**Experimental Design**

Sampling versus experiments

- similar to sampling and inventory design in that information about forest variables is gathered and analyzed
- experiments presuppose intervention through applying a *treatment* (an action or absence of an action) to a unit, called the *experimental unit*. The experimental unit is an item on which the treatment is applied.
- The goal is to obtain results that indicate cause and effect.



Definitions of terms and examples

- For each experimental unit, measures of the *variables of interest* (i.e., *response* or *dependent variables*) are used to indicate treatment impacts.
- Treatments are randomly assigned to the experimental units.
- *Replication* is the observation of two or more experimental units under identical experimental conditions.
- A *factor* is a grouping of related treatments.

Examples:

1. 1,000 seedlings in a field. Half of the seedlings get a "tea bag" of nutrients, others do not, *randomly* assigned.

   Experimental unit: the seedling.

   Treatments are: no tea bag, and tea bag.

   Factor: only one – fertilizer (none, tea bag)

   Replications: 500 seedlings get each treatment

2. 300 plant pots in a greenhouse: Each plant gets either 1) standard genetic stock; 2) genetic stock from another location; 3) improved genetic stock.

   Treatments:

   Experimental Unit:

   Factor(s):

   Replications:



3. The number of tailed frogs in different forest types is of interest. There are six areas. Three are cut and the other three are not cut.

   Treatments:

   Experimental Unit:

   Factor(s):

   Replications:

4. Two forest types are identified, Coastal western hemlock and interior Douglas fir. For each, a number of samples are located, and the growth of each tree in each sample is measured.

   Treatments:

   Experimental Unit:

   Factor(s):

   Replications:

What does it mean that *treatments are randomly assigned to experimental units*?

- Haphazard vs. random allocation
- Practical problems and implications

Other terms:

- The *null hypothesis* is that there are no differences among the treatment means. For more than one factor, there is more than one hypothesis
- The sum of squared differences (termed, *sum of squares*) between the average for the response variable by treatment versus the average over all experimental units represents the variation attributed to a factor.
- The *degrees of freedom*, associated with a factor, are the number of treatment levels within the factor minus one.



Example of hypotheses:

Factor A, fertilizer: none, medium, heavy (3 levels)

Factor B, species: spruce, pine (2 levels)

Number of possible treatments: 6 e..g, spruce, none is one treatment.

Experimental Unit: 0.001 ha plots

Replicates planned: 2 per treatment (cost constraint). How many experimental units do we need?

Variable of interest: Average 5-year height growth for trees in the plot

Null hypotheses:

There is no different between the 6 treatments. This can be broken into:

1) There is no interaction between species and fertilizer.
2) There is no difference between species.
3) There is no difference between fertilizers.

- *Experimental error* is the measure of variance due to chance causes, among experimental units that received the same treatment.
- The degrees of freedom for the experimental error relate to the number of experimental units and the number of treatment levels.
- The impacts of treatments on the response variables will be detectable only if the impacts are measurably larger than the variance due to chance causes.
- To reduce the variability due to causes other than those manipulated by the experimenter, relatively homogenous experimental units are carefully selected.



- Random allocation of a treatment to an experimental unit helps insure that the measured results are due to the treatment, and not to another cause.

Example: if we have applied the no fertilizer treatment to experimental units on north facing sites, whereas moderate and heavy fertilizer treatments are applied only to south facing sites, we would not know if differences in average height growth were due to the application of fertilization, the orientation of the sites, or both. The results would be *confounded* and very difficult to interpret.

## Variations in experimental design

*Introduction of More Than One Factor:*

- Interested in the interaction among factors, and the effect of each factor.
- A treatment represents a particular combination of levels from each of the factors.
- When all factor levels of one factor are given for all levels of each of the other factors, this is a *crossed experiment*.

  Example: two species and three fertilization levels = six treatments using a crossed experiment.



*Fixed, Random, or Mixed Effects:*

- *Fixed factors*: the experimenter would like to know the change that is due to the particular treatments applied; only interested in the treatment levels that are in the experiment (e.g., difference in growth between two particular genetic stocks) [*fixed effects*]
- *Random factors*: the variance due to the factor is of interest, not particular levels (e.g., variance due to different genetic stocks—randomly select different stock to use as the treatment) [*random effects*]
- Mixture of factor types: Commonly, experiments in forestry include a mixture of factors, some random and some fixed [*mixed effect*].

*Restricted Randomization Through Blocking: Randomized Block (RCB), Latin Square, and Incomplete Blocks Designs:*

- Randomize treatments with blocks of experimental units
- Reduces the variance by taking away variance due to the item used in blocking (e.g., high, medium and low site productivity
- Results in more homogeneous experimental units within each block.



*Restricted Randomization Through Splitting Experimental Units:*

- Called "split plot"
- An experimental unit is split. Another factor is randomly applied to the split.

Example: The factor fertilizer is applied to 0.001 ha plots. Each of the 0.001 ha plot is then split into two, and two different species are planted in each. Fertilizer is applied to the whole plot, and species is applied to the split plot. Species is therefore randomly assigned to the split plot, not to the whole experimental unit.

*Nesting of Factors*

- Treatment levels for one factor may be particular to the level of another factor, resulting in nesting of treatments.

Example, for the first level of fertilizer, we might use medium and heavy thinning, whereas, for the second level of fertilizer, we might use no thinning and light thinning.



*Hierarchical Designs and Sub-Sampling:*

- Commonly in forestry experiments, the experimental unit represents a group of items that we measure. E.g. several pots in a greenhouse, each with several plants germinating from seeds.
- Treatments are randomly assigned to the larger unit (e.g, to each plot not to each seedling). The experimental unit is the larger sized unit.
- May want variance due to the experimental unit (pots in the example) and to units within (plants in the example). These are 1) nested in the treatment; 2) random effects; and 3) hierarchical
- A common variation on hierarchical designs is measuring a sample of items, instead of measuring all items in an experimental unit.

*Introduction of Covariates*

- The initial conditions for an experiment may not be the same for all experimental units, even if blocking is used to group the units.
- Site measures such as soil moisture and temperature, and starting conditions for individuals such as starting height, are then measured (called covariates) along with the response variable
- These covariates are used to reduce the experimental error.
- Covariates are usually interval or ratio scale (continuous).



Designs in use

- The most simple design is one fixed-effects factor, with random allocation of treatments to each experimental unit, with no 1) blocking; 2) sub-sampling; 4) splits; or 5) covariates
- Most designs use combinations of the different variations. For example, one fixed-effects factor, one mixed-effects factor, blocked into three sites, with trees measured within plots within experimental units (sub-sampling/hierarchical), and measures taken at the beginning of the experiment are used as covariates (e.g., initial heights of trees.

Why?

- Want to look at interactions among factors and/or is cheaper to use more than one factor in one experiment than do two experiments.
- Experiments and measurements are expensive – use sampling within experimental units to reduce costs
- Finding homogeneous units is quite difficult: blocking is needed

BUT can end up with problems:

- some elements are not measured,
- random allocation is not possible, or
- measures are correlated in time and/or space.

In this course, start with the simple designs and add complexity.



Main questions in experiments

Do the treatments affect the variable of interest?

For fixed effects: Is there a different between the treatment means of the variable of interest? Which means differ? What are the means by treatment and confidence intervals on these means?

For random effects: Do the treatments account for some of the variance of the variables of interest? How much?

## Completely Randomized Design (CRD)

- Homogeneous experimental units are located
- Treatments are randomly assigned to experimental units
- No blocking is used
- We measure a variable of interest for each experimental unit

## CRD: One Factor Experiment, Fixed Effects

Main questions of interest

Are the treatment means different?

Which means are different?

What are the estimated means and confidence intervals for these estimates?



Notation:

Population: $y_{ij} = \mu + \tau_j + \varepsilon_{ij}$ OR $y_{ij} = \mu_j + \varepsilon_{ij}$

$y_{ij}$ = response variable measured on experimental unit *i* and treatment *j*

*j*=1 to *J* treatments

$\mu$ = the grand or overall mean regardless of treatment

$\mu_j$ = the mean of all measures possible for treatment *j*

$\tau_j$ = the difference between the overall mean of all measures possible from all treatments and the mean of all possible measures for treatment *j*, called the *treatment effect*

$\varepsilon_{ij}$ = the difference between a particular measure for an experimental unit *i*, and the mean for the treatment *j* that was applied to it

$$\varepsilon_{ij} = y_{ij} - \mu_j$$

For the experiment:

$y_{ij} = \bar{y}_{\bullet\bullet} + \hat{\tau}_j + e_{ij}$ OR $y_{ij} = \bar{y}_{\bullet j} + e_{ij}$

$\bar{y}_{\bullet\bullet}$ = the grand or overall mean of all measures from the experiment regardless of treatment; under the assumptions for the error terms, this will be an unbiased estimate of $\mu$

$\bar{y}_{\bullet j}$ = the mean of all measures for treatment *j*; under the assumptions for the error terms, this will be an unbiased estimate of $\mu_j$

$\hat{\tau}_j$ = the difference between the mean of experiment measures for treatment *j* and the overall mean of measures from all treatments; under the error term assumptions, will be an unbiased estimate of $\tau_j$

$e_{ij}$ = the difference between a particular measure for an experimental unit *i*, and the mean for the treatment *j* that was applied to it

$$e_{ij} = y_{ij} - \bar{y}_{\bullet j}$$

$n_j$ = the number of experimental units measured in treatment *j*

$n_T$ = the number of experimental units measured over all treatments = $\sum_{j=1}^{J} n_j$



Example: Fertilization Trial

A forester would like to test whether different site preparation methods result in difference in heights. Twenty five areas each 0.02 ha in size are laid our over a fairly homogeneous area. Five site preparation treatments are randomly applied to 25 plots. One hundred trees are planted (same genetic stock and same age) in each area. At the end of 5 years, the heights of seedlings in each plot were measured, and averaged for the plot.

*i* = a particular 0.02 ha area in treatment *j*, from 1 to 5.

Response variable $y_{ij}$ : 5-year height growth (one average for each experimental unit)

Number of treatments: *J*=5 site preparation methods

$n_T$ = the number of experimental units measured over all treatments = $\sum_{j=1}^{5} n_j$ =25

$n_1 = n_2 = n_3 = n_4 = n_5$ =5 experimental units measured each treatment

Schematic of Layout:

| | | | | |
|---|---|---|---|---|
| 3 | 4 | 4 | 5 | 1 |
| 1 | 2 | 3 | 5 | 2 |
| 2 | 1 | 2 | 4 | 2 |
| 5 | 4 | 3 | 1 | 5 |
| 4 | 3 | 1 | 5 | 3 |

Data Organization and Preliminary Calculations

For easy calculations by hand, the data could be organized in a spreadsheet as:

| Obs: | Treatment, $j$=1 to $J$ | | | | | |
|---|---|---|---|---|---|---|
| $i$=1 to $n_j$ | 1 | 2 | 3 | … | $J$ | |
| 1 | $y_{11}$ | $y_{12}$ | $y_{13}$ | … | $y_{1J}$ | |
| 2 | $y_{21}$ | $y_{22}$ | $y_{23}$ | … | $y_{2J}$ | |
| 3 | $y_{31}$ | $y_{32}$ | $y_{33}$ | … | $y_{3J}$ | |
| … | … | … | … | … | … | |
| $n$ | $y_{n1}$ | $y_{n2}$ | $y_{n3}$ | … | $y_{nJ}$ | |
| Sum | $y_{\cdot 1}$ | $y_{\cdot 2}$ | $y_{\cdot 3}$ | … | $y_{\cdot J}$ | y.. |
| Averages | $\bar{y}_{\bullet 1}$ | $\bar{y}_{\bullet 2}$ | $\bar{y}_{\bullet 3}$ | | $\bar{y}_{\bullet J}$ | $\bar{y}_{\bullet\bullet}$ |

$$y_{\bullet j} = \sum_{i=1}^{n_j} y_{ij} \quad \bar{y}_{\bullet j} = \frac{y_{\bullet j}}{n_j} \quad y_{\bullet\bullet} = \sum_{i=1}^{J}\sum_{i=1}^{n_j} y_{ij} \quad \bar{y}_{\bullet\bullet} = \frac{y_{\bullet\bullet}}{n_T}$$ NO

TE: may not be the same number of observations for each treatment.



Example:

$J$= 5 site preparation treatments randomly applied to n=25 plots.

Response Variable: Plot average seedling height after 5 years

Plot Average Heights (m)

| | Treatments | | | | | Overall |
|---|---|---|---|---|---|---|
| Observation | **1** | **2** | **3** | **4** | **5** | |
| 1 | 4.6 | 4.9 | 4.0 | 3.4 | 4.3 | |
| 2 | 4.3 | 4.3 | 3.7 | 4.0 | 3.7 | |
| 3 | 3.7 | 4.0 | 3.4 | 3.0 | 3.7 | |
| 4 | 4.0 | 4.6 | 3.7 | 3.7 | 3.0 | |
| 5 | 4.0 | 4.3 | 3.0 | 3.4 | 3.4 | |
| SUMS | 20.600 | 22.100 | 17.800 | 17.500 | 18.100 | 96.100 |
| Means | 4.120 | 4.420 | 3.560 | 3.500 | 3.620 | 3.844 |
| $n_j$ | 5 | 5 | 5 | 5 | 5 | 25 |

Example Calculations:

$$\bar{y}_{\bullet 1} = \sum_{i=1}^{5} y_{i1} = (4.6 + 4.3 + 3.7 + 4.0 + 4.3)/5 = 4.12$$

$$\bar{y}_{\bullet\bullet} = \frac{\sum_{j=1}^{5}\sum_{i=1}^{5} y_{ij}}{\sum_{j=1}^{5} n_j} = (20.6 + 22.1 + 17.8 + 17.5 + 18.1)/5 = 96.1/25 = 3.844$$

We then calculate:

1) Sum of squared differences between the observed values and the overall mean (*SSy*):

$$SSy = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_{\bullet\bullet}\right)^2 \quad df = \sum_{j=1}^{J} n_j - 1$$

Also called, sum of squares total (same as in regression)

2) Sum of squared differences between the treatment means, and the grand mean, weighted by the number of experimental units in each treatment ($SS_{TR}$)

$$SS_{TR} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}\right)^2 = \sum_{j=1}^{J} n_j\left(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}\right)^2 \quad df = J - 1$$

3) Sum of squared differences between the observed values for each experimental unit and the treatment means (*SSE*)

$$SSE = \sum_{j=1}^{J}\sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_{\bullet j}\right)^2 \qquad df = n_T - J$$

$$SSy = SS_{TR} + SSE$$



Alternative formulae for the sums of squares that may be easier to calculate are:

$$SSy = \sum_{j=1}^{J}\sum_{i=1}^{n_j} y_{ij}^2 - \frac{y_{\bullet\bullet}^{\ 2}}{n_T}$$

$$SS_{TR} = \sum_{j=1}^{J} n_j \bar{y}_{\bullet j}^2 - \frac{y_{\bullet\bullet}^{\ 2}}{n_T}$$

$$SSE = SSy - SS_{TR}$$

For the example, differences from treatment means (m):

| | Treatments | | | | | Overall |
|---|---|---|---|---|---|---|
| Obs. | **1** | **2** | **3** | **4** | **5** | |
| 1 | 0.480 | 0.480 | 0.440 | -0.100 | 0.680 | |
| 2 | 0.180 | -0.120 | 0.140 | 0.500 | 0.080 | |
| 3 | -0.420 | -0.420 | -0.160 | -0.500 | 0.080 | |
| 4 | -0.120 | 0.180 | 0.140 | 0.200 | -0.620 | |
| 5 | -0.120 | -0.120 | -0.560 | -0.100 | -0.220 | |
| SUMS | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Sum of Squares | | | | | | |
| Error | 0.468 | 0.468 | 0.572 | 0.560 | 0.908 | 2.976 |
| $n_j$ | 5 | 5 | 5 | 5 | 5 | 25 |
| $s^2_j$ | 0.117 | 0.117 | 0.143 | 0.140 | 0.227 | |

Example Calculations:

$$SSE \text{ for treatment } 1 = \sum_{i=1}^{5} (y_{i1} - \bar{y}_{\bullet 1})^2$$

$$= (4.6-4.12)^2 + (4.3-4.12)^2 + (3.7-4.12)^2 + (4.0-4.12)^2 + (4.0-4.12)^2 = 0.468$$

$$s^2_1 = \frac{SSE \text{ for } treatment\ 1}{n_1 - 1} = \frac{0.468}{5-1} = 0.117$$

$$SSE = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet j})^2$$

$$= SSE \text{ for } treatment\ 1 + \text{SSE for } treatment\ 2 + ... + \text{SSE for } treatment\ 5$$

$$= 0.468 + 0.468 + 0.572 + 0.560 + 0.908 = 2.976$$



Differences from grand mean (m)

| | Treatments | | | | | Overall |
|---|---|---|---|---|---|---|
| Obs. | **1** | **2** | **3** | **4** | **5** | |
| 1 | 0.756 | 1.056 | 0.156 | -0.444 | 0.456 | |
| 2 | 0.456 | 0.456 | -0.144 | 0.156 | -0.144 | |
| 3 | -0.144 | 0.156 | -0.444 | -0.844 | -0.144 | |
| 4 | 0.156 | 0.756 | -0.144 | -0.144 | -0.844 | |
| 5 | 0.156 | 0.456 | -0.844 | -0.444 | -0.444 | |
| SUMS | 1.380 | 2.880 | -1.420 | -1.720 | -1.120 | 0.000 |
| Sum of Squares | | | | | | |
| Total | 0.849 | 2.127 | 0.975 | 1.152 | 1.159 | 6.262 |
| $n_j$ | 5 | 5 | 5 | 5 | 5 | 25 |

$$SSy = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{\bullet\bullet})^2$$

$$= SSy \text{ for } treatment\ 1 + \text{SSy for } treatment\ 2 + ... + \text{SSy for } treatment\ 5$$

$$= 0.849 + 02.127 + 0.975 + 1.152 + 1.159 = 6.262$$

Difference between treatment means and grand mean (m)

| | Treatments | | | | | Overall |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | |
| Mean | 4.120 | 4.420 | 3.560 | 3.500 | 3.620 | |
| Difference | 0.276 | 0.576 | -0.284 | -0.344 | -0.224 | 0.000 |
| Sum of Squares | | | | | | |
| Treatment | 0.076 | 0.332 | 0.081 | 0.118 | 0.050 | 3.286 |
| $n_j$ | 5 | 5 | 5 | 5 | 5 | 25 |

Example Calculations:

$$SS_{TR} = \sum_{j=1}^{J} n_j \left(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}\right)^2 = \left(5 \times (4.120 - 3.844)^2\right) + \left(5 \times (4.420 - 3.844)^2\right)$$
$$+ \left(5 \times (3.560 - 3.844)^2\right) + \left(5 \times (3.500 - 3.844)^2\right) + \left(5 \times (3.620 - 3.844)^2\right)$$
$$= 3.286$$



Test for differences among treatment means

The first main question is: Are the treatment means different?

$H_0$: $\mu_1 = \mu_2 = \ldots = \mu_J$

$H_1$: not all the same

*OR:*

$H_0$: $\tau_1 = \tau_2 = \cdots = \tau_J = 0$

$H_1$: not all equal to 0

*OR:*

$H_0$: $(\phi_{TR+}\sigma^2_\varepsilon) / \sigma^2_\varepsilon = 1$

$H_1$: $(\phi_{TR+}\sigma^2_\varepsilon) / \sigma^2_\varepsilon > 1$

Where $\sigma^2_\varepsilon$ is the variance of the error terms;

$\phi_{TR}$ is the effect of the fixed treatments (see page 234 for more details on what this is).

If the treatment does not account for any of the variance in the response variable, then treatment effects are likely all = 0, and all the treatment means are likely all the same.

Using an analysis of variance table:

| Source | df | SS | MS | F | p-value |
|---|---|---|---|---|---|
| Treatment | $J$-1 | $SS_{TR}$ | $MS_{TR}=SS_{TR}/(J-1)$ | $F=MS_{TR}/MSE$ | Prob F> $F_{(J-1),(nT-J),(1-\alpha)}$ |
| Error | $n_T-J$ | $SSE$ | $MSE=SSE/(n_T-J)$ | | |
| Total | $n_T$-1 | $SSy$ | | | |

$$F=\frac{SS_{TR}/(J-1)}{SSE/\sum_{j=1}^{J}(n_j-1)}=\frac{SS_{TR}/(J-1)}{SSE/(n_T-J)}=\frac{MS_{TR}}{MSE}$$

Under $H_0$, and the assumptions of analysis of variance, this follows an F-distribution. If

$$F > F_{(J-1,n_T-J,1-\alpha)}$$

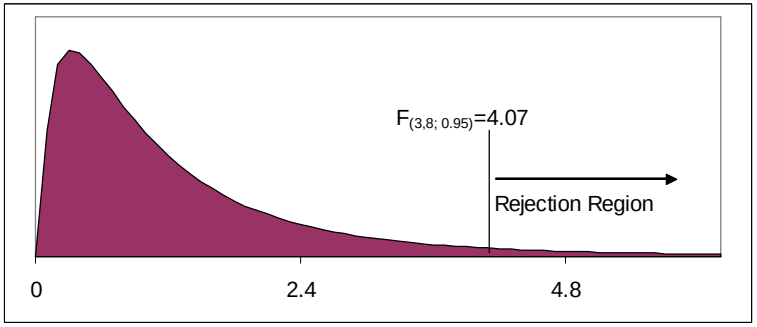
We reject $H_0$ and conclude that there is a difference between the treatment means.

Notice that this is a one-sided test, using 1-α

This is because we are testing if the ratio of variances is > 1.



For example, if we have 4 treatments, and 12 experimental units, and we want α=0.05:




If the calculated F is larger than 4.07, we reject $H_0$: The treatments means are likely different, unless a 5% error has occurred.

OR: We take our calculated F value from our experiment and plot it on this F curve. Then, find the area to the right of this value (p-value). We reject a hypothesis if the probability value (p-value) for the test is less than the specified significance level.

For the example:

If assumptions of ANOVA are met then interpret the F-value.

$H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$H_1$: not all equal

Analysis of Variance (ANOVA) Table:

| Source | df | SS | MS | F | p-value |
|---|---|---|---|---|---|
| Treatment | 5-1=4 | 3.286 | 0.821 | 5.51 | 0.004 |
| Error | 25-5=20 | 2.976 | 0.149 | | |
| Total | 25-1=24 | 6.262 | | | |

If assumptions of ANOVA are met then interpret the F-value.
NOTE: Fcritical for alpha=0.05, df treatment=4 and df error=20 is 2.87.

Since the p-value is very smaller (smaller than alpha=0.05), we reject H0 and conclude that there is a difference in the treatment means. <u>BUT</u> this is only a good test if the assumptions of analysis of variance have been met. Need to check these first (as with regression analysis).



<u>Assumptions regarding the error term</u>

For the estimated means for this experiment to be unbiased estimates of the means in the population, and the MSE to be an unbiased estimate of the variance within each experimental unit, the following assumptions must be met:

1. Observations are independent – not related in time nor in space [independent data]
2. There is normal distribution of the y-values [or the error terms] around each treatment mean [normally distributed]
3. The variances of the y's around each treatment mean [or the error terms] are the same (homogeneous) for all treatment means [equal variance]

Similar to regression:

- a normal probability plot for the error terms can be used to check the assumption of normality, and
- a residual plot can be used to visually check the assumption of equal variance.

OR, these can be tested using (1) normality tests (as with regression); (2) Bartlett's test for equal variances (for more than one factor or for other designs with blocking, etc. this becomes difficult).



Transformations to meet assumptions

Similar to regression:

- logarithmic transformations can be used to equalize variances
- arcsine transformation can be used to transform proportions into normally distributed variables
- rank transformation can be used when data are not normally distributed and other transformations do not "work" [nonparametric analysis of variance using ranks]

Unlike regression you must transform the y-variable

Process:

- do your analysis with the measured response variable
- if assumptions of the error term are not met, transform the y-variable
- do the analysis again and check the assumptions; if not me, try another transformation
- may have to switch to another method: generalized linear models, etc.

Expected values:

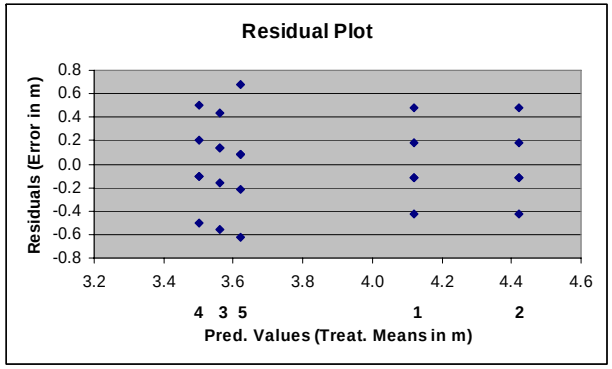
Under the assumptions of analysis of variance, MSE is an unbiased estimate of $\sigma^2_\varepsilon$ and $MS_{TR}$ is an unbiased estimate of $\phi_{TR+}\sigma^2_\varepsilon$. Therefore, this F-test will give the correct probabilities under the assumptions.

This is the same as saying that the expected value of MSE is $\sigma^2_\varepsilon$, and the expected value of $MS_{TR}$ is $\phi_{TR+}\sigma^2_\varepsilon$. The F-test is then a measure of how much larger the value is when the treatment means are accounted for.

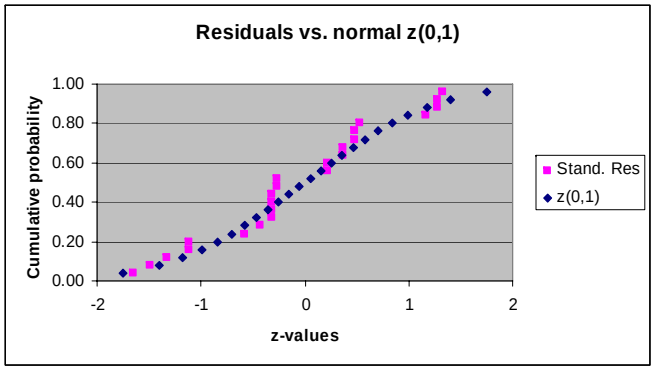


For the example, before interpreting the ANOVA table, we must check assumptions of ANOVA:

Is there equal variance across treatments? (estimated by MSE as 0.149 on our ANOVA table). Using a residual plot and EXCEL:



Are residuals normally distributed? Again using EXCEL:




Where standardized residuals are calculated by:

$$e_i\,(\text{standardized}) = \frac{e_i - 0}{\sqrt{MSE}}$$

Compare these to z-values for a standard normal distribution with a mean of zero and a variance of 1 (z(0,1))



## Differences among particular treatment means

If there are differences among means detected, which means differ?

Can use:

- Orthogonal contrasts – see textbook
- Multiple comparisons

Multiple comparisons (or contrasts):

- Many different types, e.g.
  - T-test for every pair of means; must adjust the alpha level used by dividing by the number of pairs.
  - Scheffe's multiple comparisons
  - Bonferonni's adjustments

- Try to "preserve" the alpha level used to test all the means together (the F-test)

For the example, given that there is a difference among treatment means, which pairs of means differ?

t-test for pairs of means:

- determine the number of pairs possible

$$\binom{5}{2} = \frac{5!}{3!2!} = 10 \quad \text{possible} \quad \text{pairs} \quad \text{of} \quad \text{means}$$

Comparing Treatments 2 (largest estimated mean) versus 4 (smallest estimated mean):

$H_0 : \mu_2 - \mu_4 = 0$ OR $H_0 : \mu_2 = \mu_4$

$H_1 : \mu_2 - \mu_4 \neq 0$

$$t = \frac{(\bar{y}_{\bullet 2} - \bar{y}_{\bullet 4}) - 0}{\sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_4}\right)}}$$

$$t = \frac{(4.4 - 3.5)}{\sqrt{0.149 \times \left(\frac{1}{5} + \frac{1}{5}\right)}} = 3.686$$

Under $H_0$: This follows:

$$t_{1-\alpha/2, n_T - J}$$

Using alpha=0.005 (0.05/10=0.005), for 5 treatments and 25 observations, the t-value is 3.153. Result?

Another way to assess this is to obtain the p-value for t=3.686, with 20 degrees of freedom (25-5).

This is 0.001464. Since this is less than 0.005, we reject $H_0$ and conclude that these two means differ.



<u>Confidence limits for treatment means</u>

Under the assumptions, confidence intervals for each treatment mean can be obtained by:

$$\bar{y}_{\bullet j} \pm t_{(n_T - J), 1-\alpha/2} \sqrt{\frac{MSE}{n_j}}$$

Since MSE estimates the variance that is assumed to be equal,

and the observations are normally distribution and independent.

**For the example:**

$$\bar{y}_{\bullet j} \pm t_{(n_T - 1), 1-\alpha/2} \sqrt{\frac{MSE}{n_j}}$$

$\bar{y}_{\bullet 1} = 4.1 \quad \bar{y}_{\bullet 2} = 4.4 \quad \bar{y}_{\bullet 3} = 3.6 \quad \bar{y}_{\bullet 4} = 3.5 \quad \bar{y}_{\bullet 5} = 3.6$

All $\sqrt{\frac{MSE}{n_j}}$ are all the same since $n_j$ are all equal

$$\sqrt{\frac{0.149}{5}} = 0.173 \qquad t_{20,0.975} = 2.09$$

For treatment 1:

$4.1 \pm 2.09 \times 0.173$ $\qquad$ $4.1 \pm 0.36$

$(3.74, 4.46)$

**Using SAS and R:**

For entry into statistical programs like SAS and R, the data should be organized as:

| Treatment $j$=1 to $J$ | Obs: $i$=1 to $n_j$ | Response |
|---|---|---|
| 1 | 1 | $y_{11}$ |
| 1 | 2 | $y_{21}$ |
| 1 | 3 | $y_{31}$ |
| … | … | ... |
| 1 | $n_1$ | $y_{(n1)\,1}$ |
| 2 | 1 | $y_{12}$ |
| 2 | 2 | $y_{22}$ |
| 2 | 3 | $y_{32}$ |
| … | … | ... |
| 2 | $n_2$ | $Y_{(n2)\,2}$ |
| … | ... | ... |
| $J$ | 1 | $y_{1J}$ |
| $J$ | 2 | $y_{2J}$ |
| $J$ | 3 | $y_{3J}$ |
| … | … | ... |
| $J$ | $n_J$ | $y_{(nJ)\,3}$ |



**For the example, we can put the data into an EXCEL file:**

| Treatment | Observation | AveHt |
|---|---|---|
| 1 | 1 | 4.6 |
| 1 | 2 | 4.3 |
| 1 | 3 | 3.7 |
| 1 | 4 | 4.0 |
| 1 | 5 | 4.0 |
| 2 | 1 | 4.9 |
| 2 | 2 | 4.3 |
| 2 | 3 | 4.0 |
| 2 | 4 | 4.6 |
| 2 | 5 | 4.3 |
| 3 | 1 | 4.0 |
| 3 | 2 | 3.7 |
| 3 | 3 | 3.4 |
| 3 | 4 | 3.7 |
| 3 | 5 | 3.0 |
| 4 | 1 | 3.4 |
| 4 | 2 | 4.0 |
| 4 | 3 | 3.0 |
| 4 | 4 | 3.7 |
| 4 | 5 | 3.4 |
| 5 | 1 | 4.3 |
| 5 | 2 | 3.7 |
| 5 | 3 | 3.7 |
| 5 | 4 | 3.0 |
| 5 | 5 | 3.4 |

Power of the Test:

A Type I error rate ($\alpha$, significance level), the chance of rejecting a null hypothesis when it is true (you reject when the means are actually the same) must be selected. Given:

- a particular number of experimental units
- sizes of the differences between true population means, and
- variation within the experimental units

this will set the Type II error rate ($\beta$), the chance of accepting a null hypothesis when it is false (you fail to reject when the means are actually different)

The power of the test is 1- $\beta$, the probability you will reject the null hypothesis and conclude that there is a difference in means, when there IS a difference between population means.



If the difference between population means (real treatment means) is very large, than a small number of experimental units will result in rejection of the null hypothesis.

If the number of experimental units is very large, then even a small difference between population means will be detected.

If the variation within experimental units is very small, then the difference will be detected, even with a small difference between population means, and even with only a few treatment units.

Statistical Significance is not the same as differences of Practical importance! UNLESS you:

- have some idea of within experimental unit variation from a previous study with the same conditions (e.g., MSE from a previous study)
- know the size of the difference that you wish to detect
- have selected the α level

Then:

You can calculate the number of experimental units per treatment that will result in rejection of $H_0$: when the differences are that large or greater.

Alternatively:

You can calculate the power of the test for an experiment you have already completed.

[see examples in www.forestry.ubc.ca/biometrics course materials for FRST 430/533]



Methods based on maximum likelihood rather than least squares

ML methods can be used when:

- Treatments are random rather than fixed (more on this later)
- Transformations do not result in assumptions being met
- Your dependent variable is a count, or it is a binary variable (e.g., yes or no; dead or alive; present or absent)

## CRD: Two Factor Factorial Experiment, Fixed Effects

Introduction

- Treatments can be combinations of more than one factor
- For 2-factor experiment, have several levels of Factor A and of Factor B
- All levels of Factor A occur for Factor B and vice versa (called a *Factorial Experiment*, or *crossed* treatments)

Example:

- Factor A, (three levels of fertilization: A1, A2, and A3)
- Factor B (four species: B1, B2, B3 and B4)
- Crossed: 12 treatments
- Four replications per treatment for a total of 48 experimental units
- Measured Responses: height growth in mm



*Schematic and Measured Response for the Example:*

| | | | | | |
|---|---|---|---|---|---|
| A1B1=10 | A3B2=25 | A3B4=35 | A2B2=23 | A1B2=14 | A2B3=24 |
| A1B4=24 | A2B2=22 | A1B2=15 | A2B4=28 | A3B3=32 | A3B2=25 |
| A3B2=27 | A1B4=23 | A3B3=29 | A3B2=26 | A1B3=17 | A1B1=11 |
| A3B4=35 | A1B2=13 | A1B4=22 | A1B1=11 | A2B3=24 | A3B3=30 |
| A1B3=19 | A2B1=18 | A2B4=30 | A3B3=31 | A2B3=23 | A1B4=22 |
| A3B1=22 | A2B4=29 | A3B1=23 | A2B1=18 | A1B2=15 | A3B1=23 |
| A2B2=25 | A3B4=37 | A1B1=9 | A3B1=24 | A3B4=36 | A2B4=28 |
| A1B3=17 | A2B1=18 | A2B2=20 | A2B1=18 | A2B3=26 | A1B3=18 |

A1B1=10 indicates that the response variable was 10 for this experimental unit that received Factor A, level 1 and Factor B, level 1. Treatments randomly assigned to the 48 experimental units.

*Organization of data for analysis using a statistics package:*

| A | B | result |
|---|---|---|
| 1 | 1 | 10 |
| 1 | 1 | 11 |
| 1 | 1 | 9 |
| 1 | 1 | 11 |
| 1 | 2 | 15 |
| 1 | 2 | 15 |
| 1 | 2 | 13 |
| 1 | 2 | 14 |
| 1 | 3 | 17 |
| 1 | 3 | 18 |
| 1 | 3 | 17 |
| 1 | 3 | 19 |
| 1 | 4 | 22 |
| 1 | 4 | 23 |
| 1 | 4 | 24 |
| 1 | 4 | 22 |
| 2 | 1 | 18 |
| 2 | 1 | 18 |
| 2 | 1 | 18 |
| 2 | 1 | 18 |
| 2 | 2 | 20 |
| . . . | | |
| 3 | 3 | 32 |
| 3 | 4 | 35 |
| 3 | 4 | 36 |
| 3 | 4 | 37 |
| 3 | 4 | 35 |



*Main questions*

1. Is there an interaction between Factor A and Factor B (fertilizer and species in the example)? Or do the means by Factor A remain the same regardless of Factor B and vice versa?
2. If there is no interaction, is there a difference
    a. Between Factor A means?
    b. Between Factor B means?
3. If there are differences:
    a. If there is an interactions, which treatment means differ?
    b. If there is no interaction, then which levels of Factor A means differ? Factor B means?

Notation, Assumptions, and Transformations

*Models*

Population: $y_{ijk} = \mu + \tau_{Aj} + \tau_{Bk} + \tau_{ABjk} + \varepsilon_{ijk}$

$y_{ijk}$ = response variable measured on experimental unit *i* and factor A level *j*, factor B level *k*

*j*=1 to *J* levels for Factor A; *k*=1 to *K* levels for Factor B

$\mu$ = the grand or overall mean regardless of treatment

$\tau_{Aj}$ = the *treatment effect* for Factor A, level *j*

$\tau_{Bk}$ = the *treatment effect* for Factor B, level *k*

$\tau_{ABjk}$ = the *interaction* for Factor A, level *j* and Factor B, level *k*

$\varepsilon_{ijk}$ = the difference between a particular measure for an experimental unit *i*, and the mean for a treatment:

$$\varepsilon_{ijk} = y_{ijk} - (\mu + \tau_{Aj} + \tau_{Bk} + \tau_{ABij})$$



For the experiment:

$$y_{ijk} = \bar{y}_{\bullet\bullet\bullet} + \hat{\tau}_{Aj} + \hat{\tau}_{Bk} + \hat{\tau}_{ABjk} + e_{ijk}$$

$\bar{y}_{\bullet\bullet\bullet}$ = the grand or overall mean of all measures from the experiment regardless of treatment; under the assumptions for the error terms, this will be an unbiased estimate of $\mu$

$\bar{y}_{\bullet jk}$ = the mean of all measures from the experiment for a particular treatment *jk*

$\bar{y}_{\bullet j\bullet}$ = the mean of all measures from the experiment for a particular level *j* of Factor A (includes all data for all levels of Factor B)

$\bar{y}_{\bullet\bullet k}$ = the mean of all measures from the experiment for a particular level *k* of Factor B (includes all data for all levels of Factor A)

$\hat{\tau}_{Aj}, \hat{\tau}_{Bk}, \hat{\tau}_{ABjk}$ = under the error term assumptions, will be unbiased estimates of corresponding treatment effects for the population

$e_{ijk}$ = the difference between a particular measure for an experimental unit *i*, and the mean for the treatment *jk* that was applied to it

$$e_{ijk} = y_{ijk} - \bar{y}_{\bullet jk}$$

$n_{jk}$ = the number of experimental units measured in treatment *jk*

$n_T$ = the number of experimental units measured over all treatments = $\sum_{k-1}^{K}\sum_{j=1}^{J} n_{jk}$

*Means for the example:*

Factor A: 16 observations per level

A1=16.25, A2=23.38, A3=28.75

Factor B: 12 observations per level

B1=17.08, B2=20.83, B3=24.17, B4=29.08

Treatments (A X B): 4 observations per treatment



*Sums of Squares:*

$SSy = SS_{TR} + SSE$ as with CRD: One Factor. BUT

$SS_{TR}$ is now divided into:

$SS_{TR} = \quad SSA \quad + \quad SSB \quad + \quad SSAB$

*SSy*: The sum of squared differences between the observations and the grand mean:

$$SSy = \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{n_{jk}}\left(y_{ijk} - \bar{y}_{\bullet\bullet\bullet}\right)^2 \quad df = n_T - 1$$

*SSA:* Sum of squared differences between the level means for factor A and the grand mean, weighted by the number of experimental units for each treatment:

$$SSA = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\left(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}\right)^2 \quad df = J - 1$$

*SSB:* Sum of squared differences between the level means for factor B and the grand mean, weighted by the number of experimental units for each treatment:

$$SSB = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\left(\bar{y}_{\bullet\bullet k} - \bar{y}_{\bullet\bullet\bullet}\right)^2 \quad df = K-1$$

*SSAB:* Sum of squared differences between treatment means for *jk* and the grand mean, minus the factor level differences, all weighted by the number of experimental units for each treatment:

$$SSAB = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\left((\bar{y}_{\bullet jk} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{\bullet\bullet k} - \bar{y}_{\bullet\bullet\bullet}) - (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})\right)^2$$

Since some of the estimated grand means cancel out we obtain:

$$SSAB = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\left(\bar{y}_{\bullet jk} - \bar{y}_{\bullet\bullet k} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}\right)^2$$

$$df = (J-1)(K-1)$$



*SSE*: Sum of squared differences between the observed values for each experimental unit and the treatment means:

$$SSE = \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{n_{jk}}\left(y_{ijk} - \bar{y}_{\bullet jk}\right)^2 \qquad df = n_T - JK$$

*Alternative computational formulae:*

$$SSy = \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{n_{jk}} y_{ijk}^2 - \frac{\bar{y}_{\bullet\bullet\bullet}^{\,2}}{n_T} \qquad SSA = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\bar{y}_{\bullet j\bullet}^2 - \frac{\bar{y}_{\bullet\bullet\bullet}^{\,2}}{n_T}$$

$$SS_{TR} = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\bar{y}_{\bullet jk}^2 - \frac{\bar{y}_{\bullet\bullet\bullet}^{\,2}}{n_T} \qquad SSB = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{jk}\bar{y}_{\bullet\bullet k}^2 - \frac{\bar{y}_{\bullet\bullet\bullet}^{\,2}}{n_T}$$

$$SSAB = SS_{TR} - SSA - SSB \qquad SSE = SSy - SS_{TR}$$

**[See Excel Spreadsheet for the Example]**

Assumptions and Transformations:

*Assumptions regarding the error term*

- Must meet assumptions to obtain unbiased estimates of population means, and an unbiased estimate of the variance of the error term (same as CRD: One Factor)
  - independent observations (not time or space related)
  - normality of the errors,
  - equal variance for each treatment.
- Use residual plot and a plot of the standardized errors against the expected errors for a normal distribution to check these assumptions.

*Transformations:*

As with CRD: One Factor, you must transform the y-variable

Process:

- do your analysis with the measured response variable
- if assumptions of the error term are not met, transform the y-variable
- do the analysis again and check the assumptions; if not me, try another transformation
- may have to switch to another method: generalized linear models, etc.



Test for Interactions and Main Effects

The first main question is: Is there an interaction between the two factors?

$H_0$: No interaction
$H_1$: Interaction
*OR:*

$$H_0: (\phi_{AB+}\sigma^2_{\varepsilon)} / \sigma^2_{\varepsilon} = 1$$
$$H_1: (\phi_{AB+}\sigma^2_{\varepsilon}) / \sigma^2_{\varepsilon} > 1$$

Where $\sigma^2_{\varepsilon}$ is the variance of the error terms;

$\phi_{AB}$ is the interaction effect of the fixed treatments.

Using an analysis of variance table:

| Source | df | SS | MS | F | p-value |
|---|---|---|---|---|---|
| A | $J$-1 | $SSA$ | $MSA = SSA/(J\text{-}1)$ | $F = MSA/MSE$ | Prob F> $F_{(J\text{-}1),(dfE),\ 1-\alpha}$ |
| B | $K$-1 | $SSB$ | $MSB = SSB/(K\text{-}1)$ | $F = MSB/MSE$ | Prob F> $F_{(K\text{-}1),(dfE),1-\alpha}$ |
| A X B | ($J$-1)($K$-1) | $SSAB$ | $MSAB = SSAB/(J\text{-}1)(K\text{-}1)$ | $F = MSAB/MSE$ | Prob F> $F_{dfAB,dfE,,1-\alpha}$ |
| Error | $n_T$-$JK$ | $SSE$ | $MSE = SSE/(n_T\text{-}J)$ | | |
| Total | $n_T$-1 | $SSy$ | | | |

| Source | df | MS | E[MS] |
|---|---|---|---|
| A | $J$-1 | $MSA$ | $\sigma_\varepsilon^2 + \phi_A$ |
| B | $K$-1 | $MSB$ | $\sigma_\varepsilon^2 + \phi_B$ |
| A X B | ($J$-1)($K$-1) | $MSAB$ | $\sigma_\varepsilon^2 + \phi_{AB}$ |
| Error | $n_T$-$JK$ | $MSE$ | $\sigma_\varepsilon^2$ |
| Total | $n_T$-1 | | |

See Neter et al., page 826, Table 19.8 for details on expected mean squares; $\phi$ is used here to represent fixed effects.



For the interactions:

$$F = \frac{SSAB/(J-1)(K-1)}{SSE/(n_T - JK)} = \frac{MSAB}{MSE}$$

- Under $H_0$, this follows $F_{df1,df2,\ 1-\alpha}$ where df1 is from the numerator ($J$-1)($K$-1), and df2 is from the denominator ($n_T$-$JK$)
- If the F calculated is greater than the tabular F, or if the p-value for F calculated is less than $\alpha$, reject $H_0$.
    - The means of Factor A are influenced by the levels of Factor B and the two factors cannot be interpreted separately.
    - Graph the means of all treatments
    - Conduct multiple comparisons all treatments (rather then on means of each Factor, separately
    - Not as much power (reject $H_0$ when it is false), if this occurs.

**If there are no interactions between the factors, we can look at each factor separately – fewer means, less complicated.**

Factor A:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_J$$

*OR:*

$$H_0: (\phi_{A+}\sigma^2_\varepsilon)/\sigma^2_\varepsilon = 1$$
$$H_1: (\phi_{A+}\sigma^2_\varepsilon)/\sigma^2_\varepsilon > 1$$

Where $\sigma^2_\varepsilon$ is the variance of the error terms;

$\phi_A$ is fixed effect for Factor A.



From the ANOVA table:

$$F = \frac{SSA/(J-1)}{SSE/(n_T - JK)} = \frac{MSA}{MSE}$$

- Under $H_0$, this follows $F_{df1,df2,\ 1-\alpha}$ where df1 is from the numerator ($J$-1) and df2 is from the denominator ($n_T$-$JK$)
- If the F calculated is greater than the tabular F, or if the p-value for F calculated is less than $\alpha$, reject $H_0$.
    - The means of Factor A in the population are likely not all the same
    - Graph the means of Factor A levels
    - Conduct multiple comparisons between means for the $J$ levels of Factor A, separately

The analysis and conclusions would follow the same pattern for Factor B.

*Analysis of Variance Table Results for the Example*

| Source | Degrees of Freedom | Sum of Squares | Mean Squares | F | p |
|---|---|---|---|---|---|
| A | 2 | 1258.17 | 629.08 | 514.70 | <0.0001 |
| B | 3 | 934.75 | 311.58 | 254.93 | <0.0001 |
| A X B | 6 | 17.00 | 2.836 | 2.32 | 0.0539 |
| Error | 36 | 44.00 | 1.22 | | |
| Total | 47 | 2253.92 | | | |

If assumptions met, (residuals are independent, are normally distributed, and have equal variances among treatments), we can interpret the results.



*Interpretation using $\alpha$ =0.05:*

- No significant interaction (p=0.0539); we can examine species and fertilizer effects separately.
- Are significant differences between the three fertilizer levels of Factor A (p<0.0001), and between the four species of Factor B (p<0.0001).
- The mean values based on these data are:

  A1=16.25, A2=23.38, A3=28.75

  B1=17.08, B2=20.83, B3=24.17, B4=29.08

Did not have to calculate these for each of the 12 treatments since there is no interaction.

Further analyses, for each Factor separately:

- Scheffé's test for multiple comparisons, could then be used to compare and contrast Factor level means.
  - The number of observations in each factor level are: 16 for Factor A, and 12 for Factor B
  - Use the MSE for both Factor A and for Factor B (denominator of their F-tests)
- t-tests for each pair of means could be used instead.
  - Again, use MSE, and 16 observations for Factor A versus 12 for Factor B
  - Must split alpha level used in the F-tests by the number of pairs



<u>Factor A: t-tests for pairs of means</u>

Determine the number of pairs possible

$$\binom{3}{2} = \frac{3!}{1!2!} = 3 \text{ possible pairs of means}$$

Use a significance level of 0.05/3 pairs=0.017 for each t-test

Comparing Factor Levels 1 and 2: A1 vs. A2

$$H0 : \mu_{1\bullet} - \mu_{2\bullet} = 0 \qquad H1 : \mu_{1\bullet} - \mu_{2\bullet} \neq 0$$

$$t = \frac{(\bar{y}_{\bullet 1} - \bar{y}_{\bullet 2}) - 0}{\sqrt{MSE\left(\frac{1}{\sum_{k=1}^{K} n_{1k}} + \frac{1}{\sum_{k=1}^{K} n_{4k}}\right)}}$$

$$t = \frac{(16.25 - 23.38)}{\sqrt{1.22 \times \left(\frac{1}{16} + \frac{1}{16}\right)}} = -18.258$$

Critical t value from a probability table for:

- df(error) = 36 based on ( $n_T - JK$), and 0.017 significance level (For $\alpha$ =0.05 use 0.05/3 pairs for each t-test), 2-sided test
- Using an EXCEL function: =tinv(0.017,36), returns the value of 2.50 (this assumes a 2-sided test).
- Since the absolute value of the calculated t is greater than 2.50 we reject H0.

OR

- enter your t-value, df (error), and 2 (for 2-sided) into the EXCEL function =tdist(18.258,36,2)
- Returns a p-value of <0.000. (NOTE that you must enter the positive value, and the p-value is for the two "ends" (area greater than 18.258 plus area less than -18.258)
- Since p<0.017, we reject H0

The mean of treatment A1 differs from the mean of A2.

<u>For Factor B</u>

- Recalculate the number of possible pairs for 4 factor levels

  (will be 6 pairs; divide alpha by this for each test )

- The observations per factor level is 12, rather than 16

- Df(error) and MSE are the same as for Factor A.



*A Different Interpretation using α =0.10:*

- There is a significant interaction (p=0.0539) using α =0.10;

  cannot interpret main effects (A and B) separately.

- The mean values based on these data are: [Excel]

  A1B1=10.25 A1B2=14.25 A1B3= 17.75 A1B4= 22.75
  A2B1=18.00 A2B2=22.50 A2B3= 24.25 A2B4=28.75
  A3B1= 23.00 A3B2=25.75 A3B3=30.50 A3B4=35.75

**12 mean values** as there is a significant interaction

Further analyses:

- Scheffé's test for multiple comparisons (or others) could then be used to compare and contrast treatment means (pairs or other groupings of means). The number of observations in each treatment are 4 [lower power than if there was no interaction], and use the MSE.

- Using t-tests for pairs of means, the number of observations are 4 for each *jk* treatment, use the MSE, and recalculate the number of possible pairs out of 12 treatments (will be 66 pairs! Retaining $\alpha = 0.10$, we would use $0.10/66 = 0.0015$ for each t-test )



<u>Confidence limits for factor level and treatment means</u>

Treatment means:

$$\bar{y}_{\bullet jk} \pm t_{(n-JK),1-\alpha/2}\sqrt{\frac{MSE}{n_{jk}}}$$

Factor A means:

$$\bar{y}_{\bullet j\bullet} \pm t_{(n-JK),1-\alpha/2}\sqrt{\frac{MSE}{\sum_{k=1}^{K} n_{jk}}}$$

Factor B means:

$$\bar{y}_{\bullet\bullet k} \pm t_{(n-JK),1-\alpha/2}\sqrt{\frac{MSE}{\sum_{j=1}^{J} n_{jk}}}$$

## CRD: Random and Mixed Effects

Factors in experiments can be:

- Fixed: all levels of interest are included in the experiment; we are mostly interested in testing differences and estimating means for factor levels
- Random: levels are randomly selected; not all levels of interest are included; we are mostly interested in the variance of the response variable that is DUE TO the factor
- Mixed: When there is more than one factor, there may be a mixture, with some factors that are fixed-effects and others that are mixed-effects
- Often, it is difficult to make the distinction!



Examples:

We are interested in height growth for different families (genetic stock). We select 4 families from all possible families, and include these in the experiment. Then, we get an estimate of the variance in the height growth due to changes in genetics. [One random-effect factor – family]

We are interested in seedling success depending on species and soil moisture. We select 3 species out of 12 possible species, and include moisture levels of low, medium, and high. The species are considered random-effects (we are interested estimating the variance in seedling success due to species). The moisture levels are fixed-effects (we are only interested in these specific levels that we might apply in a greenhouse to generate seedlings).

- This will effect
  - the expected values of the Mean squares, and then, the F-tests that are used
  - Tests that are done following the overall F-test
  - The conclusions that are made

For J levels of Factor A and K levels of Factor B, we have the following model:

$$y_{ijk} = \bar{y}_{\bullet\bullet\bullet} + \hat{\tau}_{A\,j} + \hat{\tau}_{Bk} + \hat{\tau}_{AB\,jk} + e_{ijk}$$

Possibilities:

- Both are fixed (covered already)
- Both are random
- One is fixed and one is random



Expected Mean Square Values Comparison:

| Mean Square | Model I Both A and B are Fixed | Model II Both A and B are Random | Model III A is Fixed B is Random |
|---|---|---|---|
| A (MSA) | $\sigma_\varepsilon^2 + \phi_A{}^*$ | $\sigma_\varepsilon^2 + nK\sigma_A^2 + n\sigma_{AB}^2$ | $\sigma_\varepsilon^2 + \phi_A + n\sigma_{AB}^2$ |
| B (MSB) | $\sigma_\varepsilon^2 + \phi_B$ | $\sigma_\varepsilon^2 + nJ\sigma_B^2 + n\sigma_{AB}^2$ | $\sigma_\varepsilon^2 + nJ\sigma_B^2$ |
| A X B (MSAB) | $\sigma_\varepsilon^2 + \phi_{AB}$ | $\sigma_\varepsilon^2 + n\sigma_{AB}^2$ | $\sigma_\varepsilon^2 + n\sigma_{AB}^2$ |
| Error (MSE) | $\sigma_\varepsilon^2$ | $\sigma_\varepsilon^2$ | $\sigma_\varepsilon^2$ |

$^*\sigma_\varepsilon^2 + \phi_A = \sigma_\varepsilon^2 + nK\dfrac{\sum_{j=1}^{J}\tau_{A\,j}^2}{J-1}$ when the number of observations (n) are all equal.

F-tests

- Sums of squares, means squares, etc are calculated the same for all three types of models
- Assumptions: Same are for fixed-effects models
- Change the F-test, so that the numerator differs from the denominator ONLY in the item that you are testing
- For means tests, use the same denominator as used for the F-test (e.g., instead of MSE for Model III, use MSAB when testing for differences in Factor A means)
- Not really relevant to test for differences among means of a Random-effects factor as we are interested in the variance due to that factor



Maximum Likelihood as an Alternative for Random-Effects and Mixed-Effects Models

- For mixed models, maximum likelihood may be a better approach than least squares methods.
- Why? Better estimates of the variances than least squares methods.

PROC MIXED in SAS

*lme* part of the package *nlme* in *R*

## Randomized Complete Block (RCB) With One Fixed-Effects Factor

Freese Handbook, page 34.

Introduction and Example

- In RCB, treatments are assigned randomly, but only within blocks of treatments
- Restricting randomization of treatments to within blocks (often called sites or trials) is used when the experimental units can be grouped by another variable that may impact the results
- In field experiments with large experimental units, blocking is often very useful in reducing error variance with only a small reduction in error degrees of freedom
- Blocks are most often random effects (we are interested in the variance due to blocks)
- The interest with RCB is with the factor, not with the blocks; the blocks are simply used to reduce the variability among experimental units



*Example: Randomized Block Design (RCB)*, with Factor A (six levels of fertilization: A1 to A6), and two sites. Randomization of Factor A is restricted to within sites.

| Site 1 | | | Site 2 | |
|---|---|---|---|---|
| A1 = 9 | A6=21 | | A4=25 | A3=19 |
| A3=15 | A2=12 | | A1=12 | A5=27 |
| A5=20 | A4=17 | | A2=16 | A6=29 |

Response variable: biomass of grasses and herbs (kg)

2 observations per treatment – 1 in each site

*Organization of data for analysis using a statistics package:*

| Site | Treatment | yjk |
|---|---|---|
| 1 | A1 | 9 |
| 1 | A2 | 12 |
| 1 | A3 | 15 |
| 1 | A4 | 17 |
| 1 | A5 | 20 |
| 1 | A6 | 21 |
| 2 | A1 | 12 |
| 2 | A2 | 16 |
| 2 | A3 | 19 |
| 2 | A4 | 25 |
| 2 | A5 | 27 |
| 2 | A6 | 29 |



Main questions of interest:

- Are the treatment means different?
- Which means are different?
- What are the estimated means and confidence intervals for these estimates?

As for CRD with one factor

The organization of the data is the same for CRD with **two** factors as with RCB, BUT the **interpretation** differs:

- It is assumed that there is no interaction between the blocks and the treatments. Not really appropriate to check this since the randomization of treatments is restricted to within blocks
- Blocks are usually considered random-effects; want to remove the effects of blocks from the analysis

Notation

Population: $y_{jk} = \mu + +\tau_{Bj} + \tau_{Ak} + \varepsilon_{jk}$

$y_{jk}$ = response variable measured on block $j$ and treatment $k$

$j$=1 to $J$ blocks; $k$=1 to $K$ treatments

$\mu$ = the grand or overall mean regardless of treatment or block

$\tau_{Ak}$ = the *treatment effect* for $k$

$\tau_{Bj}$ = the *block effect* for block $j$

$\varepsilon_{jk}$ = is actually an interaction term between block and treatment, defined as:

$$\varepsilon_{jk} = y_{jk} - (\mu + \tau_{Ak} + \tau_{Bj})$$



For the experiment:

$$y_{jk} = \bar{y}_{\bullet\bullet} + \hat{\tau}_{Bj} + \hat{\tau}_{Ak} + e_{jk}$$

$\bar{y}_{\bullet\bullet}$ = the grand or overall mean of all measures from the experiment regardless of treatment; under the assumptions for the error terms, this will be an unbiased estimate of $\mu$

$\bar{y}_{j\bullet}$ = the mean of all measures from the experiment for a particular block $j$ (includes all data for all levels of the treatment)

$\bar{y}_{\bullet k}$ = the mean of all measures from the experiment for a particular treatment $k$ over all blocks

$\hat{\tau}_{Ak}, \hat{\tau}_{Bj}$ = under the error term assumptions, will be unbiased estimates of corresponding treatment effects for the population

$e_{jk}$ = is defined as:

$$\begin{aligned} e_{jk} &= (y_{jk} - \bar{y}_{\bullet\bullet}) - (\bar{y}_{j\bullet} - \bar{y}_{\bullet\bullet}) - (\bar{y}_{\bullet k} - \bar{y}_{\bullet\bullet}) \\ &= y_{jk} - \bar{y}_{j\bullet} - \bar{y}_{\bullet k} + \bar{y}_{\bullet\bullet} \end{aligned}$$

$J$= number of blocks and also the number of measures (experimental units) for treatment $k$

$KJ$ = total number of experimental units on which the response was measured

*Sums of Squares:*

$$SSy = SS_{BLK} + SS_{TR} + SSE$$

*SSy*: The sum of squared differences between the observations and the grand mean:

$$SSy = \sum_{k=1}^{K}\sum_{j=1}^{J}\left(y_{jk} - \bar{y}_{\bullet\bullet}\right)^2 \qquad df = JK - 1$$

$SS_{TR}$ : Sum of squared differences between the treatment means, and the grand mean, weighted by the number of blocks (experimental units in each treatment)

$$SS_{TR} = \sum_{k=1}^{K} J\left(\bar{y}_{\bullet k} - \bar{y}_{\bullet\bullet}\right)^2 \qquad df = K - 1$$

$SS_{BLK}$ : Sum of squared differences between the block means, and the grand mean, weighted by the number of treatments (experimental units in each block)

$$SS_{BLK} = \sum_{j=1}^{J} K\left(\bar{y}_{j\bullet} - \bar{y}_{\bullet\bullet}\right)^2 \qquad df = J - 1$$



*SSE:* sum of squared differences between the observation and the grand mean plus the treatment and block effects.

$$SSE = SSy - SS_{TR} - SS_{BLK} \qquad df = (J-1)(K-1)$$

*Alternative computational formulae:*

$$SSy = \sum_{k=1}^{K}\sum_{j=1}^{J} y_{jk}^2 - \frac{y_{\bullet\bullet}^{\ 2}}{JK}$$

$$SS_{TR} = J\sum_{k=1}^{K} \bar{y}_{\bullet k}^2 - \frac{y_{\bullet\bullet}^{\ 2}}{JK} \qquad SS_{BLK} = K\sum_{j=1}^{J} \bar{y}_{j\bullet}^2 - \frac{y_{\bullet\bullet}^{\ 2}}{JK}$$

$$SSE = SSy - SS_{TR} - SS_{BLK}$$

Assumptions and Transformations:

- Must meet assumptions for the error term to obtain unbiased estimates of population means, and an unbiased estimate of the variance of the error term
  - independent observations (not time or space related)
  - normality of the errors,
  - equal variance for each treatment.
- Use residual plot and a plot of the standardized errors against the expected errors for a normal distribution to check these assumptions. To meet assumptions you might have to transform the y-variable, as with other designs

Differences among treatment means

The main question is: Is there a difference between treatment means:

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_K$$

*OR:*

$$H_0: (\phi_{TR+}\sigma^2_{\varepsilon)} / \sigma^2_{\varepsilon} = 1$$
$$H_1: (\phi_{TR+}\sigma^2_{\varepsilon}) / \sigma^2_{\varepsilon} > 1$$

Where $\sigma^2_{\varepsilon}$ is the variance of the error terms;

$\phi_{TR}$ is fixed effect for the treatments.



Using an analysis of variance table:

| Source | df | SS | MS | F | p-value |
|---|---|---|---|---|---|
| Block | $J$-1 | $SS_{BLK}$ | $MSA = SS_{BLK}/(J\text{-}1)$ | | |
| Treat. | $K$-1 | $SS_{TR}$ | $MS_{TR} = SS_{TR}/(K\text{-}1)$ | $F = MS_{TR}/MSE$ | Prob F> $F_{(K\text{-}1),(dfE),1-\alpha}$ |
| Error | $(J\text{-}1)(K\text{-}1)$ | $SSE$ | $MSE = SSE/(J\text{-}1)(K\text{-}1)$ | | |
| Total | $JK$-1 | $SSy$ | | | |

| Source | df | MS | E[MS] |
|---|---|---|---|
| Block | $J$-1 | $MS_{BLK}$ | $\sigma_{\varepsilon}^2 + K\sigma_{BLK}^2$ |
| Treat. | $K$-1 | $MS_{TR}$ | $\sigma_{\varepsilon}^2 + \phi_{TR}$ |
| Error | $(J\text{-}1)(K\text{-}1)$ | $MSE$ | $\sigma_{\varepsilon}^2$ |
| Total | $n_T$-1 | | |

NOTE: Neter et al., assume blocks are fixed rather than random

$\phi$ is used here to represent fixed effects and $\sigma^2$ is used to represent random effects.

From the ANOVA table:

$$F = \frac{SS_{TR}/(K-1)}{SSE/(J-1)(K-1)} = \frac{MS_{TR}}{MSE}$$

- Under $H_0$, this follows $F_{df1,df2,\,1-\alpha}$ where df1 is from the numerator ($K$-1) and df2 is from the denominator ($J$-1) ($K$-1)
- If the F calculated is greater than the tabular F, or if the p-value for F calculated is less than $\alpha$, reject $H_0$, the means of treatments in the population are likely not all the same

*Further analyses:*
Can do multiple comparisons between treatments using MSE and using J (number of blocks) as the number of observations per treatment. OR Can use t-tests of pairs of means -- must divide alpha by the number of possible pairs

*Confidence limits for treatment means*
Treatment means:

$$\bar{y}_{\bullet k} \pm t_{(dfE),1-\alpha/2}\sqrt{\frac{MSE}{J}}$$

Divide by J, since each block has a measure for each treatment.



**Randomized Block Design with other experiments**

RCB with Two Fixed Factors

- Within each block. treatments are randomly located to each experimental unit, but each treatment is a combination of two factors

*Example: Randomized Block Design (RCB)*, with three types of food (Factor A: A1 to A3), two species of fish (Factor B) and two labs (blocks). Randomization of treatments (e.g., A1, B2) is restricted to within labs.

| Lab 1 | | | Lab 2 | |
|---|---|---|---|---|
| A1B1 = 6 | A1B2=5 | | A3B1=11 | A3B2=12 |
| A3B1=10 | A2B2=8 | | A1B1=4 | A2B2=9 |
| A2B1=7 | A3B2=12 | | A2B1=8 | A1B2=5 |

Response variable: weight gain of fish (kg)

Experimental unit: one tank of fish; 6 tanks in each lab

*Organization of data for analysis using a statistics package:*

| Site | A Food | B Species | yijk |
|---|---|---|---|
| 1 | A1 | B1 | 6 |
| 1 | A1 | B2 | 5 |
| 1 | A2 | B1 | 8 |
| 1 | A2 | B2 | 7 |
| 1 | A3 | B1 | 10 |
| 1 | A3 | B2 | 12 |
| 2 | A1 | B1 | 4 |
| 2 | A1 | B2 | 5 |
| 2 | A2 | B1 | 9 |
| 2 | A2 | B2 | 8 |
| 2 | A3 | B1 | 11 |
| 2 | A3 | B2 | 12 |

*Main questions of interest—same as for RCB:*

- Is there an interaction between factors? If not, is there a difference between means for Factor A? Factor B? Which means are different? What are the estimated means and confidence intervals for these estimates?



- We are not really interested in the blocks – just used to reduce the amount of variation

*Models*

The model is a mixture between a single factor RCB and a 2-factor CRD; interpretation is more difficult

- Blocks are usually random not fixed factors
- Blocks are used to reduce variability within treatments; not of interest on their own

Population: $y_{jkl} = \mu + \tau_{BLK\,j} + \tau_{A\,k} + \tau_{Bl} + \tau_{A\,Bkl} + \varepsilon_{jkl}$

$y_{jkl}$ = response variable measured on block $j$ and treatment $kl$

$j$=1 to $J$ blocks; $k$=1 to $K$ levels for Factor A; $l$=1 to $L$ levels for Factor B

Definition of terms follows other designs

ANOVA: Blocks Random, Factor A and Factor B are Fixed

| Source | df | SS | MS | F ??? correct? |
|---|---|---|---|---|
| BLK. | $J$-1 | $SS_{BLK}$ | $MS_{BLK}= SS_{BLK}/(J\text{-}1)$ | $F= MS_{BLK}/MSE$ |
| Factor A | $K$-1 | $SS_A$ | $MS_A= SS_A/(K\text{-}1)$ | $F= MS_A/MS_{BXT}$ |
| Factor B | $L$-1 | $SS_B$ | $MS_B= SS_B/(L\text{-}1)$ | $F= MS_B/MS_{BXT}$ |
| A X B | $(K\text{-}1)(L\text{-}1)$ | $SS_{AXB}$ | $MS_{AXB}=SS_{AXB} / (K\text{-}1)(L\text{-}1)$ | $F= MSAB/MSE$ |
| Error | $(J\text{-}1)(KL\text{-}1)$ | $SSE$ | $MSE= SSE/ (J\text{-}1)(KL\text{-}1)$ | |
| Total | $n_T$-1 | $SSy$ | | |

| Source | df | MS | p-value | E[MS] |
|---|---|---|---|---|
| BLK. | $J$-1 | $MS_{BLK}$ | Prob F> $F_{(J\text{-}1),(dfE),\ 1-\alpha}$ | $\sigma_\varepsilon^2 + KL\sigma_{BLK}^2$ |
| A | $K$-1 | $MS_A$ | Prob F> $F_{(K\text{-}1),(dfBXT),1-\alpha}$ | $\sigma_\varepsilon^2 + \phi_A$ |
| B | $L$-1 | $MS_B$ | Prob F> $F_{(L\text{-}1),(dfBXT),1-\alpha}$ | $\sigma_\varepsilon^2 + \phi_B$ |
| AXB | $(J\text{-}1)(L\text{-}1)$ | $MS_{AXB}$ | Prob F> $F_{dfAXB,dfE,,1-\alpha}$ | $\sigma_\varepsilon^2 + \phi_{A\times B}$ |
| Error | $(J\text{-}1)(KL\text{-}1)$ | $MSE$ | | $\sigma_\varepsilon^2$ |
| Total | $n_T$-1 | | | |

$\phi$ is used here to represent fixed effects.

[see www.forestry.ubc.ca/biometrics course notes for FRST 430/533 for more on this design]



RCB with One fixed, one random factor

- Within each block treatments are randomly located to each experimental unit, but each treatment is a combination of two factors
- For one factor, we are interested in comparing treatment means [all levels are in the experiment]
- For the other factor, we are interested in obtaining an estimate of the variance of the response variable that is due to that factor [some levels are in the experiment]
- Must take care with correct F tests and hypotheses.

*Example: Randomized Block Design (RCB)*, with four sites, three types of fertilizer (Factor A: A1 to A3, fixed effects), two genetic families of pine trees (Factor B, random effects) .

Incomplete Block Design

- Like RCB, BUT there are not enough experimental units in each block to have every treatment in each block – incomplete
- For example:

We have 2 sites. There are 4 experimental units in each site. However, we have 5 treatments! There are not enough experimental units in site 1 to have all 5 treatments, nor is there enough experimental units in site 2 to have all 5. (REF: Chapter 28 of textbook)



RCB with replicates in each block

- Within each block there are several replicates of each treatment
- Sometimes called "Generalized RCB"

*Example: Randomized Block Design (RCB)*, with Factor A (three types of food: A1 to A3), and two labs (blocks).

Randomization of Factor A is restricted to within labs.

| Lab 1 | | Lab 2 | |
|---|---|---|---|
| A1 = 6 | A1=5 | A3=11 | A3=12 |
| A3=10 | A2=8 | A1=4 | A2=9 |
| A2=7 | A3=12 | A2=8 | A1=5 |

Response variable: weight gain of fish (kg)

Experimental unit: one tank of fish; 6 tanks in each lab

*Organization of data for analysis using a statistics package:*

| Site | Treatment | Replicate | yijk |
|---|---|---|---|
| 1 | A1 | 1 | 6 |
| 1 | A1 | 2 | 5 |
| 1 | A2 | 1 | 8 |
| 1 | A2 | 2 | 7 |
| 1 | A3 | 1 | 10 |
| 1 | A3 | 2 | 12 |
| 2 | A1 | 1 | 4 |
| 2 | A1 | 2 | 5 |
| 2 | A2 | 1 | 9 |
| 2 | A2 | 2 | 8 |
| 2 | A3 | 1 | 11 |
| 2 | A3 | 2 | 12 |

*Main questions of interest—same as for RCB:*

- Are the treatment means different? Which means are different? What are the estimated means and confidence intervals for these estimates?



*Models*

Population: $y_{ijk} = \mu + \tau_{BLK\ j} + \tau_{TR\,k} + \tau_{BLK \times TR\ jk} + \varepsilon_{ijk}$

$y_{ijk}$ = response variable measured on experimental unit *I* in block *j* and treatment *k*

*j*=1 to *J* blocks; *k*=1 to *K* treatments; *i*=1 to *n* replicates

$\mu$ = the grand or overall mean regardless of treatment or block

$\tau_{BLK\ j}$ = the *block effect* for *j*

$\tau_{TRk}$ = the *treatment effect* for block *k*

$\tau_{BLK \times TR\ jk}$ = the *interaction effect* between block *j* and treatment *k*

$\varepsilon_{ijk}$ = is error term, specific to observation *i*

The assumptions for the error term are the same as for other designs.

Using an analysis of variance table: Blocks Random, Treatments Fixed

| Source | df | SS | MS | F |
|---|---|---|---|---|
| BLK. | $J$-1 | $SS_{BLK}$ | $MS_{BLK}= SS_{BLK}/(J\text{-}1)$ | $F= MS_{BLK}/MSE$ |
| TR. | $K$-1 | $SS_{TR}$ | $MS_{TR}= SS_{TR}/(K\text{-}1)$ | $F= MS_{TR}/MS_{BXT}$ |
| BLK X TR | $(J\text{-}1)(K\text{-}1)$ | $SS_{BXT}$ | $MS_{BXT}= SS_{BXT}/(J\text{-}1)(K\text{-}1)$ | $F= MSBT/MSE$ |
| Error | $n_T\text{-}JK$ | $SSE$ | $MSE= SSE/(n_T\text{-}JK)$ | |
| Total | $n_T$-1 | $SSy$ | | |

| Source | df | MS | p-value | E[MS] |
|---|---|---|---|---|
| BLK. | $J$-1 | $MS_{BLK}$ | Prob F> $F_{(J\text{-}1),(dfE),\ 1-\alpha}$ | $\sigma_\varepsilon^2 + Kn\sigma_{BLK}^2$ |
| TR. | $K$-1 | $MS_{TR}$ | Prob F> $F_{(K\text{-}1),(dfBXT),1-\alpha}$ | $\sigma_\varepsilon^2 + n\sigma^2_{B\times T} + \phi_{TR}$ |
| BLK X TR | $(J\text{-}1)(K\text{-}1)$ | $MS_{BXT}$ | Prob F> $F_{dfBXT,dfE,,1-\alpha}$ | $\sigma_\varepsilon^2 + n\sigma^2_{B\times T}$ |
| Error | $n_T\text{-}JK$ | $MSE$ | | $\sigma_\varepsilon^2$ |
| Total | $n_T$-1 | | | |

$\phi$ is used here to represent fixed effects.



From the ANOVA table:

1. Check for a treatment by block interaction first (hopefully not there!).
2. If there is no block by treatment interaction, F test for differences among treatments.
3. If there are differences, use multiple comparisons (e.g., pairs of means t-tests) to see which treatments differ. Remember to correct alpha by dividing by the number of pairs (i.e., tests done).

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]

## Latin Square (LS) With One Fixed-Effects Factor

REF: Neter et al., Chapter 26 (White-newest edition) or Chapter 28 (Blue – older edition in the library)

Introduction and Example

- In RCB, treatments are assigned randomly, but only within blocks of treatments; blocking is in "one" direction
- The Latin Square Design extends grouping of experimental units to two variables. For example, two sites may represent north versus south facing stands, and there might be a moisture gradient within sites
- Treatments are randomly assigned in two directions; treatment appears once in every row and every column



*Example:*

*Response variable:* average 5-year height growth in each experimental unit (plot) in cm

*Treatments:* four different species, A1 to A4

*Nutrient Gradient* from East to West; *Moisture Gradient* from North to South

| | | | | | Means |
|---|---|---|---|---|---|
| | A2=40 | A1=35 | A4=53 | A3=47 | 43.75 |
| | A4=48 | A3=46 | A2=39 | A1=34 | 41.75 |
| | A1=27 | A4=53 | A3=45 | A2=41 | 41.50 |
| | A3=44 | A2=39 | A1=31 | A4=52 | 41.50 |
| Means | 39.75 | 43.25 | 42.00 | 43.50 | 42.125 |

Treatment Means:

A1: 31.75 A2: 39.75 A3: 45.50 A4: 51.50

16 experimental units

Analysis of Variance Table: Assuming that all are fixed-effects.

| *Source* | *Df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Treatment | $K$-1 | $SS_{TR}$ | $MS_{TR}$ | $MS_{TR}/MSE$ |
| Row | $J$-1 | $SS_R$ | $MS_R$ | $MS_R/MSE$ |
| Column | $L$-1 | $SS_C$ | $MS_C$ | $MS_C/MSE$ |
| Error | ($K$-1)($J$-2) | $SSE$ | $MSE$ | |
| Total | $JK$-1 | $SS_y$ | | |

Gypotheses and Tests:

Treatment: H0: $\mu_{\bullet 1\bullet} = \mu_{\bullet 2\bullet} = \mu_{\bullet 3\bullet} \cdots = \mu_{\bullet K\bullet}$

(all treatment means are the same and all treatment effects equal zero)

H1: treatment means are not all equal

Test: $F_{K\text{-}1,\ df(error)} = MS_{TR}/MSE$

Can test Row effects and Column effects, but these are really not of interest.

If there are differences among treatment means:

- you might wish to test which means differ using t-tests for pairs of treatments (must divide α by the no. of pairs) or a multiple comparison test (like Scheffé's test).
- Use the MSE from the ANOVA table for each of these.



Confidence intervals for treatment means (also use the MSE from the ANOVA):

$$\bar{y}_{\bullet k\bullet} \pm t_{1-\alpha/2,df\,(error)}\sqrt{\frac{MSE}{J}}$$

Data Organization for Analysis within SAS or R:

| Row | Column | Treatment | Response |
|---|---|---|---|
| 1 | 1 | 2 | 40 |
| 1 | 2 | 1 | 35 |
| 1 | 3 | 4 | 53 |
| 1 | 4 | 3 | 47 |
| 2 | 1 | 4 | 48 |
| 2 | 2 | 3 | 46 |
| 2 | 3 | 2 | 39 |
| 2 | 4 | 1 | 34 |
| 3 | 1 | 1 | 27 |
| 3 | 2 | 4 | 53 |
| 3 | 3 | 3 | 45 |
| 3 | 4 | 2 | 41 |
| 4 | 1 | 3 | 44 |
| 4 | 2 | 2 | 39 |
| 4 | 3 | 1 | 31 |
| 4 | 4 | 4 | 52 |

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]

## Split-Plot Experiments

Freese pp. 45 to 50.

Introduction

- As with factorial experiments, treatments can be combinations of more than one factor
- In a split-plot experiment, the experimental unit (called the "whole-plot" for one factor is subdivided, and the second factor is applied to the subdivided experimental unit (called the "split" plot).
- Can be a CRD or RCB
- Split-split plot experiment: one Factor is applied to the whole experimental unit, the second Factor is applied to a sub-divided experimental unit (split-plot), and for the third factor, the split-plot is divided once more. For more on this, see "Fundamental concepts in the design of experiments" by Charles R. Hicks.



**Example from Freese:** Randomized Block Design, with two factors, but using a split-plot for the second factor

Four plantation areas of each 12 acres (imperial units) each were selected (blocks; I, II, III and IV). Each was divided into two areas (whole plot of 6 acres each), and a burning treatment (A or B) was randomly assigned to the 2 areas in each block. Each experimental unit was then sub-divided into six areas (split-plot, 1 acre each), and planting date (a,b,c,d,e,f) was randomly assigned to each split-plot In each split-plot, 1 pound of seeds were sown. At the end of the first growing season, the number of seeds were counted.

(see schematic on page 45 of the Freese book).

*Main questions:*

1. Is there an interaction between Factors?
2. If there is an interaction, look at treatment means for differences.
3. If there is no interaction:
    a. Are there differences between levels for Factor A? (whole plot)
    b. Are there differences between levels for Factor B? (split plot)



*Analysis of Variance Table (for Split-Plot RCB)*

| *Source* | *df* | *SS* | *MS* |
|---|---|---|---|
| Block | $J$-1 | $SS_{BLK}$ | $MS_{BLK}$ |
| Factor A | $K$-1 | $SS_A$ | $MS_A$ |
| Exp. Err. #1 | ($J$-1)($K$-1) | $SS_{E1}$ | $MS_{E1}$ |
| Factor B | $L$-1 | $SS_B$ | $MS_B$ |
| A x B | ($K$-1)($L$-1) | $SS_{AXB}$ | $MS_{AXB}$ |
| Exp. Err. #2 | $K$($J$-1)($L$-1) | $SS_{E2}$ | $MS_{E2}$ |
| Total | $JKL$-1 | | |

What are the appropriate F-tests?

- Depends upon which are fixed and which are random-effects.
- Then, need the expected means squares in order to decide this.
- If both factor A and Factor B are fixed, then Exp. Err. #1 is used for Factor A, and Exp. Err. #2 is used for Factor B and for A X B.

**Organization of Example Data for Analysis using a Statistics Package:**

| Block | Burn_Type | Date | yjkl |
|---|---|---|---|
| I | A | a | 900 |
| I | A | b | 880 |
| I | A | c | 1530 |
| I | A | d | 1970 |
| I | A | e | 1960 |
| I | A | f | 830 |
| I | B | a | 880 |
| I | B | b | 1050 |
| I | B | c | 1140 |
| I | B | d | 1360 |
| I | B | e | 1270 |
| I | B | f | 150 |
| II | A | a | 810 |
| II | A | b | 1170 |
| II | A | c | 1160 |
| II | A | d | 1890 |
| II | A | e | 1670 |
| II | A | f | 420 |
| II | B | a | 1100 |
| II | B | b | 1240 |
| II | B | c | 1270 |
| II | B | d | 1510 |
| II | B | e | 1380 |
| II | B | f | 380 |
| III | A | a | 760 |
| III | A | b | 1060 |
| III | A | c | 1390 |
| III | A | d | 1820 |
| III | A | e | 1310 |
| III | A | f | 570 |
| III | B | a | 960 |



| | | | |
|---|---|---|---|
| III | B | b | 1110 |
| III | B | c | 1320 |
| III | B | d | 1490 |
| III | B | e | 1500 |
| III | B | f | 420 |
| IV | A | a | 1040 |
| IV | A | b | 910 |
| IV | A | c | 1540 |
| IV | A | d | 2140 |
| IV | A | e | 1480 |
| IV | A | f | 760 |
| IV | B | a | 1040 |
| IV | B | b | 1120 |
| IV | B | c | 1080 |
| IV | B | d | 1270 |
| IV | B | e | 1450 |
| IV | B | f | 270 |

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]

**CRD: Two Factor Experiment, Both Fixed Effects, with Second Factor Nested in the First Factor**

not in the Freese Handbook

Introduction and Example

- In a CRD with two factors, a crossed design shows that all levels of Factor A are crossed with all levels in Factor B. Example:
  - Response is weight gain
  - Factor A: Salmon or Trout
  - Factor B: no warming; warmed 1 degree C; warmed 2 degrees C.
  - Treatments: 6 treatments; all combinations of Factor A crossed with Factor B.
- A nested design is when Factor B has different levels, depending on which level of Factor A. Example:
  - Response: Weight gain
  - Factor A: Salmon or Trout



  - Factor B:
    - For Salmon: No warming; warmed 2 degree C
    - For Trout: No warming; warmed 1 degrees C
- Both CRD and nested designs have "No warming", but the levels of warming differ by Factor A (species) for the nested design.
- Sometimes it is difficult to decide if the experiment is crossed or nested. For example:
  - For the experiment, could evaluate this as Factor A, Salmon or Trout crossed with Factor B, Not warmed or warmed, where the level of warming differs slightly by species.

Example:

| | | | |
|---|---|---|---|
| A1B1 = 10 | A1B1 = 11 | A1B2= 13 | A2B4 = 23 |
| A1B2 = 15 | A2B3 = 18 | A2B4= 25 | A1B1 = 11 |
| A2B4 = 20 | A2B3 = 18 | A1B1= 9 | A2B3 = 18 |
| A2B4 = 22 | A1B2 = 15 | A2B3 = 18 | A1B2 = 14 |

Nested design with two factors, where the second factor is

nested in the first factor, with four replications per treatment.



Data:

| A | B | result |
|---|---|---|
| 1 | 1 | 10.00 |
| 1 | 1 | 11.00 |
| 1 | 1 | 9.00 |
| 1 | 1 | 11.00 |
| 1 | 2 | 15.00 |
| 1 | 2 | 15.00 |
| 1 | 2 | 13.00 |
| 1 | 2 | 14.00 |
| 2 | 3 | 18.00 |
| 2 | 3 | 19.00 |
| 2 | 3 | 17.00 |
| 2 | 3 | 18.00 |
| 2 | 4 | 20.00 |
| 2 | 4 | 22.00 |
| 2 | 4 | 25.00 |
| 2 | 4 | 23.00 |

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]

## CRD: One Factor Experiment, Fixed Effects with subsampling (i.e., hierarchical)

Example: Site Preparation

A forester would like to test whether different site preparation methods result in difference in heights. Fifteen areas each 0.02 ha in size are laid our over a fairly homogeneous area. Five site preparation treatments are randomly applied to 15 plots. One hundred trees are planted (same genetic stock and same age) in each area. At the end of 5 years, the heights of EACH seedling in each plot were measured.

We have three hierarchical levels:

- Treatments
- Experimental units within treatments – level at which the treatment is applied
- Trees within experimental units – are "nested" in experimental units; different trees in different experimental units

We have variation:

- Between treatments
- Between experimental units within each treatment
- Between trees within each experimental unit in each treatment



Notation

Population: $y_{ijl} = \mu + \tau_{TRj} + \varepsilon_{EUij} + \varepsilon_{SUijl}$

$y_{ijl}$ = response variable measured on sample $l$ of experimental unit $i$ and treatment $j$

$j$=1 to $J$ treatments

$\mu$ = the grand or overall mean regardless of treatment

$\tau_{TRj}$ = the treatment effect

$\mu_j$ = the mean for treatment $j$; grand mean plus the treatment effect

The difference between a particular measure for a sample $l$, an experimental unit $i$, and the mean for the treatment $j$ that was applied to it is now two parts:

$$\varepsilon_{EUij} + \varepsilon_{SUijl} = y_{ijl} - \mu_j$$

The error for the experimental unit and the error for the sample unit in the experimental unit.

Analysis Methods

Possible ways to analyze this experiment are:

1. Simplify this by calculating averages for each experimental unit and use these in the analysis of variance (would then be Completely Randomized Design: one factor, already covered)

2. Keep each sample observation, and use least squares to calculate as per CRD: one factor, but also estimate the within experimental unit variance (will cover this now)

3. Keep each sample observation, and use a mixed model and maximum likelihood, with the two "error terms" as random-effects (e.g., PROC MIXED in SAS).

Option 1 is simpler; Options 2 and 3 allow us to look at the variability within experimental unit.

**Another option you will see but NOT CORRECT!!**

- Keep each sample observation and treat this as one experimental unit as if this was a CRD: one factor experiment.

Since the treatment was NOT applied at this level, this **analysis would not be correct**. Treatments are randomly assigned to the experimental unit level. **The degrees of freedom and the estimated error variance used in the F-test would not be correct. In some literature, the samples are termed "pseudo-replications".**



Example:

- Have three temperatures: low, medium, and high ($J$=3)
- For each, we have two experimental units (batches) (n=2)
- Randomly assign temperatures to each batch
- For each batch, we have three loaves of bread (m=2)
- The response variable is crustiness of bread.

Data:

| temp | batch | observation | yijl |
|---|---|---|---|
| low | 1 | 1 | 4 |
| low | 1 | 2 | 7 |
| low | 1 | 3 | 5 |
| low | 2 | 1 | 12 |
| low | 2 | 2 | 8 |
| low | 2 | 3 | 10 |
| medium | 1 | 1 | 14 |
| medium | 1 | 2 | 13 |
| medium | 1 | 3 | 11 |
| medium | 2 | 1 | 9 |
| medium | 2 | 2 | 10 |
| medium | 2 | 3 | 12 |
| high | 1 | 1 | 14 |
| high | 1 | 2 | 17 |
| high | 1 | 3 | 15 |
| high | 2 | 1 | 16 |
| high | 2 | 2 | 19 |
| high | 2 | 3 | 18 |

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]

## RCB: One Factor Experiment, Fixed Effects with subsampling

- Blocked (random or fixed-effect, usually random)
- Fixed-effect factor A (we will label this as TR for treatment)
- Experimental units – level at which the block with factor A combinations are applied; may be one experimental unit or more than one (generalized RCB or RCB with replicates)
- Sampling units – number of items measured within each experimental unit.

Notation for a Generalized RCB with subsampling:

Population:

$$y_{ijl} = \mu + \tau_{BLKj} + \tau_{TRk} + \tau_{BLK\times TRjk} + \varepsilon_{EUijk} + \varepsilon_{SUijkl}$$

$y_{ijkl}$ = response variable measured on sample $l$ of experimental unit $i$, block $j$, and treatment $k$

The difference between a particular measure for a sample $l$, an experimental unit $i$, and the mean for the block $j$ and treatment $k$ combination that was applied to it is now two parts:

$$\varepsilon_{EUijk} + \varepsilon_{SUijkl}$$

The error for the experimental unit and the error for the sample unit in the experimental unit.



Analysis Methods

Possible ways to analyze this experiment are:

1. Simplify this by calculating averages for each experimental unit and use these in the analysis of variance (would then be Generalized Randomized Complete Block Design: one factor, already covered)

2. Keep each sample observation, and use least squares or to calculate as per Generalized Random Complete Block: one factor, but also estimate the within experimental unit variance (will cover this now)

3. Keep each sample observation, and use a mixed model and maximum likelihood, with the two "error terms" as random-effects (e.g., PROC MIXED in SAS).

Option 1 is simpler; Options 2 and 3 allow us to look at the variability within experimental unit.

**Another option you will see but NOT CORRECT!!**

- Keep each sample observation and treat this as one experimental unit
- Since the treatment was NOT applied at this level, this **analysis would not be correct**. Treatments are randomly assigned to the experimental unit level. **The degrees of freedom and the estimated error variance used in the F-test would not be correct. In some literature, the samples are termed "pseudo-replications".**

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]

**Analysis of Covariance (ANCOVA)**

For experimental designs covered so far:

- The response variable (y) is a continuous variable
- A number of class variables (x's) are used (effects) to explain the variation in the response variable, via a linear model
- We are interested in differences in means for each class variable (fixed-effects) or in the variance in the response variable that is due to the class variable (random-effects).

For linear regression analysis, covered in the beginning of the course:

- The dependent variable (y) is a continuous variable
- A number of continuous predictor variables (x's) are used to explain the variation in the dependent variable in a linear equation.
- We also introduced class variables (x's also) to help explain the variation in the dependent variable, represented by:
    - Dummy variables to alter the intercept
    - Interactions between dummy variables and continuous predictor variable to alter the slope.



Analysis of covariance is an experimental design, where we add continuous explanatory variables (called covariates) to help explain the variability in the response variable, for example:

- Record the initial weight of all fish prior to adding different foods. Use this initial weight as a covariate
- Record soil moisture of all plots in a field prior to applying different treatments. Use this soil moisture as a covariate.
- The covariates help "even-out" conditions that we were not able to control in trying to obtain homogeneous treatment units, and explain some of the variation in the response variable. We use these covariates to "adjust" the factor level means to a common value (usually the mean) of the covariate.

Blocking does this in a similar fashion, but:

- Blocking restricts the randomization of treatments to experimental units (treatments assigned randomly within blocks)
- Blocks are class variables.

For analysis of covariance:

- the slopes are considered the same over all treatments (common slope), in order to assess the impacts of different factors (called homogeneity of slopes)
- This must be tested, as the slope of y versus x may vary by treatment

Model:

We add a covariate to whichever experimental design we wish to use. For example, using an RCB with two fixed-effect factors, we add in the covariate.

Population:

$$y_{jkl} = \mu + \beta(x_{jkl} - \bar{x}) + \tau_{BLK\,j} + \tau_{Ak} + \tau_{Bl} + \tau_{ABkl} + \varepsilon_{jkl}$$

$y_{jkl}$ = response variable measured on block $j$ and treatment $kl$

$j$=1 to $J$ blocks; $k$=1 to $K$ levels for Factor A; $l$=1 to $L$ levels for Factor B; and definition of terms follows other designs.

$x_{jkl}$ is a measurement of the covariate; $\beta$ is the slope of the line between y and x.



Example:

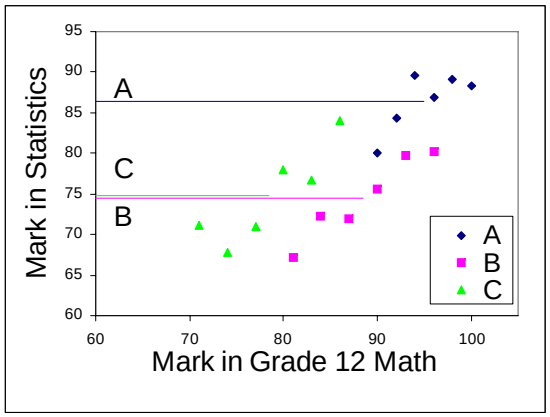
A university would like to evaluate three ways of teaching basic statistics:
(A) stats dept. method (3 lectures),
(B) computer method (3 lectures plus lab using statistical software with no lab write-up),
(C) applied science method (3 lectures plus written lab).
"Success" is measured as a grade in a common examination for all students.

The response (exam grade) might be related to abilities before taking the course:

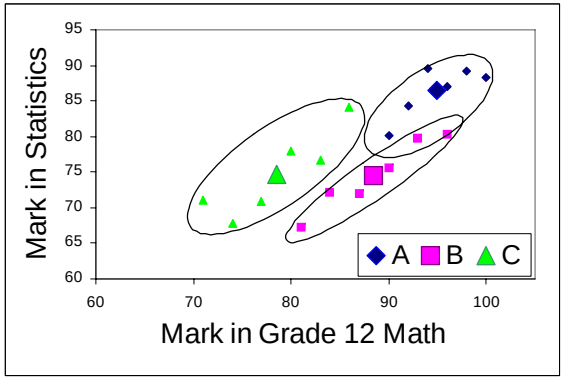
- Grade in Math 12 is used as a covariate (x variable) and obtained for each student.
- Then students are randomly assigned to one of the three class types.

The Math 12 grade is then used to "adjust" the grade in the common exam.

Looking at the trends, between Mark in Stats (y) versus Mark in Grade 12 math(x), the slopes appear to be similar.

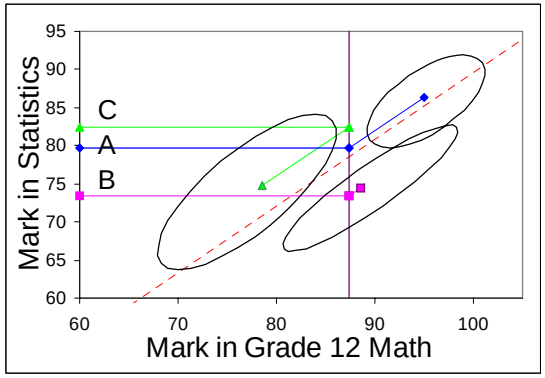



Ignoring the Grade 12 March, the mark in Statistics is higher for A, and B and C are similar.





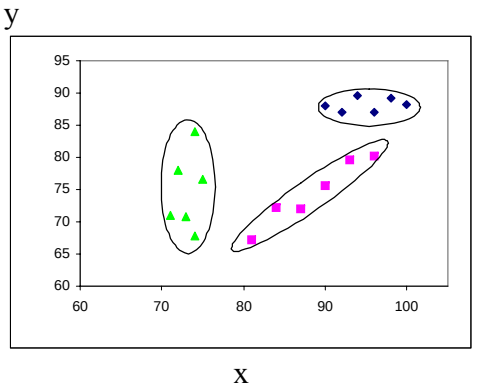

Using the covariate, and adjusting the means along the y vs x trend line to the average Mark in Grade 12 Math, C and A are similar, and B is different




If the Math grade was not used as a covariate, the conclusion would be much different.

Variations in ANCOVA:

1. Slopes are not equal




- Harder to interpret, as with any interaction
    - Use graphs to show relationships
    - Switch to a regression approach to finding equations using the continuous and class variables (represented as dummies) and interpret these results.

2. More than one covariate. Can add in more than one continuous variable.

[see www.forestry.ubc.ca/biometrics and course materials for FRST 430/533, for more detaila on this design]