

# Latin squares, sample size, and power

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

## 28.4 Latin squares

Latin squares designs use two blocking variables (assumed not to interact!), however every combination of blocking levels only gets one treatment.

Must have  $r$  levels of  $\left\{ \begin{array}{l} \text{blocking variable 1} \\ \text{blocking variable 2} \\ \text{treatment} \end{array} \right\}$ .

A common example: In industrial settings, productivity (e.g. number of units produced) depends on the day of the week, machine used, operator, et cetera. These are all useful blocking variables.

If there are 4 shifts, then can only block on 4 days of the week in a Latin squares design!

## Example from manufacturing

Each of the 4 days has all 4 treatments on different shifts, every shift has all 4 treatments on different days. For example, say  $i = 1, 2, 3, 4$  denotes 4 different brands of injection molded plastic, the treatment.

	Day of the week			
	Mon	Tue	Wed	Thu
	$j = 1$	$j = 2$	$j = 3$	$j = 4$
Shift 1 $k = 1$	$i = 3$	$i = 4$	$i = 1$	$i = 2$
Shift 2 $k = 2$	$i = 1$	$i = 2$	$i = 4$	$i = 3$
Shift 3 $k = 3$	$i = 2$	$i = 1$	$i = 3$	$i = 4$
Shift 4 $k = 4$	$i = 4$	$i = 3$	$i = 2$	$i = 1$

Whole experiment over in one week!

## Gives $n_T = 16$ observations

	Day of the week			
	Mon	Tue	Wed	Thu
	$j = 1$	$j = 2$	$j = 3$	$j = 4$
Shift 1 $k = 1$	$Y_{311}$	$Y_{421}$	$Y_{131}$	$Y_{241}$
Shift 2 $k = 2$	$Y_{112}$	$Y_{222}$	$Y_{432}$	$Y_{341}$
Shift 3 $k = 3$	$Y_{213}$	$Y_{123}$	$Y_{333}$	$Y_{443}$
Shift 4 $k = 4$	$Y_{414}$	$Y_{324}$	$Y_{234}$	$Y_{144}$

In a complete blocked design, we would instead need all four treatments from each level of shift and day of the week, i.e.  $n_T = 64$  observations (four in each of the 16 cells above instead of one).

If only one type of plastic can be run for a given shift and day, the experiment would require 4 weeks to complete instead of one.

$$Y_{ijk} = \mu + \underbrace{\alpha_j}_{\text{treatment}} + \underbrace{\beta_j + \gamma_k}_{\text{blocks}} + \epsilon_{ijk},$$

a three-way additive ANOVA model. We must assume that the treatment and blocks don't interact.

Model is fit via  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{Y}$  as usual.

## Background music (pp. 1187–1188)

$r = 5$  types of music (A, B, C, D, E): slow, instrumental, and vocal; medium instrumental and vocal; fast, instrumental and vocal; medium, instrumental only; and fast, instrumental only.  $Y_{ijk}$  is day's productivity measure for a crew of bank tellers at a particular bank on  $i$ th week, day  $j$ , treatment  $k$ , where all of these indices run from 1 to 5. The results are on p. 1187.

```
data music;
input product week day treatment @@;
datalines;
  18  1  1  4  13  2  1  3  7  3  1  1  17  4  1  5  21  5  1  2
  17  1  2  3  34  2  2  2  29  3  2  4  13  4  2  1  26  5  2  5
  14  1  3  1  21  2  3  5  32  3  3  2  24  4  3  3  26  5  3  4
  21  1  4  2  16  2  4  1  27  3  4  5  31  4  4  4  31  5  4  3
  17  1  5  5  15  2  5  4  13  3  5  3  25  4  5  2  7  5  5  1
;

* gives ANOVA table on p. 1191 (in two pieces, overall ANOVA and Type III SS);
* create lines plot to duplicate results towards bottom p. 1192;
proc glm plots=all;
  class week day treatment;
  model product= week day treatment;
  lsmeans treatment / pdiff adjust=tukey alpha=0.05 cl;
```

- Saves money!!! Every block gets each treatment only once; not as many experimental units needed:  $r^2$  vs.  $r^3$ .
- Good for preliminary studies. By assuming additivity we get simple model interpretation and quick, crude estimates of treatment effects.
- Fractional factorial designs take this idea to an extreme.

# Disadvantages

- Number of levels of both blocking variables needs to be the same & must equal the number of treatment levels. This can be generalized though.
- Additivity not always reasonable & this is hard to test with incomplete block designs. Unlike RCB designs where all that was missing is replication (to estimate an interaction), with an incomplete block design we are missing necessary information.
- $r = 2 \Rightarrow \text{dfE} = 0$  and  $r = 3 \Rightarrow \text{dfE} = (r - 1)(r - 2) = 2$ . Really need  $r \geq 4$  to proceed.

# Generalizations

- More blocking variables lead to Graeco-Latin squares design.
- Random block effects.
- Factorial treatment structure.
- Adding levels to one or more blocking variables. No longer a square but analysis proceeds similarly.

	Day of the week				
	Mon	Tue	Wed	Thu	Fri
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
Shift 1 $k = 1$	$i = 3$	$i = 4$	$i = 1$	$i = 2$	$i = 3$
Shift 2 $k = 2$	$i = 1$	$i = 2$	$i = 4$	$i = 3$	$i = 2$
Shift 3 $k = 3$	$i = 2$	$i = 1$	$i = 3$	$i = 4$	$i = 1$
Shift 4 $k = 4$	$i = 4$	$i = 3$	$i = 2$	$i = 1$	$i = 4$

- Replications.
- Latin square crossover designs.

## 16.10 Sample size and power for one-way ANOVA

Let  $Y_{ij} = \mu_i + \epsilon_{ij}$  where  $i = 1, \dots, r$  and  $j = 1, \dots, n$ . A simple, balanced one-way model with  $n_T = rn$ .

Want to test  $H_0 : \mu_1 = \dots = \mu_r$ . Power of the test is  $1 - \beta = P(\text{reject } H_0 | H_0 \text{ not true})$ .

As statisticians, we are often asked one of two related questions:

- *Given these values of  $\mu_1, \dots, \mu_r$ ,  $\sigma$ ,  $\alpha$ , and  $n_T = nr$ , what is my power  $1 - \beta$ ?*
- *Given these values of  $\mu_1, \dots, \mu_r$ ,  $\sigma$ ,  $\alpha$ , what sample sizes  $n$  do I need within each group to achieve a given power  $1 - \beta$ ?*

Typically  $1 - \beta = 0.8, 0.9$ , or  $0.95$ .

Sometimes you will be given  $\delta = \frac{\max\{\mu_i\} - \min\{\mu_i\}}{\sigma}$  instead of  $\mu_1, \dots, \mu_r$  and  $\sigma$ , a *minimally detectable standardized difference*. In this case, just fix  $\sigma = 1$ .

## Some notation

Define  $\phi = \frac{1}{\sigma} \sqrt{\frac{n}{r} \sum_{i=1}^r (\mu_i - \bar{\mu}_{\bullet})^2}$  where  $\bar{\mu}_{\bullet} = \frac{1}{r} \sum_{i=1}^r \mu_i$ . Note that  $\phi \geq 0$  and  $\phi = 0 \Leftrightarrow H_0$  true.

$\phi$  is measure of “how different” the  $\mu_i$ 's are from each other.

We reject  $H_0$  if  $F^* > F(1 - \alpha; r - 1, n_T - r)$  where  $F^* = MSTR/MSE$ . If  $\phi > 0$  is true, i.e.  $H_a$  is true, then

$$F^* \sim F(r - 1, n_T - r, \phi),$$

a non-central F distribution with non-centrality parameter  $\phi$ . This is the distribution that is used to compute power and/or sample size calculations.

# Russ Lenth's JAVA Applets

The best free power and sample calculators I've found are on the web at <http://homepage.stat.uiowa.edu/~rlenth/Power/>. My browser doesn't run these directly due to security settings, so I downloaded them and run them directly off of my desktop.

Let  $\sigma_{\mu} = \sqrt{\frac{1}{r-1} \sum_{i=1}^r (\mu_i - \bar{\mu}_{\bullet})^2}$  be the standard deviation of the  $r$  cell-means. This is what goes into Russ Lenth's one-way ANOVA Applet.

Consider textbook problem 16.27 on p. 728. Here,  $\mu_1 = 22$ ,  $\mu_2 = 28$ ,  $\mu_3 = 22$ , and  $\sigma = 1.6$ . In R we find  $\sigma_{\mu} = 3.464$ . In Russ Lenth's one-way ANOVA JAVA applet plug 3.464 for SD[treatment], 1.6 for SD[Within], 3 for levels[treatment], 12 for n[Within], and 0.01 for Significance level. This yields Power[treatment] equal to 1.00.

## Minimum distance to detect

Your book discusses powering a study to detect  $\Delta = \max\{\mu_i\} - \min\{\mu_i\}$ , the smallest difference one needs to detect (reject  $H_0$ ) with a prespecified power  $1 - \beta$  and significance level  $\alpha$ . If this is actually the *largest* difference among the means, then the smallest the standard deviation occurs when two of the means are at the maximum and the minimum, and the remaining means are at  $\bar{\mu}_\bullet$ . Then  $\sigma_\mu = \sqrt{2(\Delta/2)^2/(r-1)}$ . This provides a conservative sample size  $n$  to use. If the  $r-2$  remaining means are actually different than  $\bar{\mu}_\bullet$  the power will be larger than using  $\sigma_\mu = \sqrt{2(\Delta/2)^2/(r-1)}$ .

For textbook problem 16.29,  $r = 5$ ,  $\Delta = 10, 15, 20, 30$ ,  $\beta = 0.05$ ,  $\alpha = 0.01$  and  $\sigma = 10$ . We compute  $\sigma_\mu = 3.536$  for  $\Delta = 10$  and  $\sigma_\mu = 5.303$  for  $\Delta = 15$ . Using Russ Lenth's software, we require  $n = 50$  for  $\Delta = 10$  and  $n = 23$  for  $\Delta = 15$ .

On p. 862 balanced two-way power calculations are discussed in Sec. 19.11. On p. 909 R.C.B. designs are discussed in Sec. 21.9. On p. 1021 multi-factor studies are discussed in Sec. 24.7. On p. 1193 Latin square designs are discussed in Sec. 28.6. All of these use non-central F distributions to compute power.

Russ Lenth's power and sample-size Applets can handle all of these. His approach is slightly different than your book's, and requires the use of averaged effects.

For example, in a R.C.B. design define

$$\sigma_{\rho} = \sqrt{\frac{1}{n_b - 1} \sum_{i=1}^{n_b} (\bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet})^2} \text{ and } \sigma_{\tau} = \sqrt{\frac{1}{r - 1} \sum_{j=1}^r (\bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet})^2},$$

the standard deviations of the block and treatment effects about an overall grand mean. These are used in the Randomized complete-block design Applet.

The `power` proc can help you calculate power and sample size in SAS. The following code finds the sample size  $n$  necessary to get at least 80% power for example on pp. 717–718.

```
proc power ;  
  onewayanova  
  groupmeans = 12.5 | 13 | 18 | 21  
  stddev = 3.5  
  alpha = 0.05  
  npergroup = .  
  power = .8;  
run;
```