Latin Square Design: Step-by-Step Worksheet

Davood Tofighi, Ph.D.

Background / Scenario

You want to compare **four teaching methods** (A, B, C, D) to see their effect on student performance. Meanwhile:

- **Teacher** differences (labeled T1–T4) may affect outcomes.
- **Classroom** differences (labeled R1–R4) may also affect outcomes.

You suspect large teacher-to-teacher (row) differences and classroom-to-classroom (column) differences. You decide to use a **Latin Square** design, with rows = teachers, columns = classrooms, so each row and column has exactly one instance of each teaching method.

Example Layout (one possible random assignment)

	R1	R2	R3	R4
T1	A	B	C	D
T2	B	C	D	A
T3	C	D	A	B
T4	D	A	B	C

> Each cell is the assigned teaching method for that Teacher–Classroom combination. In practice, label rows and columns randomly, and also label A/B/C/D randomly, so the order above is just an example.

Learning Objective

Goal: Understand how a Latin Square design blocks on two nuisance factors—Teachers (rows) and Classrooms (columns)—while comparing four primary treatments (teaching methods). By the end of this worksheet, you should be able to:

1. Construct a valid 4x4 Latin Square layout.
2. Outline the data collection process and identify the measured responses.
3. Compute the analysis-of-variance (ANOVA) decomposition for a Latin Square.
4. Interpret which factors (rows, columns, or treatments) have significant effects.

Step-by-Step Scaffolding Tasks

Part A: Setting Up the Latin Square

1. **Identify the row factor**: Teachers T1–T4 are assigned as the rows.
2. **Identify the column factor**: Classrooms R1–R4 are assigned as the columns.
3. **Label the four teaching methods**: A, B, C, D.
4. **Fill in a 4×4 table** so that each letter (A/B/C/D) appears exactly once in each row and once in each column.

   Guiding Question: Why must each treatment appear exactly once per row and column?

(Hint: Each row factor sees all methods once, each column factor sees all methods once. This balances out row and column differences.**)**

Part B: Data Collection Planning

1. **Measurement**: Decide how you will measure student performance. Will you use test scores, or average quiz scores? Are there multiple students in each cell?
2. **Data Table**: Sketch out your blank 4×4 grid. Each cell will hold the observed average test score (for example) under the assigned teaching method, with that particular teacher–classroom combination.
3. **Randomization**: Randomly label T1–T4 and R1–R4, and also randomly assign methods A–D to the four symbols in the square pattern.

   Guiding Question: How can randomization reduce systematic bias?

Part C: Analysis (ANOVA) Steps

Using your completed data table (once you have real or example scores):

1. **Compute row means** (Teacher means). For teacher T1, average the four cells in row T1. Repeat for T2, T3, T4.
2. **Compute column means** (Classroom means). For classroom R1, average the four cells in column R1, etc.
3. **Compute treatment means** (Method A, B, C, D). Each method appears exactly once in each row and column, for 4 total observations per method.
4. **Find the grand mean** of all 16 responses,

$$\bar{Y} = \frac{1}{16} \sum_{i,j} Y_{ij}.$$

5. **Sums of Squares**:

   - **Total SS**:

$$\text{SST} = \sum_{i=1}^{4}\sum_{j=1}^{4}(Y_{ij} - \bar{Y})^2.$$

- **Row SS** (Teachers):

$$\text{SS}_{\text{rows}} = 4\sum_{i=1}^{4}(\bar{Y}_{\text{row},i} - \bar{Y})^2.$$

- **Column SS** (Classrooms):

$$\text{SS}_{\text{cols}} = 4\sum_{j=1}^{4}(\bar{Y}_{\text{col},j} - \bar{Y})^2.$$

- **Treatment SS** (Methods A/B/C/D):

$$\text{SS}_{\text{treat}} = 4\sum_{k=A,B,C,D}(\bar{Y}_k - \bar{Y})^2.$$

- **Error SS**:

$$\text{SS}_{\text{error}} = \text{SST} - (\text{SS}_{\text{rows}} + \text{SS}_{\text{cols}} + \text{SS}_{\text{treat}}).$$

6. **Degrees of Freedom**:
   - Rows, Columns, Treatments each has 3 degrees of freedom.
   - Total df = 16–1 = 15.
   - Error df = 15–(3 + 3 + 3) = 6.
7. **Mean Squares**: MS = SS / df for Rows, Columns, Treatments, and Error.
8. **F-ratios**:
   - $F_{\text{rows}} = \frac{\text{MS}_{\text{rows}}}{\text{MS}_{\text{error}}}$.
   - $F_{\text{cols}} = \frac{\text{MS}_{\text{cols}}}{\text{MS}_{\text{error}}}$.
   - $F_{\text{treat}} = \frac{\text{MS}_{\text{treat}}}{\text{MS}_{\text{error}}}$.

   Guiding Question: What does it mean if F_{treat} is large compared to 1?

Part D: Reflection / Discussion Questions

1. **Interpretation**:
   - If the F-ratio for Treatments is significant, what does that imply about the four methods?
   - How would you follow up with pairwise comparisons?
2. **Block Factors**:
   - Did you see a significant Teacher effect?
   - Did you see a significant Classroom effect?
3. **Check Assumptions**:

- How would you evaluate normality, constant variance, or outliers in this design?
- Are there any clues that the teacher or classroom might interact with the teaching method?

Think about: If a teacher–method interaction is suspected, a Latin Square might not fully handle that. You might need a more complex design.

Optional Challenge

1. **Replication**: Suppose you can replicate the entire 4×4 Latin Square twice. How does that change your Error df, and how might it improve your ability to detect differences?
2. **Higher Factor Interactions**: If teacher–classroom–method interactions are suspected, what alternative designs could address those? (Hint: Possibly a Graeco-Latin square or more advanced row–column approach.)
3. **Data Visualization**: Plot row effects, column effects, and method effects on a single chart. Compare the relative spread in each to see visually which factor has the largest influence.