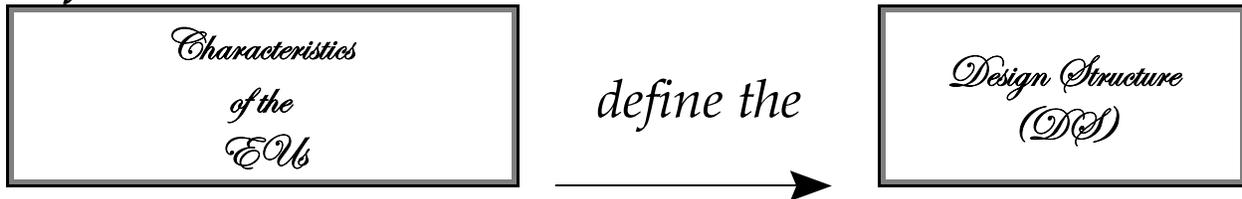
The *treatment structure* (TS) of an experimental design consists of the set of treatment (condition) levels, treatment combinations, or populations that the researcher has selected to study and or/compare.

Example:

PO: To study five diets on losing weight.

TS: One-way ↓ with five levels.

Definition . . .



The *design structure* (DS) of an experimental design consists of grouping the EUs into homogeneous groups or blocks with the objective of reducing the experimental error. EUs are grouped so that variability of the EUs *within the groups* is less than the variability among all EUs prior to grouping. The smaller the experimental error, the smaller the significant differences between treatments or conditions that can be detected.

Example:

EU: 10 males and 10 females.

DS: Randomized Complete Block (RCBD)
 ↓
 with sex as the blocking factors.

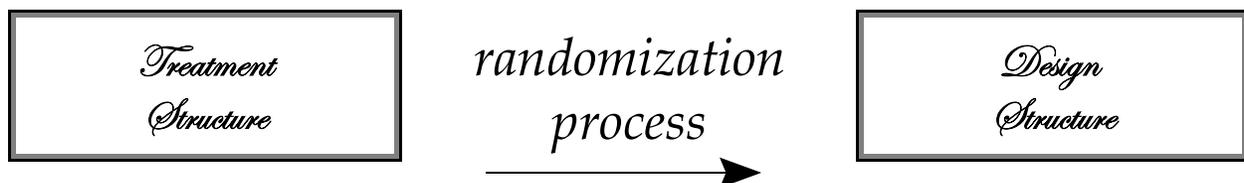
Definition . . .



The *error structure* of an experimental design consists of the variation that cannot be explained by either the treatment or design structures. The error structure provides an estimate of the random variation between EUs subjected to the *same* treatment level or combination. The error structure is the interaction of the treatment and design structures, that is, once the TS and DS have been determined the error structure is set.

An *experimental design* involves *three steps*:

- (1) selecting a *treatment* structure,
- (2) selecting a *design* structure, and
- (3) *randomly assigning* the treatment levels or combinations of the treatment structure to the EUs of the design structure.



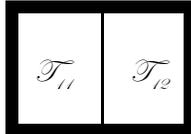
Randomization serves to:

- C provide justification for methods of statistical inference;
- C insure the estimate of the experimental error is valid;
- C provide unbiased estimates of treatment differences.

Let's look at some examples of how the randomization process works, one example with no restrictions on the randomization process (no blocking) and one where the randomization process is restricted (blocking).

NO restriction on the randomization process.

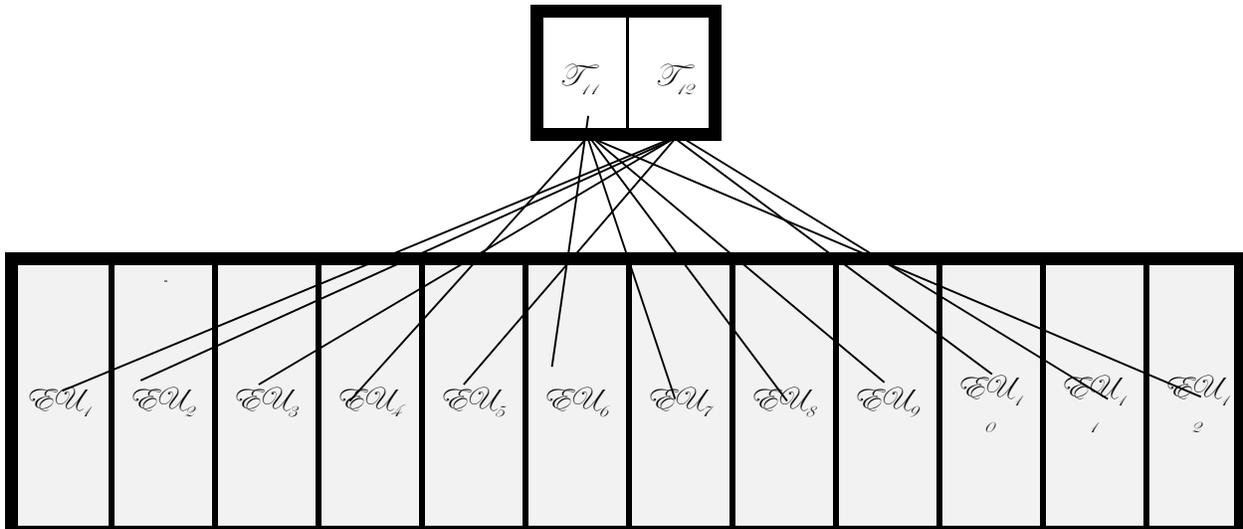
Treatment Structure: one treatment with 2 levels



Design Structure: 12 homogeneous EUs



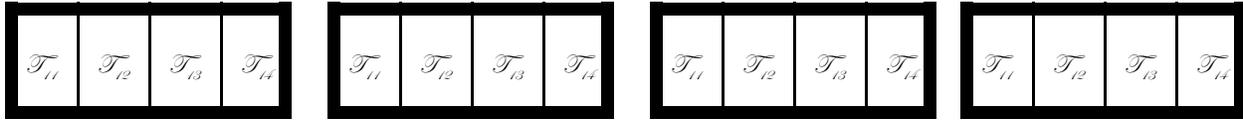
Randomization Process



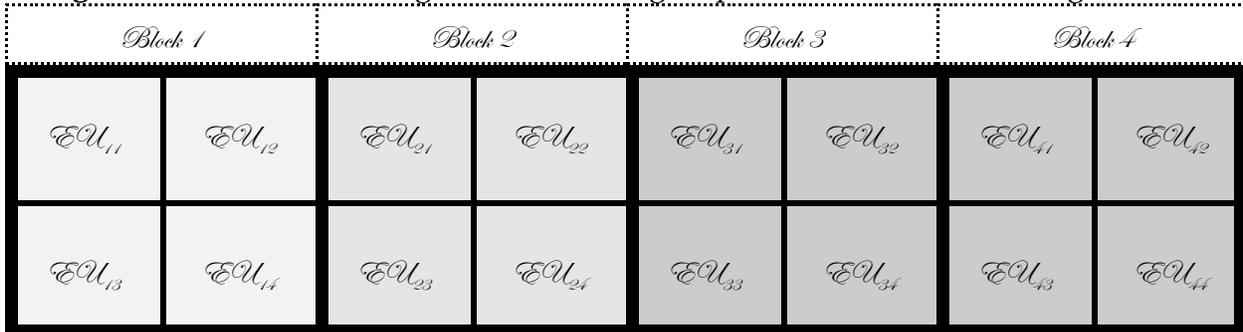
Completely Randomized Design Structure with a One-way Treatment Structure

Restriction on the randomization process.

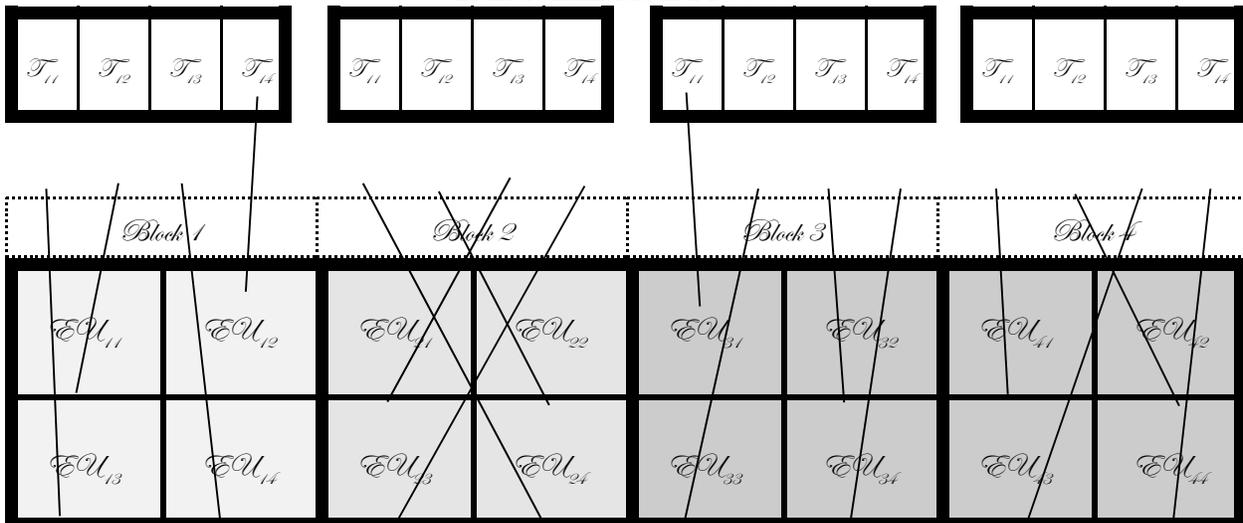
Treatment Structure: one treatment with 4 levels



Design Structure: 12 inhomogeneous EUs are grouped into blocks of 4 homogeneous EUs



Randomization Process



Completely Randomized Block Design Structure with a One-way Treatment Structure
Experimental Design II *ED Review*

An *experimental design* will determine:

C *how to generate the data*

the randomization process that links the TS to the DS determines what EU gets what treatment level or combination

C *the appropriate test statistic*

the ED will determine the structure of the ANOVA table and the appropriate errors terms for evaluating treatment differences

C *how far the conclusions can be generalized*

depending on whether the treatments of the TS are fixed or random the conclusions are valid for only those treatment levels included in the study or extend to the population from which the treatment levels were selected

For example, in the SRRC investigation if we forget about the storage time for the moment, the *treatment structure* is a *two-way*, the treatments are:

- C soil type with 2 levels, and
- C preparation method with 2 levels.

An abbreviated notation refers to this TS as a

(2 x 2).


*There is a number for each
treatment where the number
refers to the number of levels
in that treatment.*

This *design structure* is a randomized complete block design (RCBD). That is, there was a restriction on the randomization process.

What makes it a RCBD is that:

- C each block (panelist) received all four treatment combinations; and
- C within each block (panelist) the order of the four treatment combinations was randomly assigned.

A conceptual model is used to describe the real-world experimental design. This model consists of these three components,

$$y = \begin{array}{c} \textit{Treatment} \\ \textit{Structure} \end{array} + \begin{array}{c} \textit{Design} \\ \textit{Structure} \end{array} + \begin{array}{c} \textit{Error} \\ \textit{Structure} \end{array}.$$

where y is the response variable.

The portion of the *variability* in the *response variable* that can be explained is quantified in the *treatment and design structures* (**deterministic** components). The portion of the variability that is not explained (that is in part a reflection of our ignorance) is quantified in the *error structure* (**random** or **stochastic** components).

Some common treatment structures include:

C one-way

one treatment with 2 or more levels

C two-way

two treatments each with 2 or more levels

C factorial

more than 2 treatments each with 2 or more levels

C fractional factorial

not all treatments combinations evaluated

C factorial with controls.

all treatment combinations are evaluated plus there is a control (s)

Some common design structures include:

- C completely randomized (CRD) ← *no restriction on randomization process*
assume all EUs are all homogeneous
- C randomized complete block (RCBD) ← *randomization process restricted*
EUs are grouped using a single blocking factor
- C Latin square (LSD)
EUs are grouped using two blocking factors
- C incomplete block,
EUs are grouped using one or more blocking factors where blocks do not receive every treatment combination
- C various combinations of the above designs lead to repeated measures, split plots, and nested experiments.

More information on experimental design is available at the Mid South Area's website for statistical services, <http://msa.ars.usda.gov/statmsa/> under Workshop IV.

Three ED references are:

Vecchio, R. J. (1997). *Understanding of Design of Experiments: A Primer for Technologists*, Hanser/Gardner Publications, Inc., Cincinnati, Ohio.

A clearly written, uncomplicated, book that covers a wide range of topics, from the fundamentals to fractional factorials to Taguchi's contributions. The book is pretty much equation free, written by a chemist for the sole purpose of explaining the basic principles of designed experiments.

Milliken, George. and Johnson, Dallas (1984). *Design and Analysis of Messy Data, Volume I: Designed Experiments*, Van Nostrand Reinhold Company, New York.

A more advanced text on classical experimental design. Emphasis is on experiments involving complicated design structures, missing data, outliers, and data that fail to meet the usual assumptions. It is assumed the reader has had a course in analysis of variance as well as some experience in analyzing data. The techniques presented in the book are implemented using the statistical computer package SAS.

Petersen, Roger (1985). *Design and Analysis of Experiments*, Marcel Dekker, New York

A very well written and easy to understand book on classical experimental design and response surfaces. It is assumed the reader has some knowledge of statistical inference, significant tests, analysis of variance, and simple regression. The techniques presented in the book are not implemented using any statistical computer packages.

Let us examine the logic behind ANOVA by simplifying the SRRC investigation even further and treating it as a CRD with a two-way treatment structure. We are going to get rid of the panelists and pretend we have an instrument that can measure the nine flavor and five texture attributes.

In this workshop we are going to use only one of the 14 sensory attributes (response variable). The statistical method where all response variables are analyzed simultaneously is referred to as a Multivariate Analysis of Variance (MANOVA).

Definition . . .

C MANOVA is the multivariate generalization of the *analysis of variance* (ANOVA). It is a technique for testing the equality of a mean *vector* of dependent (response) variables based on a set of independent (explanatory) variables from the treatment and design structures of an *experimental design*.

Why MANOVA?

It might be the case that a combination of variables is necessary to show the difference between groups. In this case, multiple univariate analyses might not detect a real difference.

The experimental setup is the same . . .

The melons were randomly selected from each soil type location and sent to the SRRC facility. At the facility, the melons were culled for defects, washed, peeled, and cut into 2 to 3 cm x 2.5 cm cubes. Ten cubes from a particular soil type were randomly assigned to a juice catcher container and a container randomly assigned to one of the two preparation methods. After preparation, the juice catcher containers were randomly assigned to the shelves within a refrigerator and stored for 4 days. Each of the four treatment combinations had n replicates (4 ~~A~~ EUs or juice catcher containers).

After 4 days, the juice catcher containers were randomly selected from the refrigerator for measurement. Each of the three cubes was measured and a mean calculated.

The observed value for a single EU can be represented symbolically with a *means model*,

$$y_{ir} = \mu_i + c_{ir} \tag{1}$$

subscript for the treatment combinations

subscript for the replications



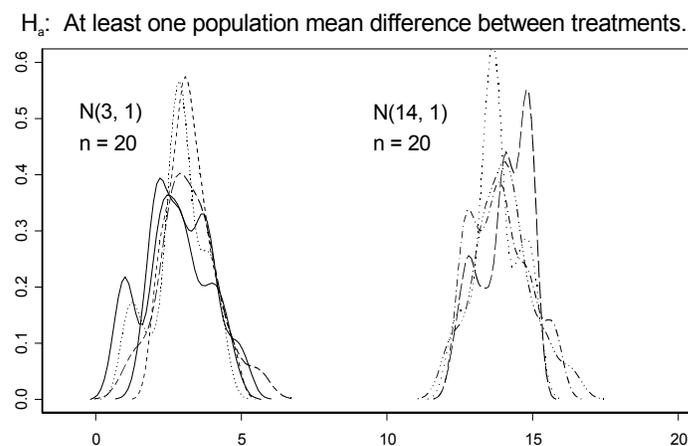
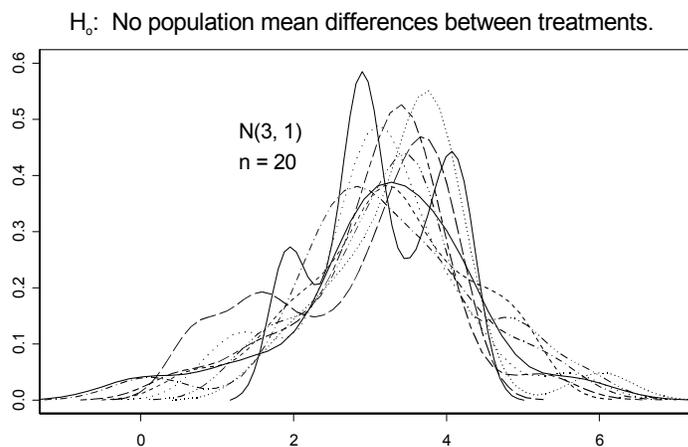
where $\{ i = 1, 2, 3, 4 \}$, $\{ r = 1, 2, \dots, n \}$ and,

y_{ir} = the observed sweetness for the r^{th} juice catcher container from the i^{th} treatment combination;

μ_i = is the response expected when treatment combination, i , is assigned to a randomly selected EU;

c_{ir} = the observed error for the r^{th} juice catcher container from the i^{th} treatment combination.

The null hypothesis of an ANOVA is one of “no difference,” i.e., no difference between treatment combination means. Stated another way, the groups imposed by the design and treatment structures are all sampled from a *single normal distribution*, $N(\mu, \sigma^2)$, with mean μ and variance σ^2 .



The logic behind ANOVA is to obtain an estimate of the population variance, F^2 , for each component in the conceptual model,

$$\begin{array}{c}
 \text{variability between treatments} \\
 \downarrow \\
 \text{total variability} \rightarrow y_{ir} = \mu_i + c_{ir} \\
 \uparrow \\
 \text{variability within a treatment}
 \end{array}$$

hence the term *analysis of variance*.

ANOVA is a method of partitioning the variability associated with the response variable into the components of the conceptual model.

For the SRRC investigation, a CRD with a two-way treatment structure, there are two sources of variability.

- C The *first* estimate is based on the variance *within treatment combination groups*. This is the variability that is inherent in the EU, the variability that no model can explain. The within estimate is the “yardstick” for comparing other measures of variability. The within treatment group variability is associated with the stochastic part of the model, ϵ_{ij} , and is referred to as the *experimental error*.

- C The *second* estimate of the variance is based on the variance *between the treatment combination group means*. This is the variability associated with the deterministic part of the model, μ_i .

The null hypothesis of “no difference” assumes that the EUs in each of the 4 treatment combinations have been sampled from a single normal distribution, $N(\mu, \sigma^2)$, with mean sweetness μ and variance σ^2 . Stated symbolically the null hypothesis is,

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu \quad (2)$$

versus

H_a : at least one of μ_i 's $\{ i = 1, 2, 3, 4 \}$ is not equal to μ .

In other words, under the null hypothesis

- C the *expected value* for the mean sweetness of the 1st group (none / clay) is : and the variance is F^2 ;
- C the *expected value* for the mean sweetness of the 2nd group (none / sandy) is : and the variance is F^2 ;
- C the *expected value* for the mean sweetness of the 3rd group (acid / clay) is : and the variance is F^2 ; and
- C the *expected value* for the mean sweetness of the 4th group (acid / sandy) is : and the variance is F^2 .

The observed statistics and their expected values under the null hypothesis for a CRD with a two-way treatment structure described by a means model is given in Table 1.

Table 1. Observed and expected means and variances for an ANOVA two-way treatment structure under the null hypothesis (2)

<i>Treatment Group</i>	1	2	3	4
<i>Sample Size</i>	n_1	n_2	n_3	n_4
<i>Means</i>				
<i>Observed</i>	\bar{y}_1	\bar{y}_2	\bar{y}_3	\bar{y}_4
<i>Expected</i>	:	:	:	:
<i>Variances</i>				
<i>Observed</i>	s_1^2	s_2^2	s_3^2	s_4^2
<i>Expected</i>	F^2	F^2	F^2	F^2

Table 1 is used to construct a test statistic based on what is *observed* and what is *expected under the null hypothesis*.

The idea is to construct a test statistic based on the two estimates of the population variance, F^2 : variation *within* each of the four treatment combination groups; and variation *between* the four treatment combination groups. If the null hypothesis is true, these two estimates will have the “same” value.

Within Group Variability: c_{ir}

There are 4 observed variances, s_i^2 , one for each treatment combinations group. Each is an estimate of the population variance F^2 . In the SRRC investigation, where there are n replications for each treatment combinations,

$$s_i^2 = \frac{\sum_{r=1}^n (y_{ir} - \bar{y}_i)^2}{n - 1}.$$

\longleftarrow *sum of squares, SS*
 \longleftarrow *degrees of freedom, df*

Instead of having 4 different estimates, we can obtain a better overall estimate by taking the average of the 4 observed variances,

$$\hat{\sigma}_W^2 = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2}{4}.$$

The notation $\hat{\sigma}_W^2$ denotes that this is the *estimate* of the population variance, F^2 , derived from the *observed variances* within each treatment combination group.

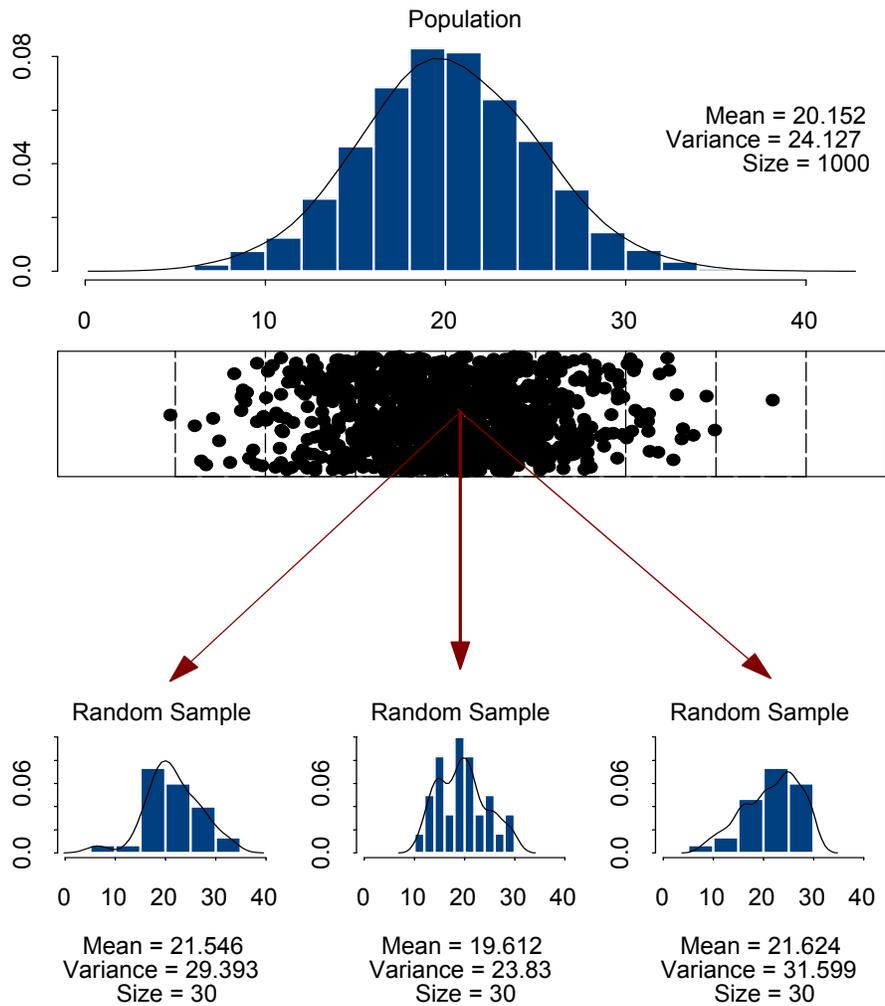
Between Group Variability: σ^2

The between variability is derived from the observed means using the *Central Limit Theorem* (CLT).

The CLT states that,

if random samples of n measurements are repeatedly drawn from a population with mean μ and finite standard deviation σ , then, when n is large, the relative frequency histogram for the sample means will be approximately normal with mean μ and standard deviation σ / \sqrt{n} .

Figure 1. Central Limit Theorem



Based on the diagram, the CLT states that the means from the samples, {21.546, 19.612, 21.624, ...} will themselves follow a normal distribution and that the variability between these sample means will be equal to the

$$\frac{\text{population variance}}{\text{sample size}} = \frac{\sigma^2}{n} = \frac{24.127}{30}.$$

Thus, under the null hypothesis (2), the 4 means in Table 1 are sampled from a single normal distribution with mean : and variance F^2/n .

This is used to construct a second estimate of F^2 by,

- C treating the 4 means as observations,
- C calculating the variance of the 4 mean observations, and
- C multiplying the results by n since the variance between the means is an estimate of F^2/n and we want an estimate of F^2 .

Recall, that we are able to do this based on the assumption of the null hypothesis, the groups imposed by the treatment structure are all sampled from a *single normal distribution*, $N(\mu, F^2)$, with mean μ and variance F^2 .

If we let \bar{y} denote the overall mean of the 4 mean observations, then

$$\hat{\sigma}_B^2 = n \left[\frac{\sum_{i=1}^4 (\bar{y}_i - \bar{y})^2}{4 - 1} \right],$$

where $\hat{\sigma}_B^2$ denotes that this is the *estimate* of the population variance, F^2 , derived from the variability between the *observed means*.

Under the assumption of the null hypothesis (2) we would expect the ratio of the two variance estimates to be close to 1.0,

$$E \left[\frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2} \right] \approx 1.0. \quad (3)$$

That is, both the numerator and denominator are estimates of the population variance, F^2 .

If the null hypothesis (2) is rejected, at least one of the group means is different from the population mean. Under this scenario we would expect $\hat{\sigma}_B^2 > \hat{\sigma}_W^2$ and therefore (3) would be greater than 1.

In ANOVA terminology,

C $\hat{\sigma}_B^2$ is the mean squares between groups or the treatment mean squares, MS_T ;

C $\hat{\sigma}_W^2$ is the mean squares within groups or the error mean squares, MS_E ; and

C (3) is the test statistic, $\frac{MS_T}{MS_E} = F\text{-statistic}$.

Traditionally, this information is presented in a table that identifies how the variability of the response variable has been partitioned into deterministic and random components. This is called the *analysis of variance table*.

Analysis of Variance Table a CRD

Source of Variation	degrees of freedom	Sum of Squares	Mean Squares	Test Statistic
Treatment Structure	df_T	SS_T	$MS_T = SS_T / df_T$	MS_T / MS_E
Error Structure	df_E	SS_E	$MS_E = SS_E / df_E$	
Total	$df_T + df_E$	SS_{TOTAL}		

where $\frac{SS_T}{df_T} = MS_T$, $\frac{SS_E}{df_E} = MS_E$, and $F = \frac{MS_T}{MS_E}$

On a nontechnical level the basic idea of *degrees of freedom* (*df*) are the number of pieces of information (number of EUs) that are free to vary.

Definition . . .

C Associated with a *SS* are its *degrees of freedom*. The number of *degrees of freedom* for a *SS* of a set of n observations is based on the algebraic identity

$$\sum_{r=1}^n (y_{ir} - \bar{y}_i) = 0.$$

We are free to specify $(n - 1)$ of the deviations, the n^{th} deviation is fixed to make the sum equal to 0.

For example, in the SRRC investigation, the degrees of freedom for the treatment structure are, $df_T = 3$. This is based on the sum of squares (see page 44) associated with the treatment structure,

$$\sum_{i=1}^4 (\bar{y}_i - \bar{\bar{y}})^2.$$

We are free to specify $(4 - 1)$ of the deviations, the 4th deviation is fixed to satisfy the identity,

$$\sum_{i=1}^4 (\bar{y}_i - \bar{\bar{y}}) = 0.$$

Based on the equation (see page 38) for s_i^2 , the degrees of freedom for the error structure is $df_E = 4*(n - 1)$.

The total degrees of freedom are just the number of EUs minus one. For this investigation the number of EUs are, $(4*n - 1)$.

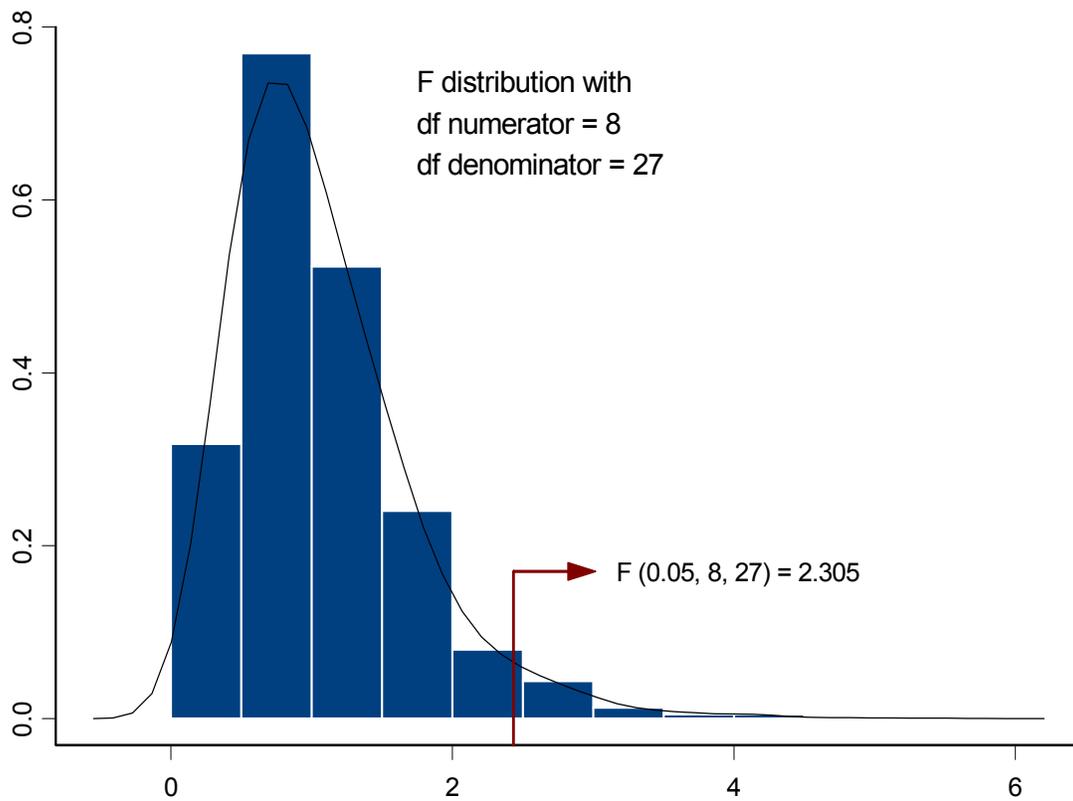
The test statistic (3) follows a “name brand” distribution, called the *F*-distribution.

Properties of the *F*-distribution

- C An *F-value* can only assume positive values.
- C The *F*-distribution, unlike the normal, is nonsymmetric.
- C There are many *F*-distributions and each one has a slightly different shape.
- C A particular *F*-distribution is specified by the *degrees of freedom* associated with the numerator and denominator of the test statistic.

For example ,

$$\frac{MS_T}{MS_E} \sim F_{(df_T, df_E)}$$



Analysis of Variance Table (page 18)
Completely Randomized DS with a One-way TS

<i>Source of Variation</i>	<i>degrees of freedom</i>	<i>Mean Square</i>	<i>F-Test Statistic</i>
TOTAL	12! 1 = 11		
DESIGN STRUCTURE			
TREATMENT STRUCTURE			
One-way treatment with 2 levels	2! 1 = 1	MST	MST / MSE
ERROR STRUCTURE			
Experimental Error	11! 1 = 10	MSE	

$$df_{TOTAL} = \# \text{ of EUs } - 1 \text{ for the grand mean}$$

$$df_T = \# \text{ of treatment levels } - 1$$

$$df_E = df_{TOTAL} - df_T$$

Analysis of Variance Table (page 19)
Randomized Complete Block DS with a One-way TS

<i>Source of Variation</i>	<i>degrees of freedom</i>	<i>Mean Square</i>	<i>F-Test Statistic</i>
TOTAL	16! 1 = 15		
<i>DESIGN STRUCTURE</i>			
Four Blocks	4! 1 = 3	MSD	
<i>TREATMENT STRUCTURE</i>			
One-way with 4 levels	4! 1 = 3	MST	MST / MSE
<i>ERROR STRUCTURE</i>			
Error	15! 3! 3 = 9	MSE	

$$df_{TOTAL} = \# \text{ of EUs} ! 1 \text{ for the grand mean}$$

$$df_D = \# \text{ of blocks} ! 1$$

$$df_T = \# \text{ of treatment levels} ! 1$$

$$df_E = df_{TOTAL} ! df_T ! df_D$$

In developing the logic behind the ANOVA note that we have made the following *assumptions*:

C the samples are *independent random samples*;

C each sample is selected from a *normal population*;
and

C each sample comes from a population with a *common variance*, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma^2$.

Stated another way c_{ir} - i.i.d. $N(0, \sigma^2)$.

We have also made some *assumptions* about blocking.

- C The degrees of freedom for error in the RCBD example can be calculated as the interaction between the design and treatment structures. For example in the ANOVA table on page 53,

$$df_E = df_D \times df_T = (3 - 1)(3 - 1) = 9.$$

- C If the treatments behave differently within the blocks, that is - if they interact, the experimental error will be large and it will be harder to detect differences. This defeats the purpose of blocking which is to reduce the experimental error.
- C This leads to one of the *assumptions* of blocking, the treatments and the blocks do not interact, the treatments behave in a similar manner within each block.

Two ANOVA references are:

Ott, Lyman (1988). *An Introduction to Statistical Methods and Data Analysis*, 3rd edition, PWS-Kent Publishing Company, Boston.

This book is appropriate for someone with little or no exposure to statistical methods and data analysis. It assumes a background in high-school algebra and no prior knowledge of statistics.

Milliken, George and Johnson, Dallas (1984). *Analysis of Messy Data, Volume I: Designed Experiments*, Van Nostrand Reinhold Company, New York.

A more advanced text on classical experimental design. Emphasis is on experiments involving complicated design structures, missing data, outliers, and data that fail to meet the usual assumptions. It is assumed the reader has had a course in analysis of variance as well as some experience in analyzing data. The techniques presented in the book are implemented using the statistical computer package SAS.

Three classics are listed below, they might be hard to learn from if you are just starting out, but they are great references.

Cochran, W. G. and Cox, Gertrude (1957). *Experimental Design*, 2nd Ed, John Wiley & Sons, Inc, New York, New York.

Snedecor George, W. and Cochran, W. G. (1967). *Statistical Methods*, 6th Ed, Iowa State University Press, Ames, Iowa.

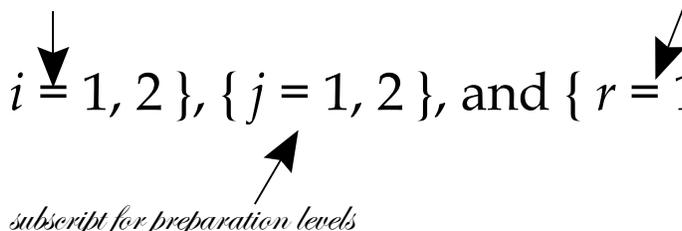
Steel, Robert G. D., and Torrie, J. H. (1960). *Principles and Procedures of Statistics*, McGraw-Hill Book Company, New York, New York.

The more common way to symbolically describe an experimental design is to use an *effects model*. In an effects model, each effect in the treatment and design structures is represented by a term in the model. There are also terms for treatment interactions.

For example, continuing to treat the SRRC investigation as a CRD with a two-way treatment structure (described on page 27), the measurement on a single EU is represented as,

$$y_{ijr} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + c_{ijr} \tag{4}$$

where $\{ i = 1, 2 \}$, $\{ j = 1, 2 \}$, and $\{ r = 1, 2, 3, 4 \}$.

subscript for soil levels *subscript for replications*

subscript for preparation levels

(4) is referred to as a *fixed effects model*, the terms are:

y_{ijr} = the observed sweetness for the r^{th} juice catcher container from the i^{th} soil and the j^{th} preparation;

μ = represents the overall mean effect;

α_i = represents the effect on the mean from the i^{th} soil type;

β_j^* = represents the effect on the mean from the j^{th} preparation;

$(\alpha\beta^*)_{ij}$ = represents any additional effects that might result from using α_i and β_j^* at the same time on an EU; and

c_{ijr} = represents the random effect associated with the r^{th} juice catcher container from the i^{th} soil type receiving the j^{th} preparation.

Where does the designation *fixed effects model* come from?

In addition to each term in the effects model being classified as part of the treatment or design structure, each can be classified as a *fixed* or *random effect*.

Whether or not an effect is classified as fixed or random depends on how the researcher selected the levels of a particular treatment or factor.

Definition . . .

- C A factor is considered a *fixed effect* when it has been *systematically selected* because of the researcher's interests. It is assumed that all treatment levels about which inferences are to be made are included in the experiment.

A factor with *fixed levels* consists of a series of identifiable populations, each with its own mean. Interest lies in estimating the mean of each of each population.

Examples of *Fixed Effects*

<i>Treatment</i>	<i>Levels</i>
Diets	Protein Carbohydrate
Cooler Temperature	34° F 40° F 46° F
Fertilizer	2 mg per plot 4 mg per plot 6 mg per plot 8 mg per plot
Fan	On Off

the researcher is interested in only these treatment levels - inferences extend only to these levels

What are the assumptions of the model (4):

1. The *fixed effects*, since they are deviations from : , all sum to 0. For example,

$$\sum_{i=1}^2 \gamma_i = 0, \sum_{j=1}^2 \delta_j = 0, \text{ and } \sum_{i=1}^2 \sum_{j=1}^2 (\gamma \delta)_{ij} = 0.$$

2. The c_{ijr} are a random sample from a populations which is normally distributed, has a mean of 0, and has a common variance F^2 . These assumptions are symbolized as

$$c_{ijr} - \text{i.i.d. } N(0, \sigma_c^2).$$

Some additional comments:

- C (4) is referred to as a fixed effects model even though it contains c_{ijr} which is a random effect, it is a fixed effects model since every term except the error term is fixed;
- C the estimation goal for fixed effects is to estimate the treatment means and the difference among treatment means;
- C if the experiment were to be repeated the same treatments would be included.

Now let's add the panelist back into the SRRC investigation. We now have a two-way treatment structure in a randomized complete block design (RCBD). Each panelist receives each of the four treatment combinations, so panelist serves as a blocking factor.

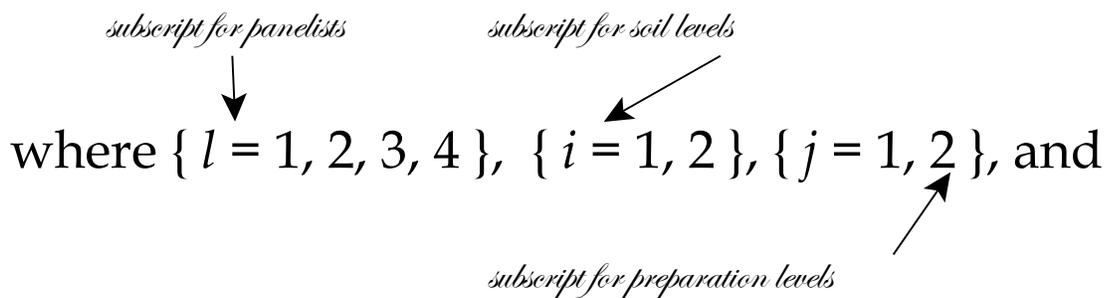
What type of effect is the block, panelist? Two questions to ask are:

- C how were the panelist selected; and
- C where are the inferences being made to, just to the panelists involved in the study or to the entire population of people who eat cantaloup?

The measurement on a single EU is now represented as,

$$y_{lij} = \mu + \alpha_l + (\alpha_i + \alpha_j + (\alpha_{ij}) + c_{lij}) \quad (5)$$

where $\{ l = 1, 2, 3, 4 \}$, $\{ i = 1, 2 \}$, $\{ j = 1, 2 \}$, and

subscript for panelists *subscript for soil levels*

subscript for preparation levels

α_l = represents the random effect associated with the l^{th} panelist.

Definitions . . .

- C A factor is considered a *random effect* when it has been *randomly selected* from a population of components. Inferences are made to the entire population.

A factor with random levels consists of a single population from which the levels being investigated are a sample. Interest lies in the variability within the population from which the sample came (*variance component*), or perhaps in a prediction of the mean of a particular level (*BLUP*).

Examples of *Random Effects*

<i>Treatment</i>	<i>Levels</i>	
Panelist	Susie Lorinda Hubert Jake	
High Schools in EBR	BR Magnet McKinley Woodale Episcopal	<i>the researcher has randomly selected these treatment levels from the population - inferences extend back to the population</i>
States	Utah Georgia Alaska Maine	

What are the assumptions of the model (5):

1. The *fixed effects*, since they are deviations from : , all sum to 0. That is,

$$\sum_{i=1}^2 \gamma_i = 0, \sum_{j=1}^2 \delta_j = 0, \text{ and } \sum_{i=1}^2 \sum_{j=1}^2 (\gamma \delta)_{ij} = 0.$$

2. The *random effects* are

$$c_{ijr} - \text{i.i.d. } N(0, F_c^2)$$

and

$$\$_l - \text{i.i.d. } N(0, F_{\$}^2),$$

*referred to as
a variance
components*

where $\$_l$ and c_{ijr} are distributed independently of each other.

Some additional comments:

- C (5) is referred to as a fixed effects model even though it contains β_l and c_{ijr} which are random effects;
- C when the random effects are only in the design and error structures, the model is still referred to as a fixed effect model;
- C a “rule of thumb” is that the components of the design structure are random effects; and
- C if the experiment were to be repeated a different combination of panelists would be included.

Now let's add the day variable back into the SRRC investigation. This is now a case where there is more than one size of experimental unit (EU) in an investigation.

The experimental designs that have *several sizes of experimental units* (SSEU) are *repeated measures* designs, *split plot* designs, and some *nested* designs.

Characteristics of experimental designs with SSEU are:

- C a treatment structure that is at least a two-way,
- C a separate treatment, design, and error structures for each EU.

In the SRRC investigation, the EUs are:

Large EU (LEU)
juice catcher container

DS	RCBD panelist is the block
TS	Two-way (2x2) C soil type, clay and sand C preparation method, acid and none

Small EU (SEU)
time interval within juice catcher container

DS	RCBD juice catcher container is the block
TS	One-way C day with four levels 0, 4, 7, and 10

In a repeated measures designs, the interest is in

- C between-LEU effects, the values that change only from container to container and remain the same for all observations on a single container, for example soil and preparation effects averaged over the within EU measurements;
- C within-LEU effects, the values that change from measurement to measurement, for example day effect averaged over soil type and preparation method;
- C between-LEU and within-LEU interaction, for example how soil type and preparation method change over time.

The measurement on a single EU is now represented as,

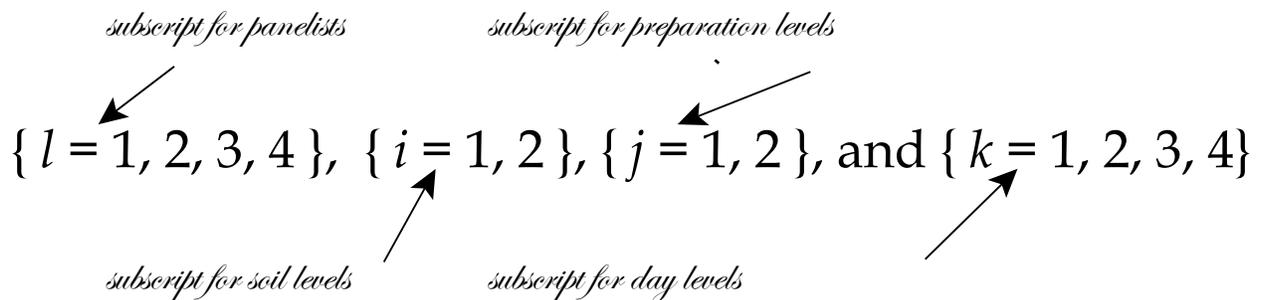
$$y_{lijk} = \mu + \underbrace{\$}_l + \underbrace{(\alpha)}_i + \underbrace{(\beta)}_j + \underbrace{((\alpha\beta))}_{ij} + \underbrace{c}_{lijk} \tag{6}$$

juice catcher container EU

$$+ \underbrace{J}_k + \underbrace{((J))}_{ik} + \underbrace{(\beta J)}_{jk} + \underbrace{((\beta J))}_{ijk} + \underbrace{\epsilon}_{lijk}$$

time interval EU

where



J_k = represents the effect on the mean as a result of being sampled on the k^{th} day, and

ϵ_{lijk} = represents the random effect associated with the r^{th} container from the i^{th} soil type receiving the j^{th} preparation on the k^{th} day.

What are the assumptions of the model (6)?

1. The *fixed effects*, since they are deviations from : , all sum to 0. For example,

$$\sum_{i=1}^2 \gamma_i = 0, \sum_{j=1}^2 \delta_j = 0, \dots, \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^4 (\gamma \delta \tau)_{ijk} = 0.$$

2. The *random effects*

$$c_{ijr} - \text{i.i.d. } N(0, F_c^2),$$

$$\$l - \text{i.i.d. } N(0, F_{\$}^2), \text{ and}$$

$$,_{lijk} - \text{i.i.d. } N(0, F_{,}^2).$$

Is this an appropriate assumption for repeated measurements on the same EU??

What makes repeated measures data analysis distinct if the *covariance structure*, Σ .

Making the assumption of independent errors, ϵ_{ijk} - i.i.d. $N(0, \sigma^2)$, means that all observations within a given container are uncorrelated and have equal variances. However, when measurements are made over time a more appropriate assumption might be that two measurements taken at adjacent times are more correlated than two measurements taken several time periods apart.

In a repeated measures analysis one must specify what the relationship is between the measurements taken at different points in time, that is the covariance matrix Σ must be specified.

For example, the population covariance matrix for the SRRC investigation would have the following structure.

E	Day 0	Day 4	Day 7	Day 10
Day 0	<i>variance</i>	covariance	covariance	covariance
Day 4	covariance	<i>variance</i>	covariance	covariance
Day 7	covariance	covariance	<i>variance</i>	covariance
Day 10	covariance	covariance	covariance	<i>variance</i>

$\sigma_{(10)}$ is the covariance between the Day 0 and Day 10 where $\sigma_{(10)} = \rho_{(10)} \sigma_1 \sigma_4$

σ_4^2 is the variance the for Day 10

How do we interpret \mathbf{E} , the covariance matrix?

- C The *diagonal elements* are the within container population variances for the days, $\{k = 1, 2, 3, 4\}$.
- C The *off diagonals elements* are the within container population covariances between days,

$$F_{kk'} = D_{kk'} F_k F_{k'}$$

where $D_{kk'}$ is the within container population correlation between the k^{th} and k'^{th} day. The covariance tells us the extent to which EUs within a container who have a high attribute measurement for the k^{th} day also tend to have large RR for k'^{th} day.

There are a number of “name brand” covariance structures that can be specified. Each structure comes with different assumptions and a different number of parameters to estimate.

Uncorrelated Y assumes ϵ_{ijk} - i.i.d. $N(0, \sigma^2)$, that is, the repeated measurements are uncorrelated and have equal variances. There is one parameter to estimate, the variance.

$$E_{UC} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Compound Symmetry (CS) Σ assumes the repeated measurements are equally correlated and have equal variances. There are two parameters to estimate, the variance and correlation.

$$E_{CS} = \sigma_{\epsilon}^2 \begin{bmatrix} 1 & D & D & D \\ D & 1 & D & D \\ D & D & 1 & D \\ D & D & D & 1 \end{bmatrix}$$

Heterogeneous Compound Symmetry (CSH) Σ assumes the same correlation structure as CS but heterogeneous variances. There are $k + 1$ parameters to estimate, k variances and 1 correlation.

Autoregressive Order One (AR(1)) Y assumes that repeated measurements w time periods apart are D^w correlated and have equal variances. There are two parameters to estimate, the variance and correlation.

$$E_{AR(1)} = \sigma_{\epsilon}^2 \begin{bmatrix} 1 & D^1 & D^2 & D^3 \\ D^1 & 1 & D^1 & D^2 \\ D^2 & D^1 & 1 & D^1 \\ D^3 & D^2 & D^1 & 1 \end{bmatrix}$$

Heterogeneous Autoregressive Order One (ARH(1)) Y assumes the same correlation structure as AR(1) but heterogeneous variances. There are $k + 1$ parameters to estimate, k variances and 1 correlation.

Huynh-Feldt Condition Υ also referred to as the “spherical structure,” is similar to CSH in that it assumes heterogeneous variances, but the covariances are constructed by taking arithmetic means. There are $k + 1$ parameters to estimate, k covariances and one covariance.

$$E_{HF} = \begin{bmatrix} \lambda + 2\gamma_1 & \binom{1}{1} + \binom{2}{2} & \binom{1}{1} + \binom{3}{3} & \binom{1}{1} + \binom{4}{4} \\ \binom{1}{1} + \binom{2}{2} & \lambda + 2\gamma_2 & \binom{2}{2} + \binom{3}{3} & \binom{2}{2} + \binom{4}{4} \\ \binom{1}{1} + \binom{3}{3} & \binom{2}{2} + \binom{3}{3} & \lambda + 2\gamma_3 & \binom{3}{3} + \binom{4}{4} \\ \binom{1}{1} + \binom{4}{4} & \binom{2}{2} + \binom{4}{4} & \binom{3}{3} + \binom{4}{4} & \lambda + 2\gamma_4 \end{bmatrix}$$

Unstructured Σ assumes there is no structure. There are $k + \frac{k(k-1)}{2}$ parameters to estimate, k variances and $\frac{k(k-1)}{2}$ covariances.

$$E_{UN} = \begin{bmatrix} \sigma_1^2 & F_{12} & F_{13} & F_{14} \\ F_{12} & \sigma_2^2 & F_{23} & F_{24} \\ F_{13} & F_{23} & \sigma_3^2 & F_{34} \\ F_{14} & F_{24} & F_{34} & \sigma_4^2 \end{bmatrix}$$

The strategy for selecting a covariance structure is to fit several and to compare them using various criteria. The decision making process is assisted by using two model-fitting criteria the,

C Akaike's Information Criteria (AIC)

C Schwarz's Bayesian Criterion (BIC).

Both criteria take into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve this particular degree of fit. Each imposes a penalty for increasing the number of parameters.

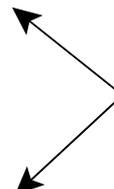
The AIC is defined as

$$-2\log(L) + 2m, \text{ and.}$$

the BIC is defined as

$$-2\log(L) + m\log(n),$$

a function of the model



where L is the maximum likelihood, m is the number of parameters, and n is the sample size.

Lower values of the criteria indicate the preferred model.

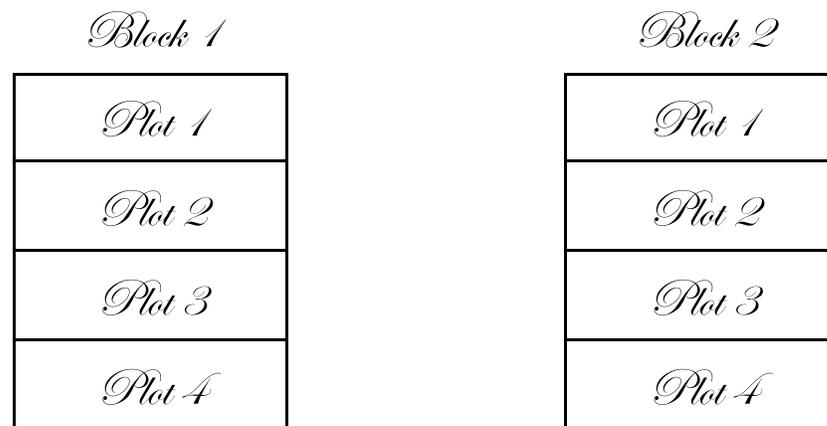
Experimental Design II

Repeated Measures

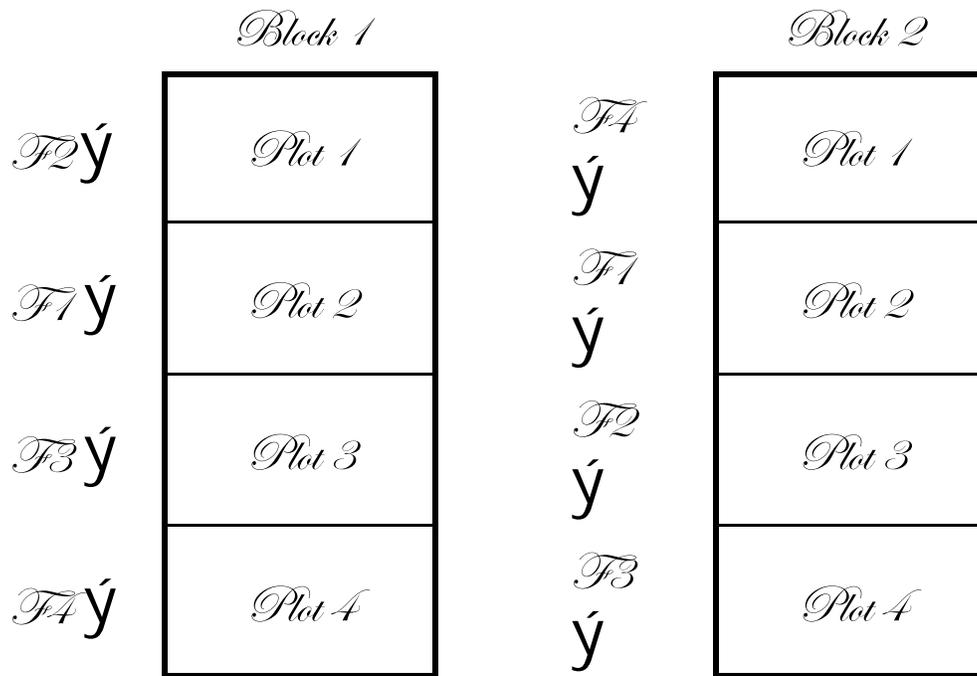
Source of Variation		degrees of freedom	Mean Square	F-Test Statistic
TOTAL		$(4 \times 4 \times 4) ! 1 = 63$		
DESIGN STRUCTURE (within container): Randomized Complete Block with 16 blocks	DESIGN STRUCTURE (between container): Randomized complete block			
	Panelist	$4 ! 1 = 3$	MSD_1	
	TREATMENT STRUCTURE (between container): Two-way (2x2)			
	Soil Type	$2 ! 1 = 1$	MST_1	MST_1 / MSE_1
	Preparation Method	$2 ! 1 = 1$	MST_2	MST_2 / MSE_1
	Soil Type x Preparation Method	$2 ! 1 = 1$	MST_3	MST_3 / MSE_1
	ERROR STRUCTURE (between container)			
	Error(container)	$16 ! 3 ! 1 ! 1 ! 1 = 9$	MSE_1	
	TREATMENT STRUCTURE (within container): One-way with 4 levels			
	Day	$4 ! 1 = 3$	MST_4	MST_4 / MSE_2
Day x Soil Type	$(4 ! 1)(2 ! 1) = 3$	MST_5	MST_5 / MSE_2	
Day x Preparation Method	$(4 ! 1)(2 ! 1) = 3$	MST_6	MST_6 / MSE_2	
Day x Soil Type x Preparation Method	$(4 ! 1)(2 ! 1) = 3$	MST_7	MST_7 / MSE_2	
ERROR STRUCTURE (within container)				
Error(within container)	$63 ! (4 + 1 + 1 + 1 + 9) ! 3 ! 3 ! 3 ! 3 = 36$	MSE_2		

Now let's introduce a split plot example.

A researcher is interested in evaluating two varieties and four fertilizer amounts on wheat yield. The field is divided into two blocks. Within each block four plots are laid out.

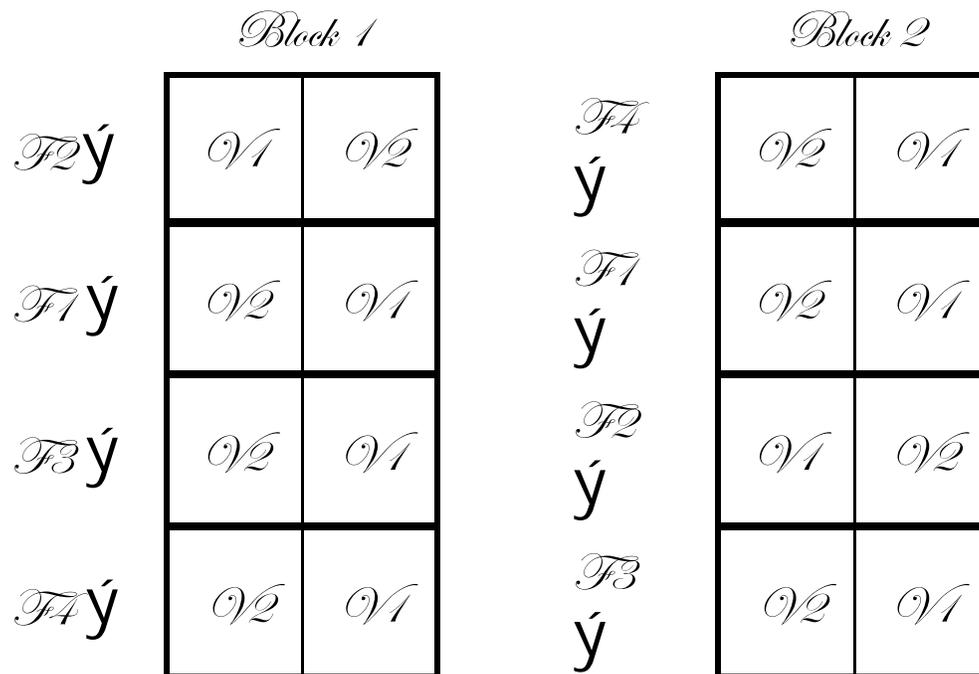


The four levels of fertilizer are randomly assigned to the four plots within each block.



This is a randomized complete block design (2 blocks) with a one-way treatment structure, fertilizer with 4 levels. There are 8 EUs, plots.

Within each of the eight plots, two sub-plots are laid out. The two varieties of wheat are randomly assigned to the sub-plots within each plot.



This is a randomized complete block design (8 plots) with a one-way treatment structure, variety with 2 levels. There are 16 EUs, sub-plots.

Characteristics of a split plot are:

- C a treatment structure that is at least a two-way;
- C two or more EUs where each EU has it's own treatment, design, and error structures;
- C all treatments are randomly assigned and at least two of the treatments require a separate randomization process (this is what distinguishes a repeated measures from a split plot, a repeated measures has at least one treatment that cannot be randomly assigned).

Large EU (LEU)
plots

DS	RCBD two blocks
TS	One-way (4) C four levels of fertilizer

Small EU (SEU)
sub-plots with each plot

DS	RCBD eight plots
TS	One-way (2) C two wheat varieties

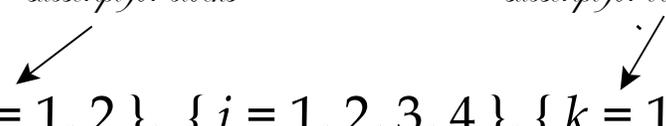
The measurement on a single EU is represented as,

$$y_{jik} = \mu + \beta_j + F_i + (F)_{ji} + V_k + (FV)_{ik} + \epsilon_{jik} \quad (7)$$

plot EU

sub-plot EU

where

subscript for blocks *subscript for varieties*

 $\{ j = 1, 2 \}, \{ i = 1, 2, 3, 4 \}, \{ k = 1, 2 \},$ and
subscript for fertilizer levels

y_{jik} = represents the observed response of variety k grown with fertilizer regime i in block j ;

μ = represents the overall mean effect;

$\$l$ = represents the *random* effect associated with the l^{th} block;

F_i = represents the effect on the mean from the i th fertilizer level;

$(\$F)_{ji}$ = represents the plot (LEU) error; and

V_k = represents the effect on the mean from the k^{th} variety;

$(FV)_{ji}$ = represents any additional effects that might result from using F_i and V_k at the same time on an EU; and

ϵ_{jik} = represents the sub-plot (SEU) error.

What are the assumptions of the model (7)?

1. The *fixed effects*, since they are deviations from μ , all sum to 0. For example,

$$\sum_{i=1}^4 F_i = 0, \sum_{k=1}^2 V_k = 0, \text{ and } \sum_{i=1}^4 \sum_{k=1}^2 (FV)_{ik} = 0.$$

2. The *random effects*

$$\mu_j - \mu - F_j - V_j - (FV)_{jk} - \epsilon_{jki} \text{ - i.i.d. } N(0, F_{\mu}^2),$$

$$(F)_{ji} - \mu_j - F_j - V_j - (FV)_{jk} - \epsilon_{jki} \text{ - i.i.d. } N(0, F_{F}^2),$$

$$\epsilon_{jki} \text{ - i.i.d. } N(0, F_{\epsilon}^2), \text{ and}$$

$(F)_{ji}$ and ϵ_{jki} are independently distributed

Some additional comments:

- C (7) is referred to as a fixed model effects since the treatment effects are all fixed;
- C when there is an interaction between a random and a fixed effect, the interaction effect is random;
- C a “rule of thumb” is that interactions between the treatment and design structure are part of the error structure.

Experimental Design II

Split Plot

Source of Variation		degrees of freedom	Mean Square	F-Test Statistic
TOTAL		$(2 \times 4 \times 2) ! - 1 = 15$		
DESIGN STRUCTURE (sub-plot): Randomized Complete Block with 8 blocks (plots)	DESIGN STRUCTURE (between plot): Randomized complete block			
	Block	$2 ! - 1 = 1$	MSD_1	
	TREATMENT STRUCTURE (between plot): One-way with 4 levels			
	Fertilizer	$4 ! - 1 = 3$	MST_1	MST_1 / MSE_1
	ERROR STRUCTURE (between plot)			
	Error(plot)	$(2 ! - 1)(4 ! - 1) = 3$	MSE_1	
	TREATMENT STRUCTURE (sub-plot): One-way with 2 levels			
	Variety	$4 ! - 1 = 3$	MST_4	MST_4 / MSE_2
	Fertilizer x Variety	$(4 ! - 1)(2 ! - 1) = 3$	MST_5	MST_5 / MSE_2
	ERROR STRUCTURE (sub-plot)			
Error(sub-plot)	$15 ! - (1 + 3 + 3) ! - 3 ! - 3 = 2$	MSE_2		

Two repeated measures and split plot references are:

Milliken, George and Johnson, Dallas (1984). *Analysis of Messy Data, Volume I: Designed Experiments*, Van Nostrand Reinhold Company, New York.

A more advanced text on classical experimental design. Emphasis is on experiments involving complicated design structures, missing data, outliers, and data that fail to meet the usual assumptions. It is assumed the reader has had a course in analysis of variance as well as some experience in analyzing data. The techniques presented in the book are implemented using the statistical computer package SAS.

Littell, Ramon, Milliken, George, Stroup, Walter, and Wolfinger, Russell (1996). *SAS System for Mixed Models*. Although this SAS manual does not mention Analyst it is a valuable reference for interpreting SAS output. The manual works through many examples in detail. New versions of SAS have been released since the book was first published.