Unit 6: A Latin Square Design

6.1. The Data

An oil company tested four different blends of gasoline for fuel efficiency according to a Latin square design in order to control for the variability of four different drivers and four different models of cars. Fuel efficiency was measured in miles per gallon (mpg) after driving cars over a standard course.

Fuel Efficiencies (mpg) For 4 Blends of Gasoline
(Latin Square Design: Blends Indicated by Letters A-D)

<table>
<thead>
<tr>
<th>Driver</th>
<th>Car Model</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>D</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>II</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>3</td>
<td>III</td>
<td>C</td>
<td>A</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>IV</td>
<td>A</td>
<td>D</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

These data are from Ott: *Statistical Methods and Data Analysis*, 4th ed., Duxbury, 1993, page 866. (Similar data are given in the 6th edition by Ott/Longnecker, in problem 15.10, page 994.)

The design is called a "Latin" Square design because Latin letters are often used to show how levels of one factor are assigned to combinations of levels of the other two factors. It is often the case that one of the effects is of principal interest (here, Blend), while the other two are "blocking" effects to control variability (here, Driver and Model). For our data, notice that each Driver tests each of the four Blends; also each Blend is tested in each Model of car. However, not all combinations of Blend-Model-Driver are present. For example, Driver 3 did not test Blend A in Model IV. In a Latin square each of the three factors must have the same number \( t \) of levels, and there are \( n_T = t^2 \) observations altogether.

Problems

6.1.1. This Latin Square design has \( 4^2 = 16 \) observations. As noted just above not all 3-way combinations of the three effects are present. How many observations would have been required in order to include all 3-way combinations of Blend, Model, and Driver?

6.1.2. Suppose that there were 5 Blends, 5 Drivers, and 5 Models. A full "factorial" design with one observation on each 3-way combination would require 125 observations. Make a table showing how you could assign Blends A-E to make a Latin Square design for this situation.

6.1.3. For your convenience, the 16 observations in the fuel efficiency study are listed below (reading across the rows of the table above):

15.5 33.9 13.2 29.1 16.3 26.6 19.4 22.8 10.8 31.1 17.1 30.3 14.7 34.0 19.7 21.6.

Put these observations into c1 of a Minitab worksheet. Use the patterned data procedure to put subscripts for Driver into c2 and Model into c3 (use 1 for I, 2 for II, etc.). The subscripts for Blend (use 1 for A, 2 for B, etc.) will have to be entered directly into the worksheet one at a time because the Latin Square pattern cannot be entered automatically by Minitab. (Use the original data table in
Label the four columns appropriately: MPG, Driver, Model, Blend.

When you are finished entering the data, print out the results. Here is what you should get (without the spaces between some of the lines, which have been introduced for easy reading):

MTB > print c1-c4

<table>
<thead>
<tr>
<th>ROW</th>
<th>MPG</th>
<th>Driver</th>
<th>Model</th>
<th>Blend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.5</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>33.9</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>13.2</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>29.1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>16.3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>26.6</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>19.4</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>22.8</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>10.8</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>31.1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>17.1</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>30.3</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>14.7</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>34.0</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>19.7</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>21.6</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

6.1.4. Because this experiment was done by an oil company it is reasonable to assume that the main issue is whether there are differences among the four Blends, and that Blend is a fixed effect because four particular blends are currently under study. Also, it may be reasonable to assume that Models of car and Drivers were chosen at random from among available models and drivers. Subject to these assumptions, write the model for this experiment. In specifying the range of the subscripts for Driver you can use "$i = 1, 2, 3, 4." and for Model you can use "$j = 1, 2, 3, 4." However, for Blend you need to say something like "the 4 values of the Blend subscript $k$ are assigned to the 16 observations according to a Latin Square design," in order to make it clear that there are only 16 observations.

6.1.5. The design space of this Latin Square is really a very carefully chosen subset of a cube. There are $4^3 = 64$ possible combinations of the levels of the three factors, of which only 16 are observed. The table at the beginning of this unit can be viewed as a two-dimensional representation of the cube with dimensions $z = \text{Driver}$ (height), $y = \text{Model}$ (width), and $x = \text{Blend}$ (depth). Minitab makes three-dimensional scatterplots; in Minitab 15 they can be conveniently rotated.

GRAPH > 3D > Groups, z='Driver", y='Model', x='Blend', Group=Blend;
Data view: symbols and lines.
TOOLS > Toolbars > 3D Graph, rotate about appropriate axes

The initial view is shown first, followed by the view that results when the plot is rotated so that its $z$-axis is perpendicular to the viewer (so that the connecting lines disappear behind the plotting symbols). This is like viewing the initial view from the top.

In Minitab (release 15 or higher), use the tools to rotate such a plot so that, first, the $y$-axis is perpendicular to your view, and then the $x$ axis. Also, with some fussing, you should be able to orient the cube to match the data table. Show a copy of one or both results. (If visible on your copy of this unit, the colors of the symbols for Blend match the colors of the letters in the data table in section 6.1.)
3D Scatterplot of Driver vs Model vs Blend

Blend

1

2

3

4

Driver

Model

Blend

1

2

3

4
6.2. ANOVA for a Latin Square Design — Interpretation of Results

Although Latin square designs have the particular kind of "balance" described in of the previous section, these designs are not balanced in the sense that each combination of the three factors occurs equally often. Some combinations occur once (e.g., Driver 1, Model I, and Blend D). But there are only 16 observations, and so most of the combinations do not occur at all (e.g., Driver 1, Model I, and Blend A). For this reason Minitab's "balanced ANOVA" procedure does not work for Latin square designs.

The word cube may be more descriptive of these designs than square. Imagine this particular design as a cube with 64 cells and try to visualize which 16 cells are filled (i.e. contain an observation) and which are not. What would the cube look like when viewed from each face? In a Latin Square design all three effects must have the same number $t$ of levels, called the order of the Latin Square. Here $t = 4$.

Because the Latin Square design is not balanced as required by Minitab's "Balanced ANOVA" procedure, we must use the "General Linear Model" procedure in order to obtain an ANOVA table for the fuel efficiency data.

STAT > ANOVA > General linear model

MTB > glm MPG = Driver Model Blend;
SUBC> random Driver Model;
SUBC> resid c5;
SUBC> ems.

General Linear Model

<table>
<thead>
<tr>
<th>Factor</th>
<th>Type</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver</td>
<td>random</td>
<td>4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Model</td>
<td>random</td>
<td>4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Blend</td>
<td>fixed</td>
<td>4</td>
<td>1 2 3 4</td>
</tr>
</tbody>
</table>

Analysis of Variance for MPG, using Adjusted SS for Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driver</td>
<td>3</td>
<td>5.897</td>
<td>5.897</td>
<td>1.966</td>
<td>0.50</td>
<td>0.699</td>
</tr>
<tr>
<td>Model</td>
<td>3</td>
<td>736.912</td>
<td>736.912</td>
<td>245.637</td>
<td>61.90</td>
<td>0.000</td>
</tr>
<tr>
<td>Blend</td>
<td>3</td>
<td>108.982</td>
<td>108.982</td>
<td>36.327</td>
<td>9.15</td>
<td>0.012</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>23.809</td>
<td>23.809</td>
<td>3.968</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>875.599</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Note: Expected Mean Squares, Error Terms, and Variance Components have been omitted here. See Problem 6.2.3.)

We conclude that the blends are significantly different at the 5% level, but not at the 1% level. It is not surprising that the fuel efficiencies vary among various models of cars (which might range from small compacts to large RVs). It does not appear, however, that there are any significant differences among our Drivers. (In the general population some drivers are easier on fuel than others; perhaps the drivers for this study have been carefully trained so that their driving styles are unusually similar.)

Notes on orthogonality and GLM:

- The GLM procedure gives two types of Sums of Squares: here the Sequential Sums of Squares are the same as the Adjusted Sums of Squares because the special structure of the Latin Square design ensures that the three 3-dimensional subspaces corresponding to the three effects Driver, Model, and Blend are orthogonal.
- A signal that a design is not orthogonal would be that the Sequential and Adjusted sums of squares do not agree.
- In its GLM output, Minitab "tries" to compute F-tests and P-values for all effects. These are correct for our orthogonal design, but for some severely non-orthogonal designs, this information might be misleading.

**Problems**

6.2.1. Use the following procedure and the stacked-subscripted data to make a table that resembles the original data table of Section 1.

\[
\text{STAT > Tables > Cross tabulation,} \\
\text{classification variables: 'Driver' 'Model',} \\
\text{Summaries, associated variables 'MPG' 'Blend',} \\
\text{checkbox, data} \\
\text{MTB > table 'Driver' 'Model';} \\
\text{SUBC> data 'MPG' 'Blend'.}
\]

Now, by hand, make a table similar to the original data table, except that the rows are Drivers, the columns are Blends, and the labels within cells are Roman numerals I-IV designating Models. Verify your result in Minitab with a procedure similar to the one shown just above. Repeat (both by hand and in Minitab) for a table in which Drivers are designated within the cells.

6.2.2. Try to use Minitab's balanced ANOVA procedure (menu path STAT > ANOVA > Balanced ANOVA or command ANOVA) to analyze this block design. What happens? (A Latin Square is "balanced" in the sense that the four subspaces corresponding to the three factors and error are orthogonal. In Minitab's "Balanced ANOVA" all combinations of treatment levels must have the same number of observations; here 16 frequencies are 1 and 64 – 16 = 48 are 0.)

6.2.3. Repeat the GLM procedure shown in this section to obtain the EMS table and components of variance, and make a column of residuals. (Notice that we have not used the restrict subcommand because it is not available with GLM. For a Latin Square design restriction makes no difference in how the F-ratios are formed.) Answer the following questions:

(a) Which mean square is in the denominator of each F-ratio. How do the EMSs lead you to the conclusion that this is correct?
(b) What is strange about the component of variance for Driver? How do you interpret this result in practice?
(c) What is the P-value of the Anderson-Darling test for normality of the residuals, and how do you interpret it?

6.2.4. Multiple comparison procedures for a Latin Square design. We have established that there are significant differences among the Blends. By hand, establish the pattern of significant differences using Fisher's LSD procedure and Tukey's HSD procedure (both at the 5% level). The formulas are similar to the ones for a one-way ANOVA, except that you must use MS(Error) from the ANOVA table as the variance estimate (i.e., instead of $s_w^2$) and that $n = t$. (The exact formula for LSD is given in O/L 6e, page 1146; use second formula, not the one for use with missing data.)

Note: For Latin square designs, the Minitab GLM procedure performs Tukey's HSD and other multiple comparisons (but not Fisher's LSD). The output format of GLM is somewhat different from the one Minitab uses for comparisons in one-way ANOVAs; both confidence intervals and tests are available. We suggest that you do the Fisher and Tukey comparisons using a calculator and then verify your answers for Tukey comparisons using GLM. (As we see in Section 6.3, you will get incorrect results if you try to use a one-way ANOVA procedure on data from a Latin Square design, so that is not a feasible procedure for getting LSDs for our data).
6.2.5. Missing data in a Latin Square design.

Suppose that the MPG value for Driver 3/Model II is missing. Copy c1 'MPG' to c11 and name the copied column 'MPGM'. Replace the appropriate observation with a * (the asterisk is Minitab's symbol for a missing observation), and run glm c11 = c2 c3 c4. How can you tell from the output that the Latin square design with a missing value is not an orthogonal design? Try running both glm c11 = c2 c4 c3 and glm c11 = c4 c2 c3. What changes and what remains the same? Interpret Minitab's (approximate) F-ratios.

Note: What Minitab calls a sequential sum of squares (Seq SS), SAS calls a Type I sum of squares. What Minitab calls an adjusted sum of squares (Adj SS), SAS calls a Type III sum of squares.

Within reason, this method can be used for more than one missing value. The more missing values, the harder it is to draw valid inferences. Some extreme configurations of missing values cannot be run successfully.

6.3. Connections With Simpler ANOVA Designs

If we incorrectly ignored both the Driver and the Model structure of the design, we might treat the data as a one-way ANOVA with Blend as the only effect to be tested. This procedure is shown below.

**INAPPROPRIATE ANALYSIS**

STAT > ANOVA > Oneway

MTB > oneway 'MPG' 'Blend'

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blend</td>
<td>3</td>
<td>109.0</td>
<td>36.3</td>
<td>0.57</td>
<td>0.646</td>
</tr>
<tr>
<td>ERROR</td>
<td>12</td>
<td>766.6</td>
<td>63.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>15</td>
<td>875.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

INDIVIDUAL 95 PCT CI’S FOR MEAN

BASED ON POOLED STDEV

| LEVEL | N  | MEAN | STDEV | -------+---------+---------+
|-------|----|------|-------|--+---------+---------|
| 1     | 4  | 23.575| 7.818 | (----------*---------)
| 2     | 4  | 25.050| 8.388 | (----------*---------)
| 3     | 4  | 18.050| 7.344 | (----------*---------)
| 4     | 4  | 22.350| 8.375 | (----------*---------)

POOLED STDEV = 7.993

Notice that the SS(Blend) here is the same as in the Latin square ANOVA table, and that SS(Error) here gets split into SS(Model), SS(Driver), and SS(Error) in the Latin square table; similarly for degrees of freedom. Finally, notice that the bogus one-way analysis fails to detect the Blend effect.

**Problems**

6.3.1. Perform two additional incorrect one-way ANOVAs, using first Driver and then Model as the single factor. If you were forced to use a spreadsheet program that will compute only one-way ANOVAs, how could you piece together results from this program to construct a correct ANOVA table for a Latin Square design? In other words, how can you combine information from three incorrect one-way ANOVAs to give the correct analysis of a Latin Square?

6.3.2. Perform incorrect analysis on these data, treating them as if they came from a block design with Blends as the fixed effect and Models as the blocking effect (ignoring Drivers). Compare sums of squares and degrees of freedom in the resulting ANOVA table with the correct ones from the Latin Square analysis. If you were to use the F-ratios from this incorrect procedure to draw conclusions about the significance of Blend and Model effects, would your conclusions happen to
be correct or incorrect? What about the conclusions drawn from an incorrect block design that ignores Models instead of Drivers? Comment.