Displays for Statistics 5303

Lecture 36

December 2, 2002

Christopher Bingham, Instructor

612-625-7023 (St. Paul)
612-625-1024 (Minneapolis)

Class Web Page
http://www.stat.umn.edu/~kb/classes/5303

Efficiency of a Incomplete Box Design

The efficiency is usually stated relative to a RCB design with the same number r of replicates *and the same* σ.

$E_{BIBD:RCB} = g(k-1)/((g-1)k) = 1-(g-k)/((g-1)k)$

Since number of treatments = $g > k$ = block size, $E_{BIBD:RCB} < 1$.

Of course, one reason for using an incomplete block design is that σ^2 tends to be smaller for small blocks than for large. Since $E_{BIBD:RCB}$ assumes the same σ^2 for a size k block as for a size g block, it may not provide a meaningful comparison of designs if the variances are very different.

Still $E_{BIBD:RCB}$ is a useful number since it appears in a number of formulas.

$E_{BIBD:RCB}$ mainly enters through the **effective number of replicates**

$$r' = E_{BIBD:RCB} \times r < r.$$

Here are several ways r' is used.

Estimates of α_i

Let $v_{ij} = y_{ij} - \overline{y}_{\bullet j}$ = response with block mean removed. Then

$\hat{\alpha}_i = v_{i\bullet}/r'$ ((treatment sum)/r')

Treatment SS

$SS_{trt} = r'\sum_i \hat{\alpha}_i^2$

Variance and standard error of a contrast

$V(\sum_i c_i \hat{\alpha}_i) = (\sum_i c_i^2)\sigma^2/r'$

$SE(\sum_i c_i \hat{\alpha}_i) = \sqrt{\{(\sum_i c_i^2)\sigma^2/r'\}}$

These are like the complete block formulas with the effective replication r' in place of the actual replication r.

Analysis of BIBD Example 14.2 using MacAnova.

The analysis is virtually identical with a RCB design, although the "by hand" formulas are more complicated.

```
Cmd> tab14_1 <- read("","exmpl14.2")
exmpl14.2         36     3
) A data set from Oehlert (2000) \emph{A First Course in Design
) and Analysis of Experiments}, New York: W. H. Freeman.
)
) Data originally from John, P. W.~M. (1961). ``An application x
) and  balanced incomplete block design
) '' {\em Technometrics\/}~{\em 3}, 51--54.
)
) Table 14.1, p. 359
) Test of 9 different detergents.  There are three basins that
) are used simultaneously at the same rate with a different
) detergent in each basin.  Response is number of plates until
) foam disappears in a basin.
) Column 1 is session. Column 2 is treatment (kind of detergent)
) Column 3 is response (number of dishes)
) Treatments 1-4 are detergent base 1
) with (3, 2, 1, or 0) parts additive
) Treatments 5-8 are detergent base
) 2 with (3, 2, 1, or 0) parts additive
) Treatment 9 is a control.
Read from file "TP1:Stat5303:Data:OeCh14.dat"

Cmd> makecols(tab14_1,session,treatment,count)

Cmd> session <- factor(session)

Cmd> treatment <- factor(treatment)
```

The blocking factor is `session`. It must appear in the model before `treatment`.

```
Cmd> anova("count=session + treatment",fstat:T)
Model used is count=session+treatment
WARNING: summaries are sequential
               DF          SS          MS            F       P-value
CONSTANT        1       13572       13572  16469.69663    1.5466e-25
session        11      412.75      37.523     45.53320    6.0284e-10
treatment       8      1086.8      135.85    164.85393    6.8089e-14
ERROR1         16      13.185     0.82407
```

You can do pairwise multiple comparisons as for a CRD and CRB.

```
Cmd> pairwise("treatment",.05,hsd:T)
              4    -12.9
              3    -6.22
  |           2    -2.22
  | |         8   -0.222
  | |         1    0.333
    | |       7     1.67
      | |     6     3.56
        |     5     5.89
              9     10.1
```

You can check that standard errors for different pairwise comparisons are the same:

```
Cmd> contrast(treatment,vector(1,-1,rep(0,7)))
component: estimate
(1)      2.5556
component: ss
(1)      9.7963
component: se
(1)      0.7412

Cmd> contrast(treatment,vector(1,0,-1,rep(0,6)))
component: estimate
(1)      6.5556
component: ss
(1)      64.463
component: se
(1)      0.7412
```



```
Cmd> trtmeans <- tabs(count,treatment,mean:T)

Cmd> blkmeans <- tabs(count,session,mean:T)

Cmd> tabs(count,treatment,count:T) # reps of each treatment
(1)           4           4           4           4           4
(6)           4           4           4           4

Cmd> tabs(count,session,count:T) # sizes of each block
 (1)           3           3           3           3           3
 (6)           3           3           3           3           3
(11)           3           3

Cmd> r <- 4; k <- 3; b <- 12; g <- 9;

Cmd> eff <- g*(k-1)/((g-1)*k); eff
(1)       0.75

Cmd> effective_r <- r*eff; effective_r
(1)           3
```

Thus $E_{BIBD:RCB} = .75$ and $r' = 3$.

Remove block means from the data.

```
Cmd> v <- count

Cmd> for(i,1,b){ # loop over blocks
        v[session == i] <- v[session == i] - blkmeans[i];;}

Cmd> tabs(v,treatment,sum:T)/effective_r # alphahats
(1)     0.33333     -2.2222     -6.2222     -12.889      5.8889
(6)      3.5556      1.6667    -0.22222      10.111

Cmd> coefs(treatment) # same by black box
(1)     0.33333     -2.2222     -6.2222     -12.889      5.8889
(6)      3.5556      1.6667    -0.22222      10.111

Cmd> c <- vector(1, -1, rep(0,g-2)); c # pairwise contrast
(1)           1          -1           0           0           0
(6)           0           0           0           0

Cmd> mse <- SS[4]/DF[4] # estimated sigma^2

Cmd> sqrt(mse*sum(c^2)/effective_r) # standard error
(1)      0.7412

Cmd> contrast(treatment,c)$se # Standard error from black box
(1)      0.7412
```

Recovery of interblock information

The incomplete randomized block model is

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

where not all possible (i,j) occur.

Consider the block sums $y_{\bullet j} \equiv \sum_i y_{ij}$ where the sum is over all treatments in block j. Then

$$\begin{aligned} y_{\bullet j} &= \textstyle\sum_i y_{ij} = k\mu + \sum_i n_{ij}\alpha_i + k\beta_j + \sum_i n_{ij}\varepsilon_{ij} \\ &= k\mu + \textstyle\sum_i n_{ij}\alpha_i + \eta_j \end{aligned}$$

where

- $n_{ij} = 1$ if treatment i is in block j
- $n_{ij} = 0$ if treatment i is not in block j
- $\eta_j \equiv k\beta_j + \sum_i n_{ij}\varepsilon_{ij}$

Note that $\sum_i n_{ij} = k$ so that $k\mu = \sum_i n_{ij}\mu$ and

$$y_{\bullet j} = \textstyle\sum_i n_{ij}(\mu + \alpha_i) + \eta_j$$

This has the form of a multiple regression *with no constant term*

$$y_{\bullet j} = \textstyle\sum_i \tilde{\beta}_i x_{ij} + \eta_j$$

where the regression coefficients and predictor variables are

$$\tilde{\beta}_i = \mu + \alpha_i \text{ and } x_{ij} = n_{ij}$$

Provided the block effects β_j's are random and independent, η_j are independent with constant variance

$$\tilde{\sigma}^2 = k^2\sigma_\beta^{\ 2} + k\sigma^2.$$

When this is the case we can get estimates of $\mu + \alpha_i$ by least squares regression with $y_{\bullet j}$ as response variable and $x_{ij} = n_{ij}$ as predictor variables.

```
Cmd> blksums <- tabs(count,session,sum:T) # y_dotj

Cmd> nij <- tabs(count,session, treatment,count:T)

Cmd> print(nij,format:"2.0f")
nij:
 (1,1)  1  1  1  0  0  0  0  0  0  Trt 1, 2, 3 in block 1
 (2,1)  0  0  0  1  1  1  0  0  0  Trt 4, 5, 6 in block 2
 (3,1)  0  0  0  0  0  0  1  1  1  Trt 7, 8, 9 in block 3
 (4,1)  1  0  0  1  0  0  1  0  0  Trt 1, 4, 7 in block 4
 (5,1)  0  1  0  0  1  0  0  1  0   etc
 (6,1)  0  0  1  0  0  1  0  0  1
 (7,1)  1  0  0  0  1  0  0  0  1
 (8,1)  0  1  0  0  0  1  1  0  0
 (9,1)  0  0  1  1  0  0  0  1  0
(10,1)  1  0  0  0  0  1  0  1  0
(11,1)  0  1  0  1  0  0  0  0  1
(12,1)  0  0  1  0  1  0  1  0  0

Cmd> makecols(nij,x1,x2,x3,x4,x5,x6,x7,x8,x9)
```

Do regression without constant term by adding **-1** to the model:

```
Cmd> regress("blksums=x1+x2+x3+x4+x5+x6+x7+x8+x9-1")
Model used is blksums=x1+x2+x3+x4+x5+x6+x7+x8+x9-1
                Coef      StdErr          t
x1             19.75       1.147     17.219
x2            15.417       1.147     13.441
x3            13.417       1.147     11.697
x4            6.4167       1.147     5.5944
x5            26.083       1.147     22.741
x6            24.083       1.147     20.997
x7             19.75       1.147     17.219
x8            19.417       1.147     16.928
x9            30.417       1.147     26.519

N: 12, MSE: 4.3056, DF: 3, R^2: 0.99969
Regression F(9,3): 1082.4, Durbin-Watson: 0.63656
To see the ANOVA table type 'anova()'
```


To get estimates of α_i which sum to 0, you need to subtract the mean of the g coefficients from each coefficient.

```
Cmd> COEF - sum(COEF)/g # interblock estimates of alpha's
         x1           x2           x3           x4           x5
         x6           x7           x8           x9
    0.33333           -4           -6          -13       6.6667
     4.6667      0.33333 -3.5527e-15           11
```

Macro **interblock()** provides a "black box" way to get these estimates

```
Cmd> interblock(count,session,treatment)# response,block,treat
 intra est intra se   inter est inter secombined est combined se
   0.33333  0.49414     0.33333  0.91174      0.33333     0.43443
   -2.2222  0.49414          -4  0.91174      -2.6259     0.43443
   -6.2222  0.49414          -6  0.91174      -6.1718     0.43443
   -12.889  0.49414         -13  0.91174      -12.914     0.43443
    5.8889  0.49414      6.6667  0.91174       6.0655     0.43443
    3.5556  0.49414      4.6667  0.91174       3.8078     0.43443
    1.6667  0.49414     0.33333  0.91174       1.3639     0.43443
  -0.22222  0.49414 -1.7764e-15  0.91174     -0.17177     0.43443
    10.111  0.49414          11  0.91174       10.313     0.43443
```

The **intra est** column has estimates as computed by **coefs()** following **anova()**. They are *intra-block* estimates because they are implicitly computed from within block differences.

The **inter est** column has the same values as just found, the *interblock* estimates.

```
Cmd> anova("count=session + treatment",silent:T)

Cmd> coefs(treatment)
(1)     0.33333     -2.2222     -6.2222     -12.889      5.8889
(6)      3.5556      1.6667    -0.22222      10.111
```

These match column 1.

The **combined est** column is a linear combination which weights the estimates inversely proportional to their variance.

Caution: This interblock recovery works *only* when blocks are random.

- The standard errors of the inter-block estimates of treatment effects are always bigger than the intra-block estimates.
- Unless the number of blocks is large, there is unlikely to be much benefit from this procedure.

Balanced incomplete blocks designs have the property

- All treatment effects α_i can be estimated with equal accuracy
- All contrasts with the same $\sum c_i^2$ have the same standard error, *always larger* than the standard error of the contrast from a RCB with r blocks and the same σ^2.

Essentially, the loss of efficiency is spread among all contrasts equally.

This isn't always desirable.

- Some contrasts are more important than others.
- Some contrasts may be assumed to be 0

Consider a 2^3 factorial with means

$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_k$
$+ \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$, $1 \le i,j,k \le 2$

You can express μ and the main and interaction effects in terms of the μ_{ij}'s.

$\mu = (\mu_{111}+\mu_{211}+\mu_{121}+\mu_{221}+\mu_{112}+\mu_{212}+\mu_{122}+\mu_{222})/8$

$\alpha_2 = -\alpha_1 =$
$(-\mu_{111}+\mu_{211}-\mu_{121}+\mu_{221}-\mu_{112}+\mu_{212}-\mu_{122}+\mu_{222})/8$

$\beta_2 = -\beta_1 =$
$(-\mu_{111}-\mu_{211}+\mu_{121}+\mu_{221}-\mu_{112}-\mu_{212}+\mu_{122}+\mu_{222})/8$

$\alpha\beta_{22} = \alpha\beta_{11} = -\alpha\beta_{12} = -\alpha\beta_{21} =$
$(-\mu_{111}+\mu_{211}+\mu_{121}-\mu_{221}-\mu_{112}+\mu_{212}+\mu_{122}-\mu_{222})/8$

...

$\alpha\beta\gamma_{222} = -\alpha\beta\gamma_{111} = \alpha\beta\gamma_{211} = \ldots = -\alpha\beta\gamma_{221} =$
$(-\mu_{111}+\mu_{211}+\mu_{121}-\mu_{221}+\mu_{112}-\mu_{212}-\mu_{122}+\mu_{222})/8$

These are all (except for μ) contrasts in the $g = 2^3 = 8$ treatment means μ_{ijk}.

Except for the divisor 8, each effect can be found using a column of this table

	I	A	B	C	AB	AC	BC	ABC
(1)	1	-1	-1	-1	1	1	1	-1
a	1	1	-1	-1	-1	-1	1	1
b	1	-1	1	-1	-1	1	-1	1
ab	1	1	1	-1	1	-1	-1	-1
c	1	-1	-1	1	1	-1	-1	1
ac	1	1	-1	1	-1	1	-1	-1
bc	1	-1	1	1	-1	-1	1	-1
abc	1	1	1	1	1	1	1	1

or, equivalently, of this one

	I	A	B	C	AB	AC	BC	ABC
(1)	+	−	−	−	+	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	−	+	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	+	−	+	−	−
bc	+	−	+	+	−	−	+	−
abc	+	+	+	+	+	+	+	+

On p. 612-613 are several BIBD plans for $g = 8$ treatments, with block sizes $k = 2$, 3, 4, 5, 6, and 7. Using any of you can estimate all effects equally accurately.

But often ABC is not important, or doesn't need to be estimated at all because you assume $\alpha\beta\gamma_{ijk} = 0$.

Here is an incomplete block design that is not a BIB for the $g = 8$ factorial treatments in blocks of size 4

I	II
(1)	a
ab	b
ac	c
bc	abc

Blocks I has the treatments for which the ABC contrast coefficients are -1 and block II has the treatments for which the ABC contrast coefficients are +1.

The estimate of $\alpha\beta\gamma_{222}$ is

$\widehat{\alpha\beta\gamma}_{222} =$

$$(-y_{111}+y_{211}+y_{121}-y_{221}+y_{112}-y_{212}-y_{122}+y_{222})/8$$
$$= \alpha\beta\gamma_{222} + (\beta_1 - \beta_2)/2 + \sum c_{ijk}^{ABC}\varepsilon_{ijk}$$

where $c_{ijk}^{ABC} = \pm 1/8$

Note that this has block effects "contaminating" the interaction effect. This "contamination" is known as *confounding*.

Here the interaction is confounded with the difference between blocks.

How about the other contrasts?
Because the contrasts are orthogonal

$$\hat{\alpha}_2 = \alpha_2 + \sum c_{ijk}^{A}\varepsilon_{ijk}$$
$$\hat{\beta}_2 = \beta_2 + \sum c_{ijk}^{B}\varepsilon_{ijk}$$
$$\vdots$$
$$\widehat{\beta\gamma}_{jk} = \beta\gamma_{jk} + \sum c_{ijk}^{BC}\varepsilon_{ijk}, \text{ etc.}$$

These are unconfounded with block effects.

Because of the way the blocks were chosen, ABC is called the *defining contrast* for the design.

Any other column of the table of contrasts (except the column of all +1's) could be the defining contrast for a design with blocks of size $2^{k-1} = 4$. The corresponding main effect or interaction would be confounded with blocks.

Block I consists of all the treatments with -1 and block II consists of all the treatments with +1 on the defining contrast.