Analysis of variance of a non-standard experimental design

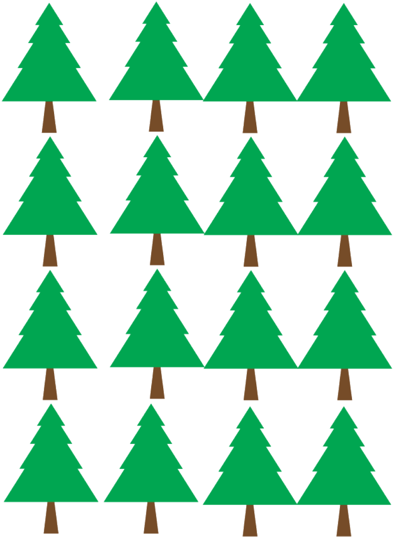
Ciro Velasco

November 1, 2024

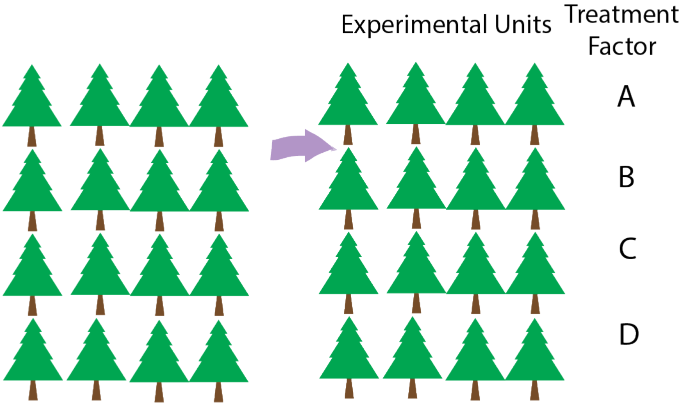
The following shows the structure of the dataset 1 of a CRD.

Rep	ET	Species	Variety	Quality
1	60	Fa	20-411	4.67
1	60	Fa	20-FA405	5.33
1	60	Fa	20-FA407	4.44
1	60	Fa	20-FA408	5.56
1	60	Fa	BARNOBLE	6.22
1	60	Fa	BARROBUSTO	6.89
1	60	Fa	BONFIRE	5.89

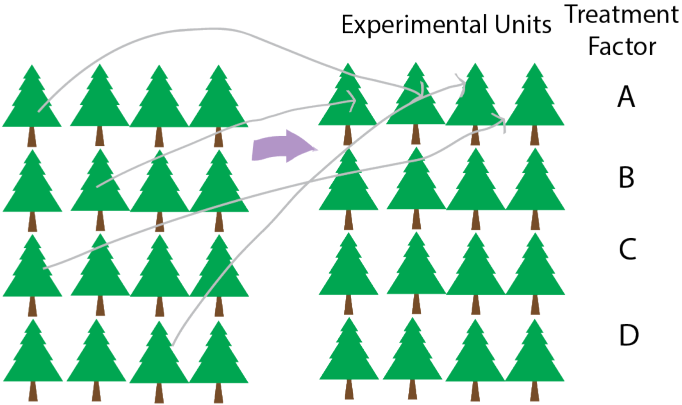
There are 16 seedlings



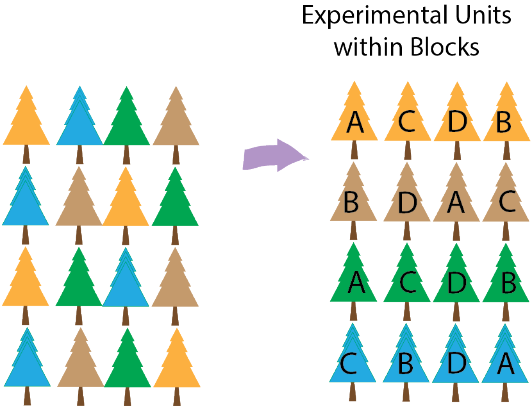
There are 16 seedlings, and one treatment factor with 4 levels



There are 16 seedlings, and one treatment factor with 4 levels, randomly assigned



Blocking factor

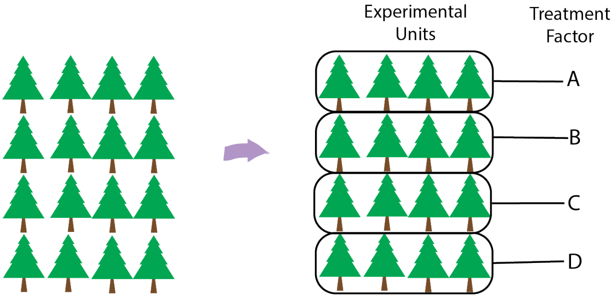


The three principles of experimental designs

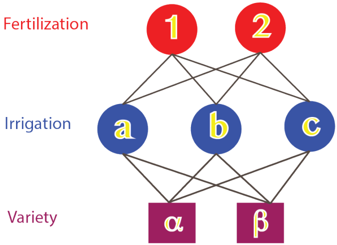
We just introduced the three most important principles of experimental designs

- Replication. To evaluate experimental variability
- Randomization. To avoid bias
- Blocking. To control for unwanted variability of the experimental units

Still another concept: Subsamples



Completely Crossed



Fertilization

1

2

Irrigation

a

b

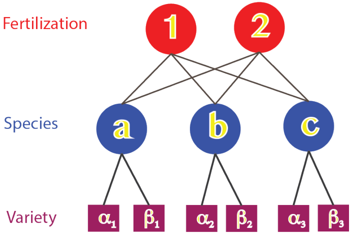
c

Variety

α

β

Crossed-Nested



Fertilization

1

2

Species

a

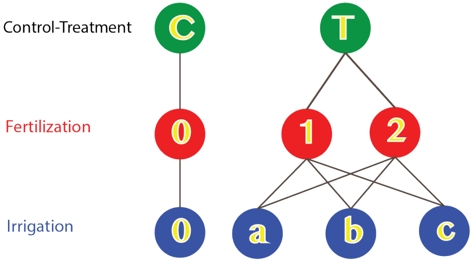
b

c

Variety

α_1 β_1 α_2 β_2 α_3 β_3

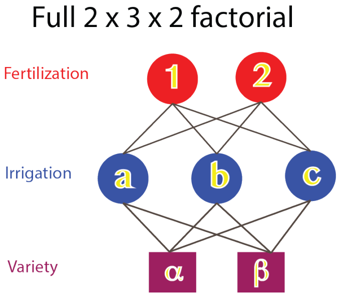
Extended Design



The general model for a full factorial with three treatment factors, which is denoted for short by $A \times B \times C$ is given by

$$Y_{ijkl} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\gamma_k+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+(\alpha\beta\gamma)_{ijk}+\epsilon_{ijkl} \tag{1}$$

That for our toy example above, A :*Fertilization* equal to 2 levels, C : *Irrigation* equal to three levels, and C : *Variety* equal to two levels.



Notes

The model above is valid under the assumption that the experiment is a Completely Randomized Design (CRD). Though, this full factorial could have been allocated into any other design, such as:

- Randomized Complete Block Design or RCBD,
- Split-Split Plot Design or SSPD,
- Incomplete Block Design or IBD,
- Etc.

The model would be different. Models are generally design specific. All of them are very standard, since the listed designs are all standard.

Crossed Nested model

For a non-standard design such as the Crossed Nested, though very common, the model is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \cancel{\gamma_{kj}} + \cancel{(\alpha\gamma)_{ikj}} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl} \tag{2}$$

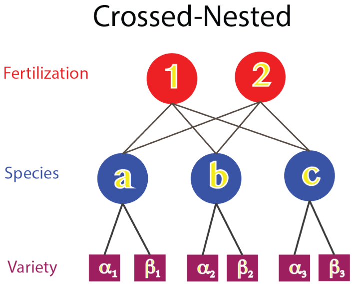
The nesting makes it imposible to estimate the canceled terms. The first becomes embedded into $(\beta\gamma)_{jk}$ and the second into $(\alpha\beta\gamma)_{ijk}$.

Note: γ is accompanied by the nesting factor all the times in nested experiments, ie. as in the model above, γ is always with β.

Thus the model is simplified as

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}. \tag{3}$$

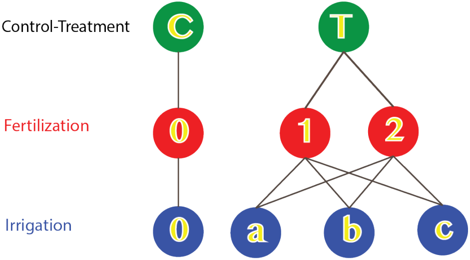
Which is the model for our toy example



α is for *Fertilization*, β is for *Species*, and γ for *Variety*.
Note: Species is the NESTING treatment factor and Variety is the NESTED treatment factor.

In the particular case of the Extended, the Control-Treatment design, the model

$$Y_{ijkl} = \mu + \alpha_i + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}. \qquad (4)$$



Notice that in this case, the nesting factor is α, thus the other terms in the model are always accompanied by this term everywhere. There is no way to separate the main effects of β, γ nor the interaction effect $(\beta\gamma)$.

The extended designs can be

- Extended RCBD
- Extended SPD
- Extended IBD
- Etc.

For instance, the model for the extended RCBD is

$$Y_{hijkl} = \mu + \theta_h + \alpha_i + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk} + \epsilon_{hijkl}, \qquad (5)$$

where θ symbolizes the blocking factor effect, which can be random or fixed.

ANOVA table for the Nested-Crossed

Let us assume that the sources of variation are the terms shown in the formula, and that for the nested factor we have total C levels, $C = C_1 + C_2 + \ldots + C_B$, where C_j is the number of levels of the nested factor for level j of factor B.

Source of Variation	Degrees of Freedom
α	$A - 1$
β	$B - 1$
$(\alpha\beta)$	$(A - 1) \times (B - 1)$
$(\beta\gamma)$	$(C_1 - 1) + (C_2 - 1) + \ldots + (C_B - 1) = (C - B)$
$(\alpha\beta\gamma)$	$(A - 1) \times (C - B)$
Error	$A \times C \times (R - 1) - 1$
Total	$A \times C \times R - 1$

Where R is the number of replications.

Dataset 1

The following data were collected from a CRD with the following treatment factors (every treatment combination when existed was replicated four times)

1. ET: Evapotranspiration with three levels (60, 80 and 100%)
2. Species: with two levels (Fa and Pp)
3. Variety: with 17 varieties for Fa, and 19 for Pp.

This dataset has the structure similar to our first toy Crossed-Nested design above.

The following shows the structure of the dataset 1.

Rep	ET	Species	Variety	Quality
1	60	Fa	20-411	4.67
1	60	Fa	20-FA405	5.33
1	60	Fa	20-FA407	4.44
1	60	Fa	20-FA408	5.56
1	60	Fa	BARNOBLE	6.22
1	60	Fa	BARROBUSTO	6.89
1	60	Fa	BONFIRE	5.89

Again, the model for this dataset is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}, \qquad (6)$$

where μ is the overall mean or in other contexts, the intercept, α represents the ET effect, β represents the Species effect, and γ the variety effect. The subscripts are as follows: $i = 1, 2, 3$ to rollover the three levels of ET. $j = 1, 2$ to run over the two levels for Species, and $k = 1, \ldots, 17$ or $k = 1, \ldots, 19$ for the varieties. And $l = 1, \ldots, 4$, the subscript for replications.
Often this model is expressed as

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + (\beta\gamma)_{k(j)} + (\alpha\beta\gamma)_{ik(j)} + \epsilon_{ijkl}, \qquad (7)$$

to more clearly specify that k runs on different indices depending on the value of j.

The ANOVA table of dataset 1 follows, note that $C = 17 + 19 = 36$, with $R = 4$.

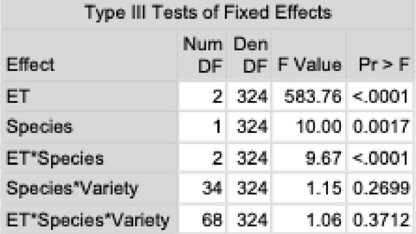
Source of Variation	Degrees of Freedom
ET	$3 - 1 = 2$
Species	$2 - 1 = 1$
ET × Species	$(3 - 1) \times (2 - 1) = 2$
Species × Variety	$(17 - 1) + (19 - 1) = 34$
ET × Species × Variety	$(3 - 1) \times (17 + 19 - 2) = 68$
Error	$3 \times 36 \times (4 - 1) - 1 = 324$
Total	$3 \times 36 \times 4 - 1 = 431$

SAS code for the Crossed-Nested data

In SAS, there are two ways to specify the same analysis.

1. Proc GLIMMIX data=mqua plots=all;
   ```
   Proc GLIMMIX data=mqua plots=all;
   class rep et species variety;
   model quality = et|species|variety(species);
   store one;
   run;
   ```
2. Proc GLIMMIX data=mqua plots=all;
   ```
   Proc GLIMMIX data=mqua plots=all;
   class rep et species variety;
   model quality = et species et×species species×variety
   et×species×variety ;
   store one;
   run;
   ```

The following shows the SAS output with results



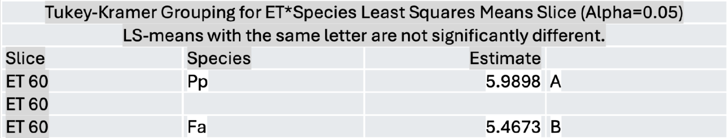
Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
ET	2	324	583.76	<.0001
Species	1	324	10.00	0.0017
ET*Species	2	324	9.67	<.0001
Species*Variety	34	324	1.15	0.2699
ET*Species*Variety	68	324	1.06	0.3712

What follows is to investigate the two factor interaction *Species* × *Variety*, using the following Postprocessing procedure in SAS:

```
Proc PLM restore=one;
slice et*species/sliceby = et linestable adjust=tukey;
effectplot interaction(sliceby=species);
run;
```

PLM results

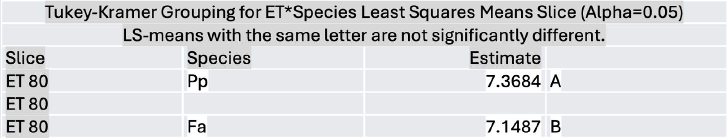
The following summarizes what is obtained from Proc PLM, in particular from invoking the slice analysis.



Tukey-Kramer Grouping for ET*Species Least Squares Means Slice (Alpha=0.05)

LS-means with the same letter are not significantly different.

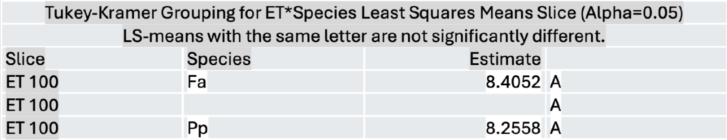
Slice	Species	Estimate	
ET 60	Pp	5.9898	A
ET 60			
ET 60	Fa	5.4673	B



Tukey-Kramer Grouping for ET*Species Least Squares Means Slice (Alpha=0.05)

LS-means with the same letter are not significantly different.

Slice	Species	Estimate	
ET 80	Pp	7.3684	A
ET 80			
ET 80	Fa	7.1487	B



Tukey-Kramer Grouping for ET*Species Least Squares Means Slice (Alpha=0.05)

LS-means with the same letter are not significantly different.

Slice	Species	Estimate	
ET 100	Fa	8.4052	A
ET 100			A
ET 100	Pp	8.2558	A

Conclusions

- Standard designs can be extended to accommodate more complicated experimental layouts
- The common design that has control as an additional factor in the experiments can be analyzed using the crossed-nested model
- SAS is an useful tool to analyze this and other more complicated designs.
- The crossed-nested designs can be extended to accommodate repeated measurements, though more interesting model can be used for this type of data.

References

- Hinkelmann, Klauss and Oscar Kempthorne (2005). Design and Analysis of Experiments: Advanced Experimental Design. Wiley Series in Probability and Statistics. Vol 1.
- Piepho, H.P., Williams, E.R., and Fleck, M. (2006). A Note on the Analysis of Designed Experiments with Complex Treatment Structure. HortScience HortSci, 41(2), 446-452. Retrieved May 20, 2024, from https://doi.org/10.21273/HORTSCI.41.2.446