Replicates and Experimental Errors

- **Replicate**: replication of same treatment

\[
y = \hat{f}(X_1, X_2, \ldots, X_m) + \epsilon = \hat{f}(X_1, X_2, \ldots, X_m) + \hat{\epsilon}
\]

- **Q**: Why do we need to understand the magnitude of exp’tal error?

  *We need to know \( \text{Var}(\epsilon) \) so that we can judge whether an effect is (statistically) significant relative to the error.*
Use of a chance mechanism (e.g., random number generators) to assign treatments to units or to run order. It has the following advantages.

- Protect against latent variables or “lurking” variables (give an example).
- Reduce influence of subjective bias in treatment assignments (e.g., clinical trials).
- Ensure validity of statistical inference (This is more technical; will not be discussed in the book. See Chapter 4 of “Statistics for Experimenters” by Box, Hunter, Hunter for discussion on randomization distribution.)

**true model:**
\[ Y = X\beta + Z\xi + \varepsilon \]

**fitted model:**
\[ Y = X\hat{\beta} + \varepsilon' \]

\[ HZ\xi = \varepsilon \]

**Q:** what if operators have an effect on response?

**Q:** Is aliasing/confounding always a bad thing?

- pros & cons
Randomization provides protection against extraneous factors that are unknown to the experimenter but may impact the response.

- **What should be randomized?**
  - Allocation of experimental materials to treatments; the order of applying treatments; the order of measuring responses; ...

Randomization is similar to a firewall or an immune system. It prevents the unknown effects from influencing the results. Randomization is the process of assigning treatments to subjects in a way that is unpredictable to the experimenter and thus ensures that the results are not biased by any unknown factors.

A block refers to a collection of homogeneous units. Effective blocking aims to minimize the variation between blocks to maximize the variation within blocks. Examples include hours, batches, lots, street blocks, or pairs of twins.

- **Can try to achieve \( X \perp Z \) by design \( \Rightarrow \) orthogonality\**

- **Block what you can and randomize what you cannot.**

- **Discuss typing experiment to demonstrate possible elaboration of the blocking idea.** See LNp.1-24.
Blocking

- **Known Z or systematic structure in EUs**
- **Block factor**: factors that are controllable and may influence the response but in which we are not directly interested. (cf. treatment factors)
- **Homogeneous EUs or same Z value**

**Q**: If operator effect is identified as significant before exp’t, what can we do?

Examples of blocking factors:
- lot-to-lot, brand-to-brand, operator-to-operator, day-to-day, ...

**Illustation: Typing Experiment**

- To compare two keyboards A and B in terms of typing efficiency. Six manuscripts 1–6 are given to the same typist.
- Several designs (i.e., orders of test sequence) are considered:
  1. **Extra variable (factor or unknown):** typing order
     - 2 levels: I, II, block factor.
     - 6 levels: I–VI
  2. Randomizing the order leads to a new sequence like this

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>I I</td>
<td>I I</td>
<td>I I</td>
<td>I I</td>
<td>I I</td>
<td>I I</td>
</tr>
<tr>
<td>II</td>
<td>I II</td>
<td>I II</td>
<td>I II</td>
<td>I II</td>
<td>I II</td>
<td>I II</td>
</tr>
</tbody>
</table>

(An improvement, but there are four with A, B and two with B, A. Why is this not desirable? Impact of learning effect.)

- **Balanced randomization**: To mitigate the learning effect, randomly choose three with A, B and three with B, A. (Produce one such plan on your own).
- Other improved plans?

jointly made by Jeff Wu (GT, USA) and S.-W. Cheng (NTHU, Taiwan)
**Reading:** textbook, 1.1, 1.2, 1.3