STAT 5200 Handout #19

Latin Square Example (Ch. 13)

Motivating Example: Traffic Signal Sequence

An experiment is carried out to compare average waiting times for 5 different traffic light sequences (that is, how long the traffic lights are "green" in each direction). Because some intersections are busier than others it is thought that some intersections might have longer average waiting times than others. So, 5 intersections are selected for the experiment. Also, during some time periods of the day there is more traffic than at other periods, and this can affect waiting times, so 5 periods of day are selected for the experiment. There are 5 × 5 × 5 = 125 combinations of traffic light sequence, intersection, and period of day. It is simply not possible to run such a large experiment so, instead, the following experiment is conducted. Each traffic light sequence is paired with each intersection exactly once, and with each period of day exactly once, for a total of 25 observations. The response variable is the average waiting time.

This is a Latin Square design because experimental units are assigned to levels of → natural groupings of exp. units treatment (Sequence) in an organized fashion within two blocking factors (Period and Intersection) which together define experimental units, and all three factors (Sequence, Period, and Intersection) have the same number of levels (5). Randomization is achieved by defining [or re-labeling] treatment levels (Sequence A, B, C, D, E).

General Discussion

Sometimes there are two sources of blocking with same # of levels (5 here)

- Within this structure, treatment levels can be assigned
  ↳ same # as both blocking factors

Visualize this structure in a square table (here for traffic signal sequence example):

Blocking Factor 2 (Period)

Blocking Factor 1 (Intersection)

	1	2	3	4	5	hypothetical 6
1	A	B	C	D	E	A
2	B	C	D	E	A	C
3	C	D	E	A	B	B
4	D	E	A	B	C	E
5	E	A	B	C	D	D

All treatment levels (Sequence A, B, C, D, E) used once within each Blocking factor level

This [Latin] square gives the design its name; combinatoric structure from Euler.

In model, sometimes refer to Blocking factors as row or column effects:

$$Y_{ijk} = \mu + T_i + B1_j + B2_k + \varepsilon_{ijk}$$

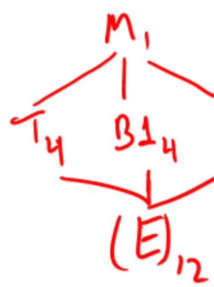
T_i: Treatment, $i = 1 \ldots 5$

$B1_j$: Blocking Factor 1 (Intersection; Row), $j = 1 \ldots 5$

$B2_k$: Blocking Factor 2 (Period; Column), $k = 1 \ldots 5$

note i, j, k have same range $1 \ldots 5$

Hasse Diagram:



why no interactions?
- not enough DF (need 16) for each pairwise int.

Recall that inference requires randomization

- In Latin Squares, achieve this by:

  randomly assign labels (or level #'s) to treatment, row, & col. variables

- See textbook page 327 for randomizing smaller squares

Ignoring block structure increases variance (MSE)

So tests tend to be less significant $\quad F_{Trt} = \frac{MS_{Trt}}{MS_E}$

Limitations of Latin Square:

- Need the same number of levels for both blocking factors and the treatment - say g
- With g levels of each factor, only have how many experimental units?

  $g^2 \Rightarrow$ so small DF for error, & not enough DF for interactions

Related to Latin Squares:

- Graeco Latin Squares (p. 343) – three sources of blocking
- Crossover Design (p.. 326) – special case where a sequence of treatments (often diets) is given to each subject over time periods

Replicating Latin Square design has challenges
- see guidelines on text p. 330 - 335

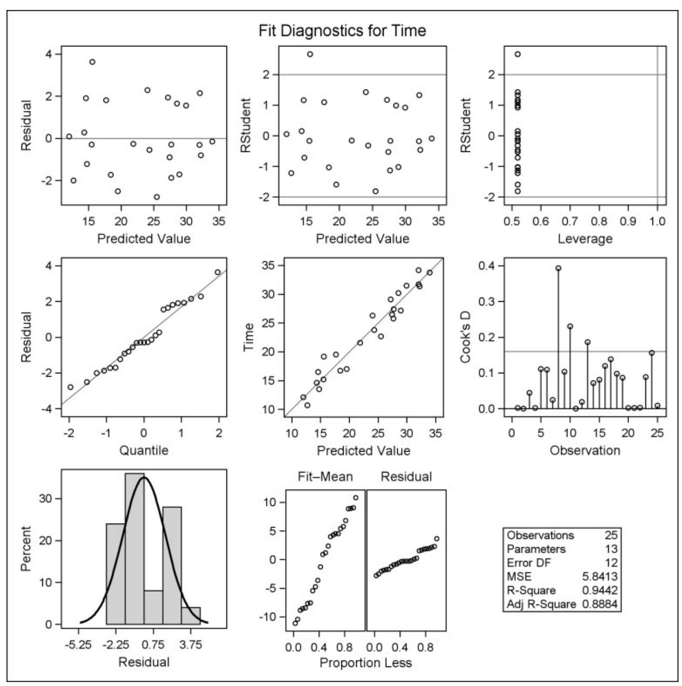
```
/* STAT 5200
   latin square design
   traffic signal sequence data
 */
/* Define options */
ods html image_dpi=300 style=journal;


/* Read in data */
data traffic;
  input Intersec Period Sequence $ Time @@;
  cards;
  1 1 A 15.2 1 2 B 33.8 1 3 C 13.5 1 4 D 27.4 1 5 E 29.1
  2 1 B 16.5 2 2 C 26.5 2 3 D 19.2 2 4 E 25.8 2 5 A 22.7
  3 1 C 12.1 3 2 D 31.4 3 3 E 17.0 3 4 A 31.5 3 5 B 30.2
  4 1 D 10.7 4 2 E 34.2 4 3 A 19.5 4 4 B 27.2 4 5 C 21.6
  5 1 E 14.6 5 2 A 31.7 5 3 B 16.7 5 4 C 26.3 5 5 D 23.8
  ;
run;

/* Fit Latin Square */
proc glm data=traffic plots=diagnostic;
  class Intersec Period Sequence;
  model Time = Intersec Period Sequence;
  lsmeans Sequence / lines adjust=tukey;
  title1 'Latin Square';
run;
```

Latin Square

Class Level Information		
Class	**Levels**	**Values**
Intersec	5	1 2 3 4 5
Period	5	1 2 3 4 5
Sequence	5	A B C D E



Dependent Variable: Time

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	1186.175200	98.847933	16.92	<.0001
Error	12	70.095200	5.841267		
Corrected Total	24	1256.270400			

R-Square	Coeff Var	Root MSE	Time Mean
0.944204	10.44998	2.416871	23.12800

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Intersec	4	18.226400	4.556600	0.78	0.5593
Period	4	1091.666400	272.916600	46.72	<.0001
Sequence	4	76.282400	19.070600	3.26	0.0498

Tukey Comparison Lines for Least Squares Means of Sequence				
LS-means with the same letter are not significantly different.				
		Time LSMEAN	Sequence	LSMEAN Number
	A	24.88	**B**	2
	A			
B	A	24.14	**E**	5
B	A			
B	A	24.12	**A**	1
B	A			
B	A	22.50	**D**	4
B				
B		20.00	**C**	3

```
/* What if Period and Intersection were ignored
   as blocking factors? */
proc glm data=traffic plots=diagnostic;
  class Sequence;
  model Time = Sequence;
  title1 'Ignoring blocking factors';
run;
```

Ignoring blocking factors

Dependent Variable: Time

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	76.282400	19.070600	0.32	0.8590
Error	20	1179.988000	58.999400		
Corrected Total	24	1256.270400			

R-Square	Coeff Var	Root MSE	Time Mean
0.060721	33.21129	7.681107	23.12800

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Sequence	4	76.28240000	19.07060000	0.32	0.8590