

# Complete Block Designs

Hao Zhang

- Sample Size Determination
- Factorial Experiments
- Example The Banana Experiment

## Sample Size Determination

Since the complete block models are similar to the two-way models for a factorial design, the sample size determination is very similar to that for a two-way models. I use the following example to demonstrate.

**Example** (Example 10.6.3 Colorfastness experiment in the textbook)

The was planned by D-Y Duan, H. Rhee, and C. Song in 1990 to investigate the effects of the number of washes on the color change of a denim fabric. The levels of the treatment factor were the number of times of laundering, and these were selected to be 1, 2, 3, 4, and 5. Because determining the color change may be subjective and three experimenters were involved, they decided to use a block design where each experimenters was a block. Thus the levels of the blocking factor denoted the experimenter, and there were  $b = 3$  blocks. They decided to use a general complete block design and allowed the block size to be  $k = vs = 5s$ , where  $s$  could be chosen.

They planned to use a block-treatment interaction model (10.6.8), and they wanted to test the null hypothesis of no treatment differences whether or not there was block  $\times$  treatment interaction. Suppose the test was to be carried out at significance level 0.05, and suppose the experimenters wanted to reject the null hypothesis with probability 0.99 if there was a true difference of  $\Delta = 0.5$  or more in the effect of the number of washes on color rating. They expected  $\sigma$  to be no larger than about 0.4.

We need to find the minimum value of  $s$ . The following SAS code is what we used for the two-way complete model.

```
data samplesize;
input r@@;
a=5;
Block b=3;
diff=0.5;
sigma=0.4;
alpha=0.05;
df1=a-1;
df2=a*b*(r-1);
ncp=b*r*diff**2/(2*sigma**2);
Falpha=finv(1-alpha, df1, df2);
power=1-probf(Falpha, df1, df2, ncp);
lines;
8 10 11 12 13
;
proc print;
var r power;
run;
```

Obs	r	power
1	8	0.94210
2	10	0.98079
3	11	0.98930
4	12	0.99415
5	13	0.99686

Therefore the number of replicates per block needs to be 12. The advantage of such programming is it gives the power of the test for any given replicate in a block.

The following SAS code demonstrate how to calculate the width of simultaneous confidence intervals for pairwise comparisons using Tukey's method.

```
data size;
input r@@;
a=5;
b=3;
mse=0.4**2;
alpha=0.05;
prob=1-alpha;
```

```

proc print;
var r width;
run;

```

```

data size;
input r@@;
a=5;
b=3;
mse=0.4**2;
alpha=0.05;
prob=1-alpha;
df=a*b*(r-1);
qT=probmcc('range',,,prob,df,a);
width=2*(qT/2**0.5)*sqrt(mse*2/(b*r));
lines;
8 10 11 12 13
;
proc print;
var r width;
run;

```

Tukey's

## Factorial Experiments

When the treatments are factorial in nature, the treatment parameter  $\tau_i$  in the complete block design models can be replaced by main-effect and interaction parameters.

For example, for an experiment with two treatment factors A and B that is designed as a randomized complete block design, the block-treatment model is

Instead of  $\tau_i$  only:  
 $\tau_{ij}$  is the combination of 2 factors

$$Y_{hijt} = \mu + \theta_h + \underbrace{\tau_{ij}}_{\text{Total}} + \epsilon_{hijt}$$

with the usual assumptions on the error variables, where  $h$  for block,  $i$  for the level of  $A$  and  $j$  for the level of  $B$  and  $t$  for replicate.

To investigate the interaction between  $A$  and  $B$  and the main effects of  $A$  or  $B$ , we can write

Breakdown:  $\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$ .

Factor A      Factor B      Interaction

It model becomes

$$Y_{hijt} = \mu + \theta_h + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{hijt}$$

The analysis of the model is similar to what we covered for the factorial designs without blocks.

It is worth to note that the block-treatment interaction model

$$Y_{hijt} = \mu + \theta_h + \tau_{ij} + (\theta\tau)_{hij} + \epsilon_{hijt}$$

is equivalent to the following model

$$Y_{hijt} = \mu + \theta_h + \alpha_i + \beta_j + (\alpha\beta)_{ij} + (\theta\alpha)_{hi} + (\theta\beta)_{hj} + (\theta\alpha\beta)_{hij} + \epsilon_{hijt}$$

Block, A, B Interaction (3 way)

Block, A Interaction (2 way)

Block, B Interaction (2 way)

Block, A, B Interaction (3 way)

In practice, some interactions with the blocks may be negligible. This results in increased degrees of freedom.

## Example The Banana Experiment

This experiment was a class project run by three students. The objective is to investigate the effects of lighting condition (two factors, 1=day/night cycle, 2=closet) and storage method (two levels, 1= hanging, 2 = counter-top) on the ripening of bananas. The experiment was run as a complete block design with the three experimenters as blocks. The number of replicates within each block is  $s = 4$ . The response variable is the percentage of blackened skin after 5 days.

Replicates

Data are provide below.

Experimenter (Block)	Light	Storage	percentage of blackened skin			
I	1	1	30	30	17	43
	1	2	43	35	36	64
	2	1	37	38	23	53
	2	2	22	35	30	38
II	1	1	49	60	41	61
	1	2	57	46	31	34
	2	1	20	63	64	34
	2	2	40	47	62	42

III	1	1	21	45	38	39
	1	2	42	13	21	26
	2	1	41	74	24	51
	2	2	38	22	31	55

It is decided that the block does not interact with the treatment factors.

SAS code and output:

```
data banana;
  input block light storage @@;
  do rep=1 to 4; drop rep;
    input y @@;
    output;
  end;
  lines;
  1 1 1 30 30 17 43
  1 1 2 43 35 36 64
  1 2 1 37 38 23 53
  1 2 2 22 35 30 38
  2 1 1 49 60 41 61
  2 1 2 57 46 31 34
  2 2 1 20 63 64 34
  2 2 2 40 47 62 42
  3 1 1 21 45 38 39
  3 1 2 42 13 21 26
  3 2 1 41 74 24 51
  3 2 2 38 22 31 55
;
run;

proc glm data=banana;
class block light storage;
model y=block light storage light*storage;
lsmeans light storage;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1514.041667	302.808333	1.58	0.1874
Error	42	8061.875000	191.949405		
Corrected Total	47	9575.916667			

R-Square	Coeff Var	Root MSE	y Mean
0.158109	34.89086	13.85458	39.70833

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	2	1255.791667	627.895833	3.27	0.0478 ✓
light	1	80.083333	80.083333	0.42	0.5218 ✓
storage	1	154.083333	154.083333	0.80	0.3754 ✓
light*storage	1	24.083333	24.083333	0.13	0.7250 ✓

light	y LSMEAN
1	38.4166667
2	41.0000000

storage	y LSMEAN
1	41.5000000
2	37.9166667

we don't want to use the p-value b/c we argue we shouldn't use a formal hypothesis test.

Instead, we compare MS of each var. with MS<sub>Error</sub>. If bigger than MS<sub>Error</sub>, then block is a good idea.

Ex) 627.895 >>> 191.949

We see that the data do not reveal any significant differences between the lighting conditions, and between the storage methods.