LECTURE 15 – LATIN SQUARE DESIGN

STAT 350

Fall 2014

Announcements

- SAMSI Undergraduate Workshop - February 26-27, 2015

  Held at SAMSI, Research Triangle Park, NC

  Deadline for applications is January 9, 2015

This year's topic: **Mathematical and Statistical Ecology**

Contrast degrees of freedom

- If we reject the null hypothesis in a one-way ANOVA and conclude at least two means are different, we often may be interested in examining relationships between the group means further with detailed research hypotheses in the form of contrasts.

- What degrees of freedom did we use for the *t*-distribution in our hypothesis tests and confidence intervals for this problem?

- Why?

- In general, what degrees of freedom should you use?

Two blocking factors

- Suppose there are two blocking factors, each with t levels.
- One treatment factor, also with t levels
- One possibility: create t^2 blocks, one for each combination of the two treatment factors
- Now do a RCB: t treatments in each of t^2 blocks implies t^3 experimental units
- Might be too expensive and/or time consuming

Latin Square Designs

- In RCB with two blocking factors, t^3 experimental units:
  - every treatment appears with every block combination
  - (turns two blocking factors into one combined blocking factor, with levels = combinations)
- In Latin square, only t^2 experimental units:
  - every treatment appear with every level of first block
  - every treatment appear with every level of second block
  - every treatment **does not** appear with every combination of the levels of the two blocks

4x4 Latin Square Layout

- Latin letters once in each row, column

	col 1	col 2	col 3	col 4
row 1	A	B	C	D
row 2	B	C	D	A
row 3	C	D	A	B
row 4	D	A	B	C

4x4 Latin Square Layout

- Latin letters once in each row, column

	col 1	col 2	col 3	col 4
row 1	A	B	C	D
row 2	B	C	D	A
row 3	C	D	A	B
row 4	D	A	B	C

- Since each treatment much appear once in each row and each column, an easy way to construct a Latin square is to
  - obtain a random permutation of letters in one row
  - shift letters over by one in subsequent rows

Analysis of variance

- ANOVA for Latin square:

Source	df	SS	MS	F	p-value
Row	$t-1$	SS**R**	MS**R**	MS**R**/MSE	not very interesting
Column	$t-1$	SS**C**	MS**C**	MS**C**/MSE	not very interesting
Treatment	$t-1$	SS**T**	MS**T**	MS**T**/MSE	
Error	$(t-1)(t-2)$	SSE	MSE		
Total	$t^2 - 1$	SSTotal			

Analysis of variance – Latin square

Source	df	SS	MS	F	p-value
Row	$t-1$	SS**R**	MS**R**	MS**R**/MSE	not very interesting
Column	$t-1$	SS**C**	MS**C**	MS**C**/MSE	not very interesting
Treatment	$t-1$	SS**T**	MS**T**	MS**T**/MSE	
Error	$(t-1)(t-2)$	SSE	MSE		
Total	$t^2 - 1$	SSTotal			

- Considerable power is lost in the tests of hypotheses for treatment comparisons unless the reduction in SSE (and MSE) due to blocking by both row and column criteria is **substantial**.

Restrictions in Latin Square

- Number of rows, columns, and treatments are all equal

Number of treatments (*t*)	Error df, *(t-1) (t-2)*
4	6 (too small?)
5	12
6	20
7	30
8	42 (too big?)

RCB design with two factors

- In general, suppose there are *a* level of blocking factor A, *b* levels of blocking factor B, and *t* treatments of interest.

- For a RCB design, *a*b*t* experimental units are needed.

Blocking Factor B

Blocking Factor A

	1	2	3	…	b
1	1,..,t	1,..,t	1,..,t	1,..,t	1,..,t
2	1,..,t	1,..,t	1,..,t	1,..,t	1,..,t
3	1,..,t	1,..,t	1,..,t	1,..,t	1,..,t
…	1,..,t	1,..,t	1,..,t	1,..,t	1,..,t
a	1,..,t	1,..,t	1,..,t	1,..,t	1,..,t

Observe **each treatment** with each **combination** of blocking factor A and blocking factor B.

Intuition behind designs

- We believe/know the blocking factors affect the response.
- We want to test if the treatment affects the response.
- We should *balance* the allocation of treatments to experimental units belonging to different blocks.
- In other words, we do not want to assign all experimental units in one level of a blocking factor to the **same** treatment because then treatment would be ***confounded*** with block.
- Solution: Latin square design (only possible when $a = b = t$)
  - Assign each treatment to each level of blocking factor A **once**
  - Assign each treatment to each level of blocking factor B **once**

RCB vs Latin square

- Suppose we have t = 5 treatments: A, B, C, D, and E.
- In addition we have two blocking factors, each with t = 5 levels.

	col 1	col 2	col 3	col 4	col 5
row 1	ABCDE	ABCDE	ABCDE	ABCDE	ABCDE
row 2	ABCDE	ABCDE	ABCDE	ABCDE	ABCDE
row 3	ABCDE	ABCDE	ABCDE	ABCDE	ABCDE
row 4	ABCDE	ABCDE	ABCDE	ABCDE	ABCDE
row 5	ABCDE	ABCDE	ABCDE	ABCDE	ABCDE

RCB design
(t^3 = 125 exp. units)

Latin square design
(t^2 = 25 exp. units)

	col 1	col 2	col 3	col 4	col 5
row 1	D	B	A	E	C
row 2	B	A	E	C	D
row 3	A	E	C	D	B
row 4	E	C	D	B	A
row 5	C	D	B	A	E

Food sales

- A Latin square design was used to investigate the amount of shelf space (six levels) on sales of powdered coffee cream. The experiment was carried out over a six-week period using different stores.

- Is the study experimental or observational?
- What is the response variable?
- What is the treatment? What are its levels? Is it quantitative? Is it fixed or random?
- What are the experimental units?
- What are the observational units?
- What are blocking factors? Are they quantitative? Are they fixed or random?

Fixed effect Latin square

- Fixed-effect Latin square model:

$$y_{ijk} = \mu + \rho_i + \gamma_j + \tau_k + e_{ijk}$$

$$i = 1, 2, ..., t \quad j = 1, 2, ..., t \quad k = 1, 2, ..., t$$

$$e_{ijk} \sim N(0, \sigma^2)$$

μ = overall mean

ρ_i = fixed effect of level *i* of row block

γ_j = fixed effect of level *j* of column block

τ_k = fixed effect of treatment *k*

e_{ijk} = experimental error

Mixed Latin square

- Mixed effect Latin square model has a random column effect

$$y_{ijk} = \mu + \rho_i + b_j + \tau_k + e_{ijk}$$

$$i = 1, 2, ..., t \quad j = 1, 2, ..., t \quad k = 1, 2, ..., t$$

$$b_j \sim N(0, \sigma_b^2) \quad e_{ijk} \sim N(0, \sigma^2)$$

μ = overall mean

ρ_i = fixed effect of level *i* of row block

b_j = random effect of level *j* of column block

τ_k = fixed effect of treatment *k*

e_{ijk} = experimental error

Random Latin square

- Random Latin square model has random row and column effects

$$y_{ijk} = \mu + a_i + b_j + \tau_k + e_{ijk}$$

$$a_i \sim N(0, \sigma_a^2) \qquad b_j \sim N(0, \sigma_b^2) \qquad e_{ijk} \sim N(0, \sigma^2)$$

μ = overall mean

a_i = random effect of level *i* of row block $i = 1, 2, ..., t$

b_j = random effect of level *j* of column block $j = 1, 2, ..., t$

τ_k = fixed effect of treatment *k* $k = 1, 2, ..., t$

e_{ijk} = experimental error

Treatment effect might also be random; same ANOVA as always, but different interpretation

Same test for treatment

- **R.** Response variable is affected by the treatment
- **A.** $H_a : \tau_1, \tau_2, ..., \tau_t$ are not all identical
- **N.** $H_0 : \tau_1 = \tau_2 = ... = \tau_t$
- **T.** $F = \dfrac{MST}{MSE}$ (the usual F-statistic)
- **D.** F-distribution with *(t-1)* and *(t-1)(t-2)* degrees of freedom
- **R., C.**

Food sales

- A Latin square design was used to investigate the amount of shelf space (six levels) on sales of powdered coffee cream. The experiment was carried out over a six-week period using different stores.

SS(weeks) = 530, SS(stores) = 6475, SS(space) = 510

Source	df	SS	MS	F	p-value
Treatment =					0.2074
Row =					
Column =					
Error		1280			
Total					

Treatment test for food sales

- **R.** The six levels of shelf space lead to different coffee cream sales
- **A.**
- **N.**
- **T.** $F = \dfrac{MST}{MSE}$ =
- **D.** F-distribution with ___ and ______ degrees of freedom
- **R.** p-value = ________
- **C.**

Treatment test for food sales

- **R.** The six levels of shelf space lead to different coffee cream sales
- **A.** $H_a : \tau_1, \tau_2, ..., \tau_t$ are not all identical
- **N.** $H_0 : \tau_1 = \tau_2 = ... = \tau_t$
- **T.** $F = \dfrac{MST}{MSE}$ **= 1.59**
- **D.** F-distribution with 5 and 20 degrees of freedom
- **R.** p-value = 0.2074
- **C.**

Were both blocking factors worthwhile?

- Relative efficiency for RCB compared RCB with **one** blocking factor to hypothetical completely randomized design with **no** blocking factor

$$RE = \frac{\text{exp. error in CRD}}{\text{exp. error in RCB}}$$

- Relative efficiency for Latin square compares Latin square with **two** blocking factors to hypothetical RCB with **one** blocking factor

$$RE = \frac{\text{exp. error in RCB w/ one factor}}{\text{exp. error in Latin square}}$$

Relative efficiency of column blocking

- If only the row blocking factor is used for blocking in hypothetical RCB, then estimated experimental error variance is

$$(MS(columns) + (t-1)MSE)/t$$

- Relative efficiency of Latin square is

$$RE = \frac{(MS(columns) + (t-1)MSE)/t}{MSE}$$

Relative efficiency of row blocking

- If only the column blocking factor is used for blocking in hypothetical RCB, then estimated experimental error variance is

$$(MS(rows) + (t-1)MSE)/t$$

- Relative efficiency of Latin square is

$$RE = \frac{(MS(rows) + (t-1)MSE)/t}{MSE}$$

Relative efficiency for food sales

- Relative efficiency of **stores** (compared to hypothetical RCB blocked on **weeks** only):

- Relative efficiency of **weeks** (compared to hypothetical RCB blocked on **stores** only):

Relative efficiency for food sales

- Relative efficiency of **stores** (compared to hypothetical RCB blocked on **weeks** only):

$$RE = \frac{(MS(stores) + (t-1)MSE)/t}{MSE} = \frac{(1295 + 5 * 64)/6}{64} = 4.21$$

- Relative efficiency of **weeks** (compared to hypothetical RCB blocked on **stores** only):

$$RE = \frac{(MS(weeks) + (t-1)MSE)/t}{MSE} = \frac{(106 + 5 * 64)/6}{64} = 1.109$$

Conclusions on relative efficiency

- Blocking on stores was **very effective**
  - there are a lot of store-to-store variation
  - RCB blocked only on weeks requires over four times as many observations
- Blocking on weeks was **not effective**
  - there is not much week-to-week variation
  - RCB blocked only on stores might require fewer observations than Latin square blocked on both

Relative efficiency versus testing

- Why do we decide whether it was worthwhile to block based on relative efficiency rather than use hypothesis tests to evaluate if there are effects of the blocking variables?