ANOVA Designs - Part II

Nested Designs (NEST)
- Design
- Linear Model
- Computation
- Example
- NCSS

Factorial Designs (FACT)
- Design
- Linear Model
- Computation
- Example
- NCSS

RCB Factorial (Combinatorial Designs)

Nested Designs

A nested design (sometimes referred to as a *hierarchical* design) is used for experiments in which there is an interest in a set of treatments and the experimental units are *sub-sampled.*

For example, consider a typical provenance study where a forest geneticist collects 5 seeds from 5 superior trees in each of 3 forests. The seeds are germinated in a greenhouse and the seedlings are measured for height growth. Graphically, the design would look like this...

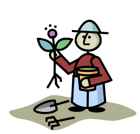
Nested Designs

Forest A B C

Trees 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Seedlings i ii iii iv v... etc.

Total of 75 seedlings.



Nested Designs

Note that in this sort of design, each parent tree and each seed is given a unique identity because it is not replicated across a treatment--it is unique to that particular treatment because of it's genotype.

This type of design is very common in genetics, systematics, and evolutionary studies where it is important to keep track of each plant obtained from specific populations, lines, or parentage.

Nested Design

-Linear Model-

The additive model for this design is:

$$y_{ijk} = \mu + \alpha_i + \beta_{(i)j} + \varepsilon_{ijk}$$

where:

μ: constant; overall mean

α_i : constant for ith treatment group; deviation from mean of i

β_{ij} : a random effect due to the ith group nested witin the jth experimental unit

ε_{ijk} : random deviation associated with each observation

NB: same basic form as RCB, but *j* subscript has been added to β and ε.

Nested Design

-Computation-

Let's look at a similar but simpler example of a provenance study:

Tree		Forest					
		A	B	C	D	E	
1		15.8	18.5	12.3	19.5	16.0	
		15.6	18.0	13.0	17.5	15.7	
		16.0	18.4	12.7	19.1	16.1	
	$T_{i1.}$	47.4	54.9	38.0	56.1	47.8	
2		13.9	17.9	14.0	18.7	15.8	
		14.2	18.1	13.1	19.0	15.6	
		13.5	17.4	13.5	18.8	16.3	
	$T_{i2.}$	41.6	53.4	40.6	56.5	47.7	
	$T_{i..}$	89.0	108.3	78.6	112.6	95.5	484.0

Nested Design

-Computation-

Sum of Squares	df	SS	MS	F
Among Forests	a-1	SS_a	MS_a	MS_a/MS_b
Btwn trees within forests	a(b-1)	SS_b	MS_b	MS_b/MS_e
Amng seedlg within trees	ab(n-1)	SS_e	MS_e	
Total	abn-1			

Nested Design

-Computation-

Source	df	SS	MS	F
Among Forests	4	129.28	32.32	22.6***
Trees (Forest)	5	7.14	1.43	7.15***
Among Seedlings	20	4.01	0.20	
Total	29			

```
> seedling<-read.csv("C:/TEMPR/Seedling.csv")
> attach(seedling)
> seedling
   Forest Tree Height
1       A   T1   15.8
2       A   T1   15.6
3       A   T1   16.0
4       A   T2   13.9
5       A   T2   14.2
6       A   T2   13.5
7       B   T3   18.5
8       B   T3   18.0
9       B   T3   18.4
10      B   T4   17.9
11      B   T4   18.1
12      B   T4   17.4
13      C   T5   12.3
14      C   T5   13.0
15      C   T5   12.7
16      C   T6   14.0
17      C   T6   13.1
18      C   T6   13.5
19      D   T7   19.5
20      D   T7   17.5
21      D   T7   19.1
22      D   T8   18.7
23      D   T8   19.0
24      D   T8   18.8
25      E   T9   16.0
26      E   T9   15.7
27      E   T9   16.1
28      E   T0   15.8
29      E   T0   15.6
30      E   T0   16.3
```

Nested Design

- Using R -

Design: 5 forests, 2 trees per forest (10 total), 3 seedlings grown from each tree. Seedlings are nested within tree are nested within forest.

NB: Difference in coding for nested design! Each tree *must* be coded differently as one tree can not occur in five different forests.

Nested Design

- Using -

Note use of virgule to designate nested effect.

```
> anova(lm(Height~Forest/Tree))

Analysis of Variance Table

Response: Height
            Df  Sum Sq Mean Sq F value    Pr(>F)
Forest       4 129.277  32.319 161.059 6.553e-15 ***
Forest:Tree  5   7.137   1.427   7.113 0.0005718 ***
Residuals   20   4.013   0.201
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

Factorial Design

Often an investigator is interested in the *combined (interactive) effect* of two types of treatments. For example, in a greenhouse study you might be interested in the effects water, fertilizer, and the combined effect of water & fertilizer on seedling biomass.

This design differs from a blocking design because neither nutrients nor water are considered extraneous sources of variability--they are both central to the hypothesis. This is an economical design because it accomplishes several things at once.

Factorial Design

A typical design such as we have just discussed might look like this graphically:

Nutrients

		Low	Med	High
Water	Low			
	Med			
	High			

Factorial Design

The sets of treatments are called *factors* or *main effects*. The different treatment within sets are called *levels*. Levels can, and usually are, categorical in nature.

In our example, *nutrients* would be Factor-A and contain 3 levels and *water* would be Factor-B and contain 3 levels. Thus, There would be a total of a×b treatment combinations (i.e., $3 \times 3 = 9$). If there were $n = 5$ seedlings per treatment, there would be $N = 45$ seedlings in the study.

This particular design permits the analysis of interactions (i.e., evaluates whether B responds the same way across all levels of A).

Factorial Design

-Model-

The additive model for this design is:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

where:

μ : constant; overall mean

α_i : constant for ith treatment group; deviation from mean of i

β_j : constant for the jth source of variation; deviation from the mean of j

$\alpha\beta_{ij}$: the interaction effect bewteen i & j for A & B

(NB: this is single term & not a product.)

ε_{ijk} : random deviation associated with each observation

Factorial Design

-Computations-

Source	df	SS	MS
Factor A	a-1	SS_a=A-CF	MS_a= SS_a/(a-1)
Factor B	b-1	SS_b=B-CF	MS_b= SS_b/(b-1)
A×B Interaction	(a-1)(b-1)	SS_{ab}=S-A-B+CF	MS_{ab}= SS_{ab}/(a-1)(b-1)
Error	ab(n-1)	SS_e=T-S	MS_e = SS_e/ab(n-1)
Total	abn-1	SS_t=T-CF	

Factorial Design
-Computations-

$$T = \sum_i \sum_j \sum_k y_{ijk}^2$$

$$A = \sum_i T_{i..}^2 / bn$$

$$B = \sum_j T_{.j.}^2 / an$$

$$S = \sum_i \sum_j T_{ij.}^2 / n$$

$$CF = T_{...}^2 / abn$$

Computations are performed in virtually the same way as we have done for previous designs, only now we add S to account for the interaction term.

Factorial Design
-Computations-

FEM, REM, Mixed	MS	F-test
A fixed, B fixed	A	MS_a / MS_e
	B	MS_b / MS_e
	A×B	MS_{ab} / MS_e
A rand, B rand	A	MS_a / MS_{ab}
	B	MS_b / MS_{ab}
	A×B	MS_{ab} / MS_e
A fixed, B rand	A	MS_a / MS_{ab}
	B	MS_b / MS_e
	A×B	MS_{ab} / MS_e
A rand, B fixed	A	MS_a / MS_e
	B	MS_b / MS_{ab}
	A×B	MS_{ab} / MS_e

The appropriate F-test is determined by the *type of factor* (fixed vs. random).

At this point, the specification of the type of factor you have determines the outcomes of the analysis!

Factorial Design
-Example-

Suppose we wished to look at seedling vigor of Ohio buckeyes and assess the variation attributable to tree (1, 2, 3, 4) and fertilizer type (A, B, C). Two nuts are sampled at random and seedlings are grown from the nuts. Vigor is scored as 1-10.

In this type of design, fertilizer type is a fixed effect and tree is random effect (we could use any 4 buckeye trees), so the MS has to be adjusted accordingly.

The data are as follows...

Factorial Design

-Example-

Fertilizer	Tree-1	Tree-2	Tree-3	Tree-4
A	2	4	3	1
	1	2	1	1
B	4	3	6	6
	5	3	7	5
C	6	8	7	5
	4	8	8	6

Factorial Design

-Example-

The resulting ANOVA table for these data would be:

Source	df	SS	MS	F_{calc}	F_{table}
Fertilizer	2	88.08	44.04	12.62	5.143
Tree	3	9.83	3.28	4.37	3.490
Fert × Tree	6	20.92	3.49	4.65	2.996
Error	12	9.00	0.75		

Factorial Design

-Example Using -

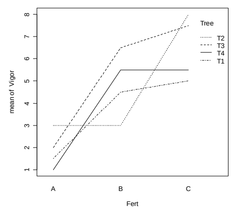
```
> vigor<-read.csv("C:/TEMPR/Vig.csv")
> vigor
   Fert Tree Vigor
1     A   T1     2
2     A   T1     1
3     A   T2     4
4     A   T2     2
5     A   T3     3
6     A   T3     1
7     A   T4     1
8     A   T4     1
9     B   T1     4
10    B   T1     5
11    B   T2     3
12    B   T2     3
13    B   T3     6
14    B   T3     7
15    B   T4     6
16    B   T4     5
17    C   T1     6
18    C   T1     4
19    C   T2     8
20    C   T2     8
21    C   T3     7
22    C   T3     8
23    C   T4     5
24    C   T4     6
> attach(vigor)
```

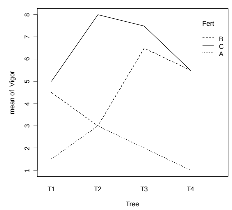
```
> summary(vigor)
 Fert  Tree        Vigor
 A:8   T1:6   Min.   :1.000
 B:8   T2:6   1st Qu.:2.750
 C:8   T3:6   Median :4.500
       T4:6   Mean   :4.417
              3rd Qu.:6.000
              Max.   :8.000
```

Factorial Design

-Example Using -

```
> interaction.plot(Fert,Tree,Vigor)
> interaction.plot(Tree,Fert,Vigor)
```





Factorial Design

-Example Using -

Note symbol for interaction design.

How does this compare to our hand calculations?

```
> anova(lm(Vigor~Fert*Tree))
Analysis of Variance Table

Response: Vigor
          Df Sum Sq Mean Sq F value    Pr(>F)
Fert       2 88.083  44.042 58.7222 6.347e-07 ***
Tree       3  9.833   3.278  4.3704   0.02680 *
Fert:Tree  6 20.917   3.486  4.6481   0.01146 *
Residuals 12  9.000   0.750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1
```

Factorial Design

-Computations-

FEM, REM, Mixed	MS	F-test
A fixed, B fixed	A	MS_a / MS_e
	B	MS_b / MS_e
	A×B	MS_{ab} / MS_e
A rand, B rand	A	MS_a / MS_{ab}
	B	MS_b / MS_{ab}
	A×B	MS_{ab} / MS_e
A fixed, B rand	A	MS_a / MS_{ab}
	B	MS_b / MS_e
	A×B	MS_{ab} / MS_e
A rand, B fixed	A	MS_a / MS_e
	B	MS_b / MS_{ab}
	A×B	MS_{ab} / MS_e

The appropriate F-test is determined by the *type of factor* (fixed vs. random).

At this point, the specification of the type of factor you have determines the outcomes of the analysis!

Factorial Design

-Example Using R-

```
> summary(aov(Vigor~Fert*Tree+Error(Fert*Tree)))

Error: Fert
     Df Sum Sq Mean Sq
Fert  2 88.083  44.042

Error: Tree
     Df Sum Sq Mean Sq
Tree  3 9.8333  3.2778

Error: Fert:Tree
          Df  Sum Sq Mean Sq
Fert:Tree  6 20.9167  3.4861

Error: Within
          Df Sum Sq Mean Sq
Residuals 12   9.00    0.75
```

A	MS_a / MS_{ab}
B	MS_b / MS_e
A×B	MS_{ab} / MS_e

One approach is to call for the basics of the AOV table and then do the F-tests manually to construct the appropriate table.

This is a bit less than satisfying. Alternatively, turn to `lme` (linear mixed effects models, in `nlme` package).

```
> vigor.lme.2<-lme(Vigor~Fert*Tree, random = ~1 | Tree, data=vigor)
> summary(vigor.lme.2)

Linear mixed-effects model fit by REML
 Data: vigor
      AIC      BIC    logLik
  66.9201 73.7088 -19.46005

Random effects:
 Formula: ~1 | Tree
        (Intercept)  Residual
StdDev:    0.942809 0.8660254

Fixed effects: Vigor ~ Fert * Tree
             Value Std.Error DF   t-value p-value
(Intercept)    1.5 1.1242281 12  1.334249  0.2069
FertB          3.0 0.8660254 12  3.464102  0.0047
FertC          3.5 0.8660254 12  4.041452  0.0016
TreeT2         1.5 1.5898987  0  0.943456     NaN
TreeT3         0.5 1.5898987  0  0.314485     NaN
TreeT4        -0.5 1.5898987  0 -0.314485     NaN
FertB:TreeT2  -3.0 1.2247449 12 -2.449490  0.0306
FertC:TreeT2   1.5 1.2247449 12  1.224745  0.2442
FertB:TreeT3   1.5 1.2247449 12  1.224745  0.2442
FertC:TreeT3   2.0 1.2247449 12  1.632993  0.1284
FertB:TreeT4   1.5 1.2247449 12  1.224745  0.2442
FertC:TreeT4   1.0 1.2247449 12  0.816497  0.4301
```

RCB Factorial Experiments

It should now be clear that by simple extension, one can make more complex experimental designs by simply combining terms in the linear model.

For example, in a typical drug interaction experiment, a study would be designed with a control (no drugs), Drug A (0/1), Drug B (0/1), and a Drug A×B interaction. These drugs are given to various subjects, one per day over 4 days. Subject is used as a block to remove this as a source of variability.

RCB Factorial Experiments

The data for this experiment response times (in msec) and are as follows (Rao 1998, Ex. 15.5, p. 715):

	Subjects			
Therapy	1	2	...	8
No drugs	18.8	18.5	...	26.5
Drug A alone	13.5	9.8	...	15.5
Drug B alone	13.6	13.4	...	15.4
Drugs A & B combo.	10.6	12.6	...	12.6

RCB Factorial Experiments

Thus, the linear model for this design would be:

$y_{ijk} = \mu + \mathrm{R}_i + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk}$

where:

μ: constant; overall mean

R_i : constant for the ith block

α_j : constant for ith treatment group; deviation from mean of i

β_k : constant for the jth source of variation; deviation from the mean of j

$\alpha\beta_{jk}$: the interaction effect bewteen j & k for A & B

ε_{ijk} : random deviation associated with each observation