Topic: Designed experiments with blocking factors

Block Designs

Motivation - sometimes the variability of responses among experimental units is large, making detection of differences among treatment means $\mu_1, \mu_2, \ldots, \mu_t$ difficult

In a randomized complete block design (RBD) to evaluate t treatments in b complete blocks, ,

1. matched sets of experimental units are formed, each consisting of t units. Goal is reduced variance of the response within a block .That is, the units within a block are homogeneous. Variance between blocks is fine., not a problem.
2. Units are randomly assigned to each of the t treatments within each block, as opposed to the *completely randomized* design where units are assigned completely at random to treatments.

in a restricted randomization)

```
%let seed0=234;    *Randomized Complete Block Design;
%let seed=368;

data trts;
   do driver=1 to 6;
   do i=1 to 3;
      u=ranuni(&seed);
      output;
   end;
   end;
   keep driver u;
run;
proc rank data=trts out=rtrts;
   by driver;
   var u;
   ranks ru;
run;
data rtrts;
   array cname{3} $ ("D","E","F");
   set rtrts;
   by driver;
   cartype=cname{ru};
   keep cartype driver;
run;
proc transpose data=rtrts out=rcbd prefix=day;
   by driver;
   var cartype;
run;
proc print data=rcbd;run;
```

Obs	driver	_NAME_	day1	day2	day3
1	1	cartype	D	F	E
2	2	cartype	E	D	F
3	3	cartype	F	E	D
4	4	cartype	D	E	F
5	5	cartype	D	F	E
6	6	cartype	E	D	F

RBD - first example

Acrophopia can be treated in several ways:

- "Contact desensitization " - activity/task demonstrated then walked through while a therapist is in constant contact with the subject.
- "Demonstration participation" - therapist talks subject thru task, no contact.
- "Live Modeling" - subject simply watches completion of task

Severity of acrophobia measured by HAT (Height Avoidance Test) scores, measured before/after therapy. Considerable heterogeneity in degree of acrophobia. So $N = 15$ subjects put into **blocks** according to original HAT score, then one from each block randomly assigned to a therapy. Δ HAT score below:

	Therapy			
Block j	Contact Desensitization	Demonstration Participation	Live Modeling	$\bar{y}_{+j}$
1	8	2	-2	2.67
2	$y_{12} = 11$	1	0	4
3	9	12	6	9
4	16	11	2	9.67
5	24	19	11	18
Avg $\bar{y}_{i+}$	13.6 A	9 A	3.4 B	

RBD example

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \text{iid } N(0, \sigma^2)$$

Source	Sum of Squares	d.f.	Mean Square	F
A: Therapies	260.9	2	130.5	15.3
B: Blocks	438	4	109.5	
Error	68.4	8	$\hat{\sigma}^2 = 8.6$	
Total	767.3	14		

(Data taken from Larsen and Marx, 1986)

$$SS[Tot] = SS[A] + SS[B] + SS[E]$$

$$767.3 = SS[Tot] = \sum_{i=1}^{3}\sum_{j=1}^{5}(Y_{ij} - \bar{Y}_{..})^2$$

$$260.9 \quad SS[A] = \sum\sum(\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$438 \quad SS[B] = \sum\sum(\bar{Y}_{.j} - \bar{Y}_{..})^2$$

$$68.4 \quad SS[E] = \sum\sum(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$$

Note that

$$y_{ij} - \bar{y}_{++} = \underbrace{(\bar{y}_{i+} - \bar{y}_{++})}_{\text{therapy effect}} + \underbrace{\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}}_{\text{block effect}} + \underbrace{y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}}_{\text{error}}$$

F-tests in the RBD

A model for RBD with fixed treatment (therapy) effects is

$$Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$$

where $i = 1, \ldots, a \quad j = 1, \ldots, b$ and $E_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

Mean squares obtained by dividing SS by df:

$$\begin{aligned} MS[A] &= \frac{SS[A]}{a-1} \\ MS[B] &= \frac{SS[B]}{b-1} \\ MS[E] &= \frac{SS[E]}{N-a-b+1} \end{aligned}$$

The primary hypothesis of interest is for a therapy effect:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \quad \text{vs} \quad H_1 : \text{not all equal.}$$

Using level α, reject H_0 if

$$F = \frac{MS[A]}{MS[E]} > F(\alpha, a-1, N-a-b+1)$$

$$Y_{ij} = \mu + \alpha_i + \beta_j + \Sigma_{ij}$$

The *EMS* for error is $E(MS[E]) = \sigma^2$, but only under the *additivity* assumption that there is no block-trt interaction. This assumption is required for inference about treatment effects in the absence of replication, common to block designs.

For the HAT scores, $F_A = MS[A]/MS[E] = 130.5/8.6 = 15.3$ which has $p < 0.01$ on $2, 8$ df, providing strong evidence of a therapy effect. Inference, including MCPs, for CONTRASTS involving fixed effects is the same in the complete RBD as it is for other factorial experiments with fixed effects. E.g.

$$\widehat{SE}(\overline{Y}_{i+}) = \sqrt{MS[E]/b}, \quad \widehat{SE}(\overline{Y}_{i+} - \overline{Y}_{j+}) = \sqrt{2MS[E]/b}$$

will change if Block effects modelled as random variables

Another example

Expt conducted to see whether or not artificial food supplements might entice rats to eat rat poison. 3200 baits placed around garbage-storage areas. Baits were mixed with $t = 4$ flavors Baits randomized to four close locations with equal access. After 2 wks, the percentage of baits devoured was recorded. Then other sets of locations in the same area were selected and the experiment was repeated for four more two-weeks periods:

Expt	Percentage of baits accepted				
	Plain	Vanilla	RoastB	Bread	Avg.
1	13.8	11.7	14.0	12.6	13.0
2	12.9	16.7	15.5	13.8	14.7
3	25.9	29.8	27.8	25.0	27.1
4	18.0	23.1	23.0	16.9	20.3
5	15.2	20.2	19.0	13.7	17.0
Avg	17.2	20.3	19.9	16.4	18.4
Std Dev.	5.3	6.8	5.6	5.1	

Source	Sum of Squares
Flavor	56.38
Expt.	495.32
Error	29.76
Total	581.46

- Calculate an F-ratio that can be used to test for a flavor effect.
- Conduct all pairwise comparisons using Tukey's HSD with FWE .05.
- Inhomogeneity of variance? Ranks of means and variances?

Multiple comparisons among means in the RBD

Scheffè simultaneous 95% confidence intervals for contrasts like

$$c_1\mu_1 + c_2\mu_2 + \cdots + c_a\mu_a$$

look like

$$c_1\bar{y}_{1+} + c_2\bar{y}_{2+} + \cdots + c_a\bar{y}_{a+} \pm \sqrt{(a-1)(F^*)MS[E]\sum \frac{c_i^2}{b}}$$

where $F^* = F(0.05, a-1, N-a-b+1)$. For simultaneous pairwise differences,

$$\bar{y}_{i+} - \bar{y}_{j+} \pm \underbrace{\sqrt{(a-1)(F^*)MS[E]\frac{2}{b}}}$$

"minimum significant difference"

For the HAT scores, $\bar{y}_{1+} = 13.6$, $\bar{y}_{2+} = 9$, $\bar{y}_{3+} = 3.4$ and

$$\sqrt{(a-1)(F^*)(MS[E])(1/5+1/5)} = \sqrt{(3-1)(4.46)(8.6)(2/5)} = 5.5$$

with $\bar{y}_{LM+}$ significantly different from the other two. (LM brings about significantly less improvement than the other two therapies.)

Tukey's Studentized Range (HSD) Test for variable: DIFF

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 8 MSE= 8.55
Critical Value of Studentized Range= 4.041
Minimum Significant Difference= 5.2843

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	TREAT
A	13.600	5	Contact Desensit
A			
A	9.000	5	Demonstration Pa
B	3.400	5	Live Modelling

means treat/

scheffe tukey

$$HSD = \frac{q}{\sqrt{2}}\sqrt{\frac{2MS(E)}{b}} = 5.28$$

Scheffe's test for variable: DIFF

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha= 0.05 df= 8 MSE= 8.55
Critical Value of F= 4.45897
Minimum Significant Difference= 5.5226

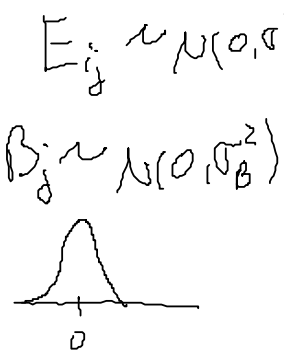
Scheffe Grouping	Mean	N	TREAT
A	13.600	5	Contact Desensit
A			
A	9.000	5	Demonstration Pa
B	3.400	5	Live Modelling

Another example, block effects are random

A study investigates the efficiency of four different unit-dose injection systems. For each system, an individual subject (pharmacist or nurse) measures the average time it takes to remove a unit of each system from its outer package, assemble it, and simulate an injection. (Data from Larsen and Marx, 1986.)

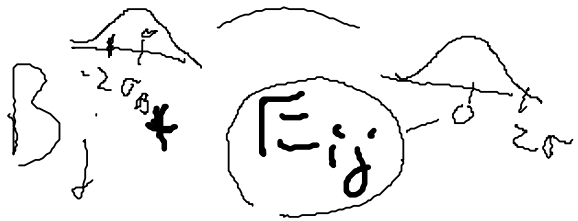
Average times (seconds) for implementing systems

Subject	Standard	Vari-Ject	Unimatic	Tubex	$\bar{y}_{+j}$
1	35.6	17.3	24.4	25.0	25.6
2	31.3	16.4	22.4	26.0	24.0
3	36.2	18.1	22.8	25.3	25.6
4	31.1	17.8	21	24	23.5
5	39.4	18.8	23.3	24.2	26.4
6	34.7	17	21.8	26.2	24.9
7	34.1	14.5	23	24	23.9
8	36.5	17.9	24.1	20.9	24.9
9	40.7	16.4	31.3	36.9	31.3
$\bar{y}_{i+}$	35.5	17.1	23.8	25.8	25.6



Model

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



```
data one;
   input subject system time;
   cards;
1 1 35.6
2 1 31.3
  ...
8 4 20.9
9 4 36.9
;
run;

proc mixed method=type3;
   class system subject;
   model time=system/ddfm=satterth;
   random subject;
   lsmeans system/adj=tukey cl pdiff;
run;
```

```
                          The Mixed Procedure
              Dependent Variable            time
              Covariance Structure          Variance Components
              Estimation Method             Type 3
              Fixed Effects SE Method       Model-Based
              Degrees of Freedom Method     Satterthwaite

                Class       Levels     Values
                system           4     1 2 3 4
                subject          9     1 2 3 4 5 6 7 8 9

                        Total Observations               36

                            Type 3 Analysis of Variance
                           Sum of
Source          DF        Squares     Mean Square   Expected Mean Square
system           3    1559.202222      519.734074   Var(Residual) + Q(system)
subject          8     177.405000       22.175625   Var(Residual) + 4 Var(subject)
Residual        24     148.472778        6.186366   Var(Residual)
```

Source	Error Term	Error DF	F Value	Pr > F
system	MS(Residual)	24	84.01	<.0001
subject	MS(Residual)	24	3.58	0.0072
Residual	.	.	.	.

Covariance Parameter Estimates

Cov Parm	Estimate
subject	3.9973
Residual	6.1864

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
system	3	24	84.01	<.0001

Least Squares Means

Effect	system	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
system	1	35.5111	1.0637	21.9	33.38	<.0001	0.05
system	2	17.1333	1.0637	21.9	16.11	<.0001	0.05
system	3	23.7889	1.0637	21.9	22.36	<.0001	0.05
system	4	25.8333	1.0637	21.9	24.29	<.0001	0.05

Effect	system	Lower	Upper
system	1	33.3044	37.7178
system	2	14.9266	19.3400
system	3	21.5822	25.9956
system	4	23.6266	28.0400

Injection systems significantly different. Estimated variance component of block effects:

$$\hat{\sigma}_B^2 = \frac{1}{a}(MS[B] - MS[E]) = \frac{1}{4}(22.2 - 6.2) = 4(\text{squared seconds})$$

Differences of Least Squares Means

Effect	system	_system	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment
system	1	2	18.3778	1.1725	24	15.67	<.0001	Tukey-Kramer
system	1	3	11.7222	1.1725	24	10.00	<.0001	Tukey-Kramer
system	1	4	9.6778	1.1725	24	8.25	<.0001	Tukey-Kramer
system	2	3	-6.6556	1.1725	24	-5.68	<.0001	Tukey-Kramer
system	2	4	-8.7000	1.1725	24	-7.42	<.0001	Tukey-Kramer
system	3	4	-2.0444	1.1725	24	-1.74	0.0940	Tukey-Kramer

Effect	system	_system	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
system	1	2	<.0001	0.05	15.9579	20.7977	15.1433	21.6122
system	1	3	<.0001	0.05	9.3023	14.1421	8.4878	14.9567
system	1	4	<.0001	0.05	7.2579	12.0977	6.4433	12.9122
system	2	3	<.0001	0.05	-9.0755	-4.2356	-9.8900	-3.4211
system	2	4	<.0001	0.05	-11.1199	-6.2801	-11.9345	-5.4655
system	3	4	0.3242	0.05	-4.4644	0.3755	-5.2789	1.1900

Note the *df* columns:

- For difference of means, pesky mean random effects wash out
- For means, random effects don't wash out:

$$\begin{aligned} \overline{Y}_{i_1+} &= \\ \overline{Y}_{i_2+} &= \\ \overline{Y}_{i_1+} - \overline{Y}_{i_2+} &= \\ SE(\overline{Y}_{i_1+}) &= \\ SE(\overline{Y}_{i_1+} - \overline{Y}_{i_2+}) &= \end{aligned}$$

Latin squares: for experiments with

Fisher's famous experiment with $3 \times 2 = 6$ combos of

- 3 levels of phosphate
- 2 levels of nitrogen

	Column					
Row	1	2	3	4	5	6
1	E	B	F	A	C	D
2	B	C	D	E	F	A
3	A	E	C	B	D	F
4	F	D	E	C	A	B
5	D	A	B	F	E	C
6	C	F	A	D	B	E

Labels:

	Phosphate		
	none	single	double
without Nitrate	A	B	C
with Nitrate	D	E	F

Data (data taken from p. 91 of Fisher's "The Design of Experiments.")

	Column						
Row	1	2	3	4	5	6	
1	633 E	527 B	652 F	390 A	504 C	416 D	520.3
2	489 B	475 C	415 D	488 E	571 F	282 A	453.3
3	384 A	481 E	483 C	422 B	334 D	646 F	458.3
4	620 F	448 D	505 E	439 C	323 A	384 B	453.2
5	452 D	432 A	411 B	617 F	594 E	466 C	495.3
6	500 C	505 F	259 A	366 D	326 B	420 E	396.0
means	513	478	454.2	453.7	442.0	435.7	462.8

Means:

Level of phosphate	Level of nitrogen	N	y Mean	y Std Dev
0	0	6	345.000000	67.7701999
0	1	6	405.166667	46.5635766
1	0	6	426.500000	72.3512267
1	1	6	520.166667	78.7589148
2	0	6	477.833333	23.9116429
2	1	6	601.833333	55.4163033

Model: $Y_{ijk} =$

```
proc glm data=both;
   title "Factorial effects of phosphate and nitrogen";
   class row col phosphate nitrogen;
   model y=row col nitrogen|phosphate;
   estimate "nitrogen effect without phosphate"     nitrogen -1 1
                                 nitrogen*phosphate -1 1;
   estimate "nitrogen effect with single phosphate" nitrogen -1 1
                                 nitrogen*phosphate 0 0  -1 1;
   estimate "nitrogen effect with double phosphate" nitrogen -1 1
                                 nitrogen*phosphate 0 0  0 0  -1 1;
   estimate "phosphate nonlinear"                   phosphate 1 -2 1;
   contrast "phosphate nonlinear"                   phosphate 1 -2 1;
   lsmeans nitrogen*phosphate/slice=phosphate;
run;
```

The GLM Procedure

Class	Levels	Values
row	6	1 2 3 4 5 6
col	6	1 2 3 4 5 6
phosphate	3	0 1 2
nitrogen	2	0 1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	326845.7500	21789.7167	14.27	<.0001
Error	20	30541.0000	1527.0500		
Corrected Total	35	357386.7500			

R-Square	Coeff Var	Root MSE	y Mean
0.914544	8.444622	39.07749	462.7500

Source	DF	Type I SS	Mean Square	F Value	Pr > F
row	5	54198.5833	10839.7167	7.10	0.0006
col	5	24467.2500	4893.4500	3.20	0.0276
nitrogen	1	77191.3611	77191.3611	50.55	<.0001
phosphate	2	164871.5000	82435.7500	53.98	<.0001
phosphate*nitrogen	2	6117.0556	3058.5278	2.00	0.1611

Exercise

Let Y_{ij} denote the observation in row i column j. Let $\overline{Y}_k$ denote the treatment mean for k^{th} treatment level. For latin square, identify these sums of squares:

$$\sum_{i=1}^{a}\sum_{j=1}^{a}(\overline{y}_{i+} - \bar{y}_{++})^2 = SS[\quad]$$

$$\sum_{i=1}^{a}\sum_{j=1}^{a}(y_{ij} - \bar{y}_{++})^2 = SS[\quad]$$

$$a\sum_{j=1}^{a}(\overline{y}_{+j} - \bar{y}_{++})^2 = SS[\quad]$$

$$\sum_{i=1}^{a}\sum_{j=1}^{a}(y_{ij} - \overline{y}_{i+} - \overline{y}_{+j} - \overline{y}_k + 2\bar{y}_{++})^2 = SS[\quad]$$

$$a\sum_{k=1}^{a}(\overline{y}_k - \bar{y}_{++})^2 = SS[\quad]$$

Note that $\overline{y}_k$ determined by the i, j combination. For $i = j = 1$ in Fisher's potatoes design, $\overline{y}_5 = 520.2$.

Test the hypothesis that neither nitrogen nor phosphate have any effect on yield. (Do the averages on slide 17 differ significantly?)

$$\begin{aligned} F &= \frac{MS(trt)}{MS(E)} \\ &= \frac{[SS(N) + SS(P) + SS(N * P)]/5}{MS(E)} \\ &= \frac{[77191 + 164871 + 6117]/5}{?} \\ &= \frac{248180/5}{?} = 32.5(df = 5, 20) \end{aligned}$$

Next we might test for $N \times P$ interaction:

$$F = \frac{MS(N \times P)}{MS(E)} = \frac{6117/}{1527} =$$

Appropriate $\alpha = .01, .05, .10$ critical values, respectively: $5.84, 3.49, 2.59$. Conclusion?

If one does not want to assume phosphate and nitrogen effects are additive, one could assess simple effects:

Least Squares Means

phosphate	nitrogen	y LSMEAN
0	0	345.000000
0	1	405.166667
1	0	426.500000
1	1	520.166667
2	0	477.833333
2	1	601.833333

phosphate*nitrogen Effect Sliced by nitrogen for y

nitrogen	DF	Sum of Squares	Mean Square	F Value	Pr > F
0	2	53844	26922	17.63	<.0001
1	2	117144	58572	38.36	<.0001

phosphate*nitrogen Effect Sliced by phosphate for y

phosphate	DF	Sum of Squares	Mean Square	F Value	Pr > F
0	1	10860	10860	7.11	0.0148
1	1	26320	26320	17.24	0.0005
2	1	46128	46128	30.21	<.0001

- Observed phosphate effects significant for each level of nitrogen
- Observed nitrogen effects significant for each level of phosphate

Main effects:

```
Level of                    ---------------y---------------
phosphate          N                  Mean             Std Dev

0                 12          375.083333          63.7216366
1                 12          473.333333          87.1303447
2                 12          539.833333          76.4803401


Level of                    ---------------y---------------
nitrogen           N                  Mean             Std Dev

0                 18          416.444444          78.904426
1                 18          509.055556         101.272766
```

Is phosphate effect *linear*? (Note $SS(\text{phosphate}) = SS(\text{lin}) + SS(\text{nonlinear})$.)

```
   contrast "phosphate linear" phosphate -1 0 1;
   contrast "phosphate nonlinear" phosphate 1 -2 1;
```

```
Contrast                 DF    Contrast SS    Mean Square   F Value   Pr > F

phosphate linear          1    162855.3750    162855.3750    106.65   <.0001
phosphate nonlinear       1      2016.1250      2016.1250      1.32   0.2641
```

How to randomize a latin square:

- randomly permute rows
- randomly permute columns
- randomly permute labels

Suppose we have $a = 4$ treatments, labelled 1,2,3,4.

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

Aside: The latin square we started with is in *reduced form*, where the first row is 1,2,3,4 and so is the first column. Any 4×4 latin square in reduced form will take either the form above or the form below:

1	2	3	4
2	1	4	3
3	4	2	1
4	3	1	2

(Two "isotopy classes")

- Advantages:
  - better control of error (smaller $MS(E)$)
- Disadvantages
  - loss of df for $MS(E)$ (few would advocate the use of a solitary 3×3 Latin Square.
  - limited existence of designs can be restrictive

Graeco-Latin Squares

(Hyper) Graeco-Latin Squares permit investigation of a treatment allowing for three (or more) blocking factors.

Multiple superimposed (mutually) orthogonal latin squares:

α	β	γ	δ
β	γ	δ	α
γ	δ	α	β
δ	α	β	δ

A	B	C	D
D	C	B	A
B	A	D	C
C	D	A	B

Orthogonal - every greek-latin combo appears exactly once.

Latin rectangle

Case 2 Example, taken from Oehlert's book:
3 methods of drug-delivery: *A*-solution, *B*-tablet, *C*-capsule.

	Period					
Subject	1		2		3	
1	A	1799	C	1846	B	2147
2	C	2075	B	1156	A	1777
3	B	1396	A	868	C	2291
4	B	3100	A	3065	C	4077
5	C	1451	B	1217	A	1288
6	A	3174	C	1714	B	2919
7	C	1430	A	836	B	1063
8	A	1186	B	642	C	1183
9	B	1135	C	1305	A	984
10	C	873	A	1426	B	1540
11	A	2061	B	2433	C	1337
12	B	1053	C	1534	A	1583