Stat 565
1-25-18

From last time:

(1)

$$SS_E = \sum\sum y_{ij}^2 - b\sum_i \bar{y}_{i\cdot}^2 - a\sum_j \bar{y}_{\cdot j}^2 + N\bar{y}_{\cdot\cdot}^2$$

$$= \sum\sum y_{ij}^2 - \frac{1}{b}\sum_i y_{i\cdot}^2 - \frac{1}{a}\sum_j y_{\cdot j}^2 + \frac{1}{N} y_{\cdot\cdot}^2$$

Let x stand for the specific y_{ij} that is missing.

let $y_{\cdot\cdot}' = y_{\cdot\cdot} - x$

$y_{i\cdot}' = y_{i\cdot} - x$

$y_{\cdot j}' = y_{\cdot j} - x$

(2)

Now,

$$SS_E = \overset{\text{everything not involving } x}{R} + x^2 - \frac{1}{b}(y_{i\cdot}' + x)^2 - \frac{1}{a}(y_{\cdot j}' + x)^2 + \frac{1}{N}(y_{\cdot\cdot}' + x)^2$$

$$\frac{\partial SS_E}{\partial x} = \cancel{2}x - \frac{\cancel{2}}{b}(y_{i\cdot}' + x) - \frac{\cancel{2}}{a}(y_{\cdot j}' + x) + \frac{\cancel{2}}{N}(y_{\cdot\cdot}' + x) \overset{\text{set}}{=} 0$$

$$x\left(1 - \frac{1}{b} - \frac{1}{a} + \frac{1}{N}\right) = \frac{y_{i\cdot}'}{b} + \frac{y_{\cdot j}'}{a} - \frac{y_{\cdot\cdot}'}{N}$$

Mult by $N = ab$

$$x\underbrace{\left(\underset{ab}{N} - a - b + 1\right)}_{(a-1)(b-1)} = a y_{i\cdot}' + b y_{\cdot j}' - y_{\cdot\cdot}'$$

(3)

$$\therefore \quad \hat{x} = \frac{a y_{i\cdot}' + b y_{\cdot j}' - y_{\cdot\cdot}'}{(a-1)(b-1)}$$

Note: If you impute a value and then run an ANOVA, the software will treat the imputed value as an actual observation.

You must <u>manually decrease df_E by 1</u> and recompute MSE & F & pvalue.

(4)

<u>The Latin Square Design</u>

As in RCBD, there is 1 treatment factor and 2 blocking factors

<u>Example</u>: Our treatment factor is temperature

We think there may be differences between batches of raw material. Also, operator will be blocked.

We will require every factor to have p levels.

Suppose $p = 3$

This should require 27 runs.

We can do it in 9 runs.

		Operator 1	Operator 2	Operator 3
Batch	1	A	B	C
	2	B	C	A
	3	C	A	B

Model: $y_{ijk} = \mu + \underbrace{\tau_i}_{\text{treatment}} + \underbrace{\alpha_j}_{\text{batch}} + \underbrace{\beta_k}_{\text{operator}} + \varepsilon_{ijk}$ $\quad i,j,k = 1,\dots,p$

6

Matrix form $(p=3)$

$$\begin{bmatrix} y_{111} \\ y_{123} \\ y_{132} \\ \hline y_{212} \\ y_{221} \\ y_{233} \\ \hline y_{313} \\ y_{322} \\ y_{331} \end{bmatrix}_{9\times1} = \left[\begin{array}{c|cc|cc|cc} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & 0 & 1 \\ \hline 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & -1 & -1 & -1 & -1 \\ \hline 1 & -1 & -1 & 1 & 0 & -1 & -1 \\ 1 & -1 & -1 & 0 & 1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 0 \end{array}\right]_{9\times7} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{7\times1} + \vec{\varepsilon}_{9\times1}$$

$$Y = X\beta + \varepsilon$$

Parameter estimates are found using least squares (7)

$\hat{\mu} = \bar{y}_{...}$, $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$, $\hat{\alpha}_j = \bar{y}_{.j.} - \bar{y}_{...}$,

$$\hat{\beta}_k = \bar{y}_{..k} - \bar{y}_{...}$$

Then $\hat{y}_{ijk} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...})$

$$= \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - 2\bar{y}_{...}$$

Decompose $SS_{TOT} = \sum\sum\sum(y_{ijk} - \hat{y}_{ijk} + \hat{y}_{ijk} - \bar{y}_{...})^2$

$$= SS_{TRT} + SS_{ROW} + SS_{COL} + SS_E$$

(8)

$$SS_{TRT} = \sum_{i=1}^{p} \frac{y_{i..}^2}{p} - \frac{y_{...}^2}{N}$$

$$SS_{ROW} = \sum_{j=1}^{p} \frac{y_{.j.}^2}{p} - \frac{y_{...}^2}{N} \qquad N = p^2$$

$$SS_{COL} = \sum_{k=1}^{p} \frac{y_{..k}^2}{p} - \frac{y_{...}^2}{N}$$

SS_E is found by $SS_{TOT} - SS_{TRT} - SS_{ROW} - SS_{COL}$

(9)

ANOVA table

Source	SS	df	MS	F
TRT	SS_{TRT}	$p-1$	$\frac{SS}{df}$	MS_{TRT}/MS_E
ROW	SS_{ROW}	$p-1$		—
COL	SS_{COL}	$p-1$		—
ERR	SS_E	$(p-1)(p-2)$		—
TOT	SS_{TOT}	$N-1$	—	—

$$df_E = p^2-1-3(p-1) = p^2-3p+2$$
$$= (p-1)(p-2)$$

(10)

Replication in the Latin Square

Case 1: Use the original design & run n replications at each point in the grid

Source	df
TRT	$p-1$
ROW	$p-1$
COL	$p-1$
REP	$n-1$
ERR	$np^2-3p-n+3$ ← $np^2-1-3(p-1)-(n-1)$
TOT	$N-1 = np^2-1$

(11)

Case 2: Use the same batches throughout, but change operators with each replication

Source	df
TRT	$p-1$
ROW	$p-1$
COL	$n(p-1)$
REP	$n-1$
ERR	$(p-1)(np-2)$
TOT	$N-1 = np^2-1$

$\leftarrow np^2-1-2(p-1)-n(p-1)-(n-1)$

(12)

Case 3: Use new batches of material & new operators in each replication

Source	df
TRT	$p-1$
ROW	$n(p-1)$
COL	$n(p-1)$
REP	$n-1$
ERR	$(p-1)[n(p-1)-1]$ by subtraction
TOT	$N-1 = np^2-1$

HW #3 due Feb 1 p. 177 4.8, 4.9

4.8. A consumer products company relies on direct mail marketing pieces as a major component of its advertising campaigns. The company has three different designs for a new brochure and wants to evaluate their effectiveness, as there are substantial differences in costs between the three designs. The company decides to test the three designs by mailing 5000 samples of each to potential customers in four different regions of the country. Since there are known regional differences in the customer base, regions are considered as blocks. The number of responses to each mailing is as follows.

	Region			
Design	**NE**	**NW**	**SE**	**SW**
1	250	350	219	375
2	400	525	390	580
3	275	340	200	310

(a) Analyze the data from this experiment.

(b) Use the Fisher LSD method to make comparisons among the three designs to determine specifically which designs differ in the mean response rate.

(c) Analyze the residuals from this experiment.

4.9. The effect of three different lubricating oils on fuel economy in diesel truck engines is being studied. Fuel economy is measured using brake-specific fuel consumption after the engine has been running for 15 minutes. Five different truck engines are available for the study, and the experimenters conduct the following randomized complete block design.

	Truck				
Oil	**1**	**2**	**3**	**4**	**5**
1	0.500	0.634	0.487	0.329	0.512
2	0.535	0.675	0.520	0.435	0.540
3	0.513	0.595	0.488	0.400	0.510

(**a**) Analyze the data from this experiment.

(**b**) Use the Fisher LSD method to make comparisons among the three lubricating oils to determine specifically which oils differ in brake-specific fuel consumption.

(**c**) Analyze the residuals from this experiment.