

## Chapter 4

# The Randomized Block Design

### 4.1 Study Suggestions

Chapter 1 of the text introduced the CRD as a device for comparing two treatments. Chapter 2 introduced the hypothesis test for analyzing data from a CRD. Chapter 3 gave two methods, computer simulation and mathematical theory, for obtaining an approximate P-value. With this solid background, Chapter 4 presents a new design, the randomized block design (RBD), a hypothesis test for this new design's data, and an easy way to obtain an approximate P-value for the test.

In an RBD the pool of subjects is divided into blocks, and within each block a CRD is performed. Thus, you already know a great deal about RBD's since you have studied CRD's extensively.

In an RBD the researcher must choose a factor for creating blocks. For example, in the AIDS-IP study in the text, the researcher used the T-cell count to form blocks. I do not attempt to give rules for choosing blocks, but the following argument may provide you with some guidance. The goal of a CRD is to learn whether the treatments differ. This goal is difficult to achieve because subjects are so variable. More precisely, there are factors other than treatment that influence a subject's response. Any factor whose value both influences the response, and can be determined without too much trouble provides a good criterion for forming blocks. Thus, obtaining expertise in the subject area is the best way to become good at choosing blocks!

There are two major new methods developed in Chapter 4: the interaction graph, and the Mantel-Haenszel (MH) test.

Make sure you can draw an interaction graph for

an RBD. Remember that there are two possible interaction graphs, one for the proportion of successes and one for the proportion of failures.

Recall the two examples of the RBD design in the text—the AIDS-IP study and the eight Infidelity studies. The AIDS-IP study has a natural ordering to its blocks, but the blocks corresponding to the different Infidelity studies are not ordered. If the blocks are ordered, then the interaction graph should present the blocks in their natural order. If the blocks are not ordered, however, then there is no “natural” way to present the blocks in the interaction graph, and you should remember this fact when you interpret the graph.

The two treatments can be compared within each block of an RBD. If there is a consistent direction to the differences between treatments across blocks (that is, if one of the treatments is better than the other treatment in every block) then the MH test is a good way to obtain an overall (that is, across blocks) comparison of the treatments. This sounds fairly simple, but there are two difficulties.

- As Exercise 2 on page 131 of the text illustrates, if there is not a consistent direction across blocks to the differences between treatments, then it is very misleading to try to obtain an overall comparison of treatments. In these situations, simply use Fisher's test in each block, knowing that this strategy may yield different conclusions in every block.
- We do not get to see the actual difference between treatments in each block, we only see the data. The data include “chance variation,” and, hence, can be misleading. Thus, a certain

amount of judgment is required in the decision of whether to use the MH test.

Regarding the second item above, there is no absolute algorithm for making the decision on whether to use the MH test. Having said that, I will now describe my favorite approach. Within each block, I compute the standardized version,  $z$ , of the test statistic for Fisher's test. If at least one of my  $z$ 's is greater than or equal to 2, and at least one  $z$  is less than or equal to  $-2$ , then I am reluctant to use the MH test. The rationale for this strategy is quite simple. Any block with  $z \geq 2$  provides convincing evidence that treatment 1 is better than treatment 2, while any block with  $z \leq -2$  provides convincing evidence that treatment 1 is worse than treatment 2. Thus, with conflicting, convincing evidence, I am reluctant to combine evidence across blocks by using the MH test.

If every block in an RBD has exactly two subjects with one subject assigned to each treatment, then the RBD is called a randomized pairs design (RPD). The MH test simplifies tremendously for an RPD, and its name changes to McNemar's test. If  $b + c \leq 14$ , then Table A.4 yields the exact P-value for McNemar's test. If  $b + c > 14$ , however, the standard normal curve can be used to obtain an approximate P-value for McNemar's test. In this latter case, the standardized value of the test statistic has a particularly simple expression, namely

$$z = \frac{b - c}{\sqrt{b + c}}.$$

My students' main difficulty with the RPD is the distinction between the Format 1 and Format 2 contingency tables. Remember that the Format 2 table displays *pairs of subjects*. The Format 2 table for the Catcher's study appears on page 138 of the text and is reproduced below for ease of reference.

	Behind		
Front	<i>S</i>	<i>F</i>	Total
<i>S</i>	6	11	17
<i>F</i>	4	4	8
Total	10	15	25

The 6 in the "*a*" position of the table means that six pairs of subjects (throws) gave a success on each

treatment (location). The 11 in the "*b*" position of the table means that 11 pairs of subjects gave a success on the first treatment (front), and a failure on the second treatment (behind). The 17 in the " $n_1$ " position of the table means that 17 pairs of subjects gave a success on the first treatment; in other words, 17 of the 25 subjects on the first treatment yielded successes.

## 4.2 Solutions to Odd-Numbered Exercises

### Solutions for Section 4.1

1. In either block, treatment 1 has a higher proportion of successes than does treatment 2. Again in either block, the difference of the proportions is 0.10. This constancy of difference across blocks is reflected in the parallel line segments in the interaction graph. A similar pattern exists in the collapsed table; namely, the proportion of successes on treatment 1 is 10 percentage points higher than the proportion of successes on treatment 2.
3. (d.) Every student researcher obtained a larger proportion of successes with the first version than with the second. This is consistent with the pattern in the Colloquium study reported in the text in Chapter 1. The interaction graph reveals this phenomenon because its curve for version 1 is higher than its curve for version 2 in every block.

The difference between success proportions is not constant from block to block, a fact that is revealed visually by the line segments of the interaction graph not being parallel.

As in every block, the collapsed table has a higher proportion of successes with version 1 than with version 2. The collapsed table is most strongly influenced by the first block, which contains over one-half of the subjects in the study.

### Solutions for Section 4.2

1. The table below is needed to perform the various tests.

Block	$a$	$E(A)$	$\text{Var}(A)$	$z$
1	55	50	12.562	1.41
2	75	70	10.553	1.54
MH:	130	120	23.115	2.08

3. The table below is needed to perform the various tests. I did not compute the value of  $z$  for students 2–5 because their exact P-value for Fisher's test can be obtained from Table A.1.

Student	$a$	$E(A)$	$\text{Var}(A)$	$z$
1	11	9.5	3.074	0.86
2	5	3.5	1.050	
3	3	2.0	0.667	
4	3	2.0	0.667	
5	3	2.5	0.694	
MH:	25	19.5	6.152	2.22

For the first student and the first alternative, the approximate P-value is the area under the standard normal curve to the right of 0.86. This area equals 0.1949. For the second student,  $R = 8$ ,  $C = 7$ , and  $O = 5$ ; the exact P-value can be obtained from Table A.1 and is equal to 0.1573. For the remaining three students the exact P-value is greater than or equal to 0.2500.

11. Let the high dose be treatment 1 and the low dose be treatment 2. The table below is needed to perform the MH test.

Count	$a$	$E(A)$	$\text{Var}(A)$	$z$
< 200	48	46.19	2.282	1.20
200–499	395	398.43	3.684	–1.79
MH:	443	444.62	5.966	–0.66

In the low-count block, the test statistic for Fisher's test is  $z = 1.20$ , and the approximate P-value for the third alternative is  $2(0.1151) = 0.2302$ . In the high-count block, the test statistic is  $z = -1.78$ , and the approximate P-value for the third alternative is  $2(0.0375) = 0.0750$ . For the MH test, the test statistic is  $z = -0.66$ , and the approximate P-value for the third alternative is  $2(0.2546) = 0.5092$ .

13. Let the low dose be treatment 1 and the placebo be treatment 2. The table below is needed to perform the MH test.

Count	$a$	$E(A)$	$\text{Var}(A)$	$z$
< 200	48	45.59	3.972	1.21
200–499	394	385.08	6.064	3.62
MH:	442	430.67	10.036	3.58

In the low-count block, the test statistic for Fisher's test is  $z = 1.21$ , and the approximate P-value for the first alternative is 0.1131. In the high-count block, the test statistic is  $z = 3.62$ , and the approximate P-value for the first alternative is less than or equal to 0.0002. For the MH test, the test statistic is  $z = 3.58$ , and the approximate P-value for the first alternative is 0.0002.

15. Let the high dose be treatment 1 and the placebo be treatment 2. The table below is needed to perform the MH test.

Count	$a$	$E(A)$	$\text{Var}(A)$	$z$
< 200	48	43.85	3.248	2.30
200–499	395	389.30	7.666	2.06
MH:	443	433.15	10.914	2.98

In the low-count block, the test statistic for Fisher's test is  $z = 2.30$ , and the approximate P-value for the first alternative is 0.0107. In the high-count block the test statistic is  $z = 2.06$ , and the approximate P-value for the first alternative is 0.0197. For the MH test, the test statistic is  $z = 2.98$ , and the approximate P-value for the first alternative is 0.0014.

17. Let the combined doses be treatment 1 and the placebo be treatment 2. The table below is needed to perform the MH test.

Count	$a$	$E(A)$	$\text{Var}(A)$	$z$
< 200	96	91.60	4.327	2.12
200–499	789	779.39	7.554	3.50
MH:	885	870.99	11.881	4.06

### Solutions for Section 4.3

3. (c.) The standardized value of the test statistic is

$$z = \frac{b - c}{\sqrt{b + c}} = \frac{14 - 1}{\sqrt{14 + 1}} = 3.36.$$

The approximate P-value for the first alternative is the area to the right of 3.36 under the standard normal curve. This area equals 0.0004.

5. (b.) Based on Margaret's suspicion that the A fork would be more effective, I will use the first alternative.  $R = 14$  and  $O = 10$ , yielding  $P = 0.090$ .

### 4.3 Exam Questions

1. Kramer performed a study last week, but cannot remember whether it was a CRD or an RPD. He obtained  $a = 10$ ,  $b = 10$ ,  $c = 5$  and  $d = 25$ .

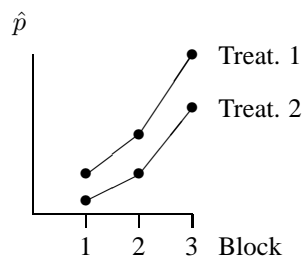
- (a) What is the value of  $\hat{p}_1$  if the study was a CRD?  
 (b) What is the value of  $\hat{p}_1$  if the study was an RPD?

2. Refer to the data presented in the previous question. Use the standard normal curve to compute the approximate P-value for the first alternative:

- (a) If the study was a CRD.  
 (b) If the study was an RPD.

3. Kramer performed a study last month, but cannot remember whether it was a CRD or an RPD, but he does remember that it was not balanced. Comment.

4. Below is an interaction graph for the proportion of successes in a Randomized Block Design.



- (a) True or false? In every block, treatment 1 had a higher proportion of successes than treatment 2.  
 (b) True or false? For either treatment, as the block number increased, the proportion of successes increased.

5. Refer to the previous question. Do you think the MH test should be used for these data? Explain your answer.

6. Suppose that an RBD has two blocks. True or false? The null hypothesis for the MH test implies, among other things, that treatment 1 is equally effective in blocks 1 and 2.

7. An RPD yields the following  $2 \times 2$  contingency table:

	Treatment 2		
Treatment 1	S	F	Total
S	8	6	14
F	3	11	14
Total	11	17	28

- (a) How many subjects were in the study?  
 (b) Compute the sample proportion of failures on the first treatment.  
 (c) Find the exact P-value for the first alternative ( $>$ ).

8. An RBD is conducted with two blocks, yielding the following information.

Block	$a$	$E(A)$	$\text{Var}(A)$
1	65	60	12.060
2	45	40	8.633

- (a) Use the standard normal curve to obtain an approximate P-value for Fisher's test with the first alternative ( $>$ ) for the data from block 1.  
 (b) Use the standard normal curve to obtain an approximate P-value for the MH test with the first alternative ( $>$ ).

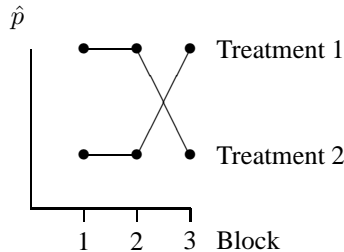
9. An RBD yields data summarized in the tables below.

Block 1			
Treatment	S	F	Total
1	10	10	20
2	32	48	80
Total	42	58	100

Block 2			
Treatment	S	F	Total
1	24	56	80
2	4	16	20
Total	28	72	100

Compute the table of row proportions for each block. Briefly explain the pattern in these data.

10. Below is an interaction graph for the proportion of successes in a randomized block design.



Remembering our convention that a success is preferred to or better than a failure, which one of the following statements is true about the evidence in the data?

- (a) Treatment 1 performed better than treatment 2 in every block.
  - (b) Treatment 1 performed worse than treatment 2 in every block.
  - (c) Treatment 1 performed better than treatment 2 in blocks 1 and 2, but performed worse in block 3.
  - (d) Treatment 1 performed worse than treatment 2 in blocks 1 and 2, but performed better in block 3.
11. An RBD with two blocks yields the following table of row proportions:

		Block 1		Total
Treatment		S	F	
1		0.30	0.70	1.00
2		0.20	0.80	1.00

		Block 2		Total
Treatment		S	F	
1		0.50	0.50	1.00
2		0.60	0.40	1.00

Draw the interaction graph for the proportions of successes.

12. Bob performs an RPD with 50 pairs of trials and plans to compare his treatments with McNemar's test.

True or false? Bob cannot use Table A.4 in the text to obtain the exact P-value because his sample size is larger than 14.

13. An RBD is conducted with three blocks, yielding the following information.

Block	$a$	$E(A)$	$\text{Var}(A)$
1	20	17.5	2.679
2	7	6.0	1.263
3	5	4.5	0.865

- (a) For the data from block 1, compute the observed value,  $z$ , of the standardized version of the test statistic for Fisher's test. (Note: Do not compute a P-value.)
  - (b) Use the standard normal curve to obtain an approximate P-value for the Mantel-Haenszel test and the first alternative ( $>$ ).
14. A medical researcher compares two treatments by using an RBD. Blocks 1 and 2 are females and males, respectively. The standardized values of Fisher's test are 3.50 and  $-3.25$  in blocks 1 and 2, respectively. Explain why use of the MH test would be inappropriate.
15. Kramer performs an RPD and obtains the results presented below.

Pair	1		2		3	
Trial	1	2	3	4	5	6
Treatment	1	2	2	1	2	1
Response	S	S	F	S	F	S

Pair	4		5		6	
Trial	7	8	9	10	11	12
Treatment	1	2	2	1	1	2
Response	S	F	F	F	F	S

Pair	7		8	
Trial	13	14	15	16
Treatment	1	2	2	1
Response	F	F	F	S

Use these data to answer the following questions.

- (a) Present these data in a  $2 \times 2$  contingency table in Format 2.  
 (b) Compute the values of  $\hat{p}_1$  and  $\hat{p}_2$ .  
 (c) Find the exact P-value for McNemar's test for the first alternative.

#### 4.4 Solutions to Exam Questions

1.  $\hat{p}_1 = 10/20 = 0.5$ ;  $\hat{p}_1 = 20/50 = 0.4$ .  
 2. (a)

$$z = \frac{\sqrt{49}[10(25) - 10(5)]}{\sqrt{20(30)(15)(35)}} = 2.49.$$

The approximate P-value is 0.0064.

(b)

$$z = \frac{10 - 5}{\sqrt{10 + 5}} = 1.29.$$

The approximate P-value is 0.0985.

3. It could not have been an RPD because an RPD must be balanced.  
 4. True; True.  
 5. Yes. The direction of evidence is consistent across blocks.  
 6. False.  
 7. (a) 56. (b)  $14/28 = 0.50$ . (c)  $R = 9$  and  $O = 6$ , giving  $\mathbf{P} = 0.254$ .  
 8. (a) For Fisher's test,

$$z = \frac{65 - 60}{\sqrt{12.060}} = 1.44,$$

giving an approximate P-value of 0.0749.

(b) By addition,  $u = 110$ ,  $E(U) = 100$ , and  $\text{Var}(U) = 20.693$ . Thus,

$$z = \frac{110 - 100}{\sqrt{20.693}} = 2.20,$$

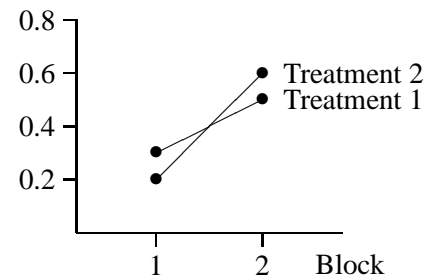
giving an approximate P-value of 0.0139.

9. The tables of row proportions are below. In each block treatment 1 has a 10-percentage-point-higher proportion of successes than treatment 2.

		Block 1		
Treatment		S	F	Total
1		0.50	0.50	1.00
2		0.40	0.60	1.00

		Block 2		
Treatment		S	F	Total
1		0.30	0.70	1.00
2		0.20	0.80	1.00

10. (d).  
 11. The interaction graph for the proportions of successes is below.



12. False. (Table A.4 requires that  $b + c \leq 14$ ; the number of pairs  $n$  can be larger.)  
 13. (a)  $z = 1.53$ .  
 (b)  $z = 1.82$ , and the approximate P-value is 0.0344.  
 14. The individual block analyses indicate that treatment 1 is superior for the females in the study and treatment 2 is superior for the males in the study. Any overall conclusion—be it that either treatment is superior or that there is no difference—will be misleading.  
 15. (a)

		Treatment 2		
Treatment 1		S	F	Total
S		1	4	5
F		1	2	3
Total		2	6	8

- (b)  $\hat{p}_1 = 5/8$  and  $\hat{p}_2 = 2/8$ .  
 (c)  $R = 5$  and  $O = 4$ , and the exact P-value is 0.188.