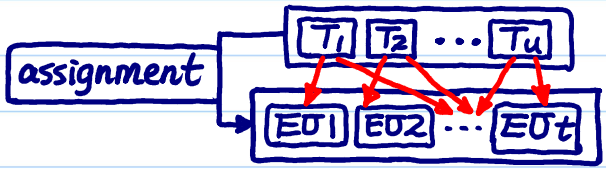
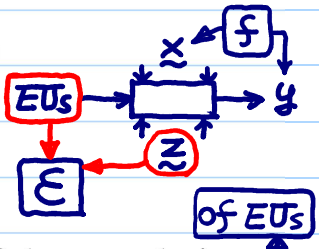
Fundamental Principles : Replication, randomization, and blocking

← Fisher's work (LNp 1-6)

multiple observations (y) per treatment





Replication

- Each treatment is applied to units that are representative of the population (example : measurements of 3 units vs. 3 repeated measurements of 1 unit).
  - Q: What are the "sources of variation" in ε ?
- Replication vs Repetition (i.e., repeated measurements).
- Enable the estimation of experimental error. Use sample standard deviation.
  - "true" error variance (Recall. test for lack of fit, LM, LNp.6-6~11)
- Decrease variance of estimates and increase the power to detect significant differences : for independent y_i's,

$\bar{y}$ → estimator of $\mu_{T_j} = E(y_{T_j})$

$$Var\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) = \frac{1}{N} Var(y_1).$$

Var(ε) ; # of replicates

Replicates and Experimental Errors

exp'ters often resist to perform replicates Why ?

replicate: replication of same treatment

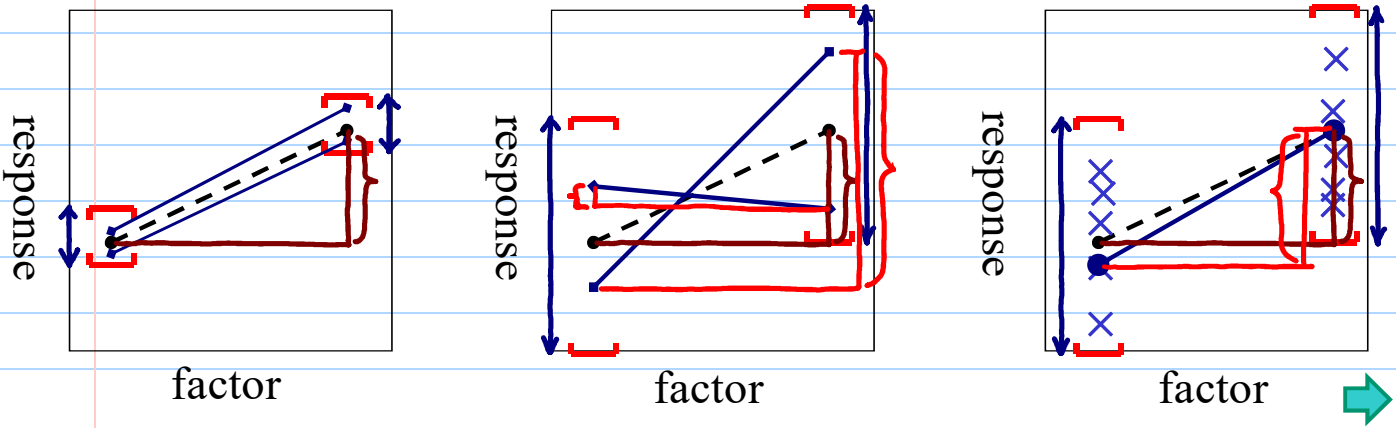
$$y = f(X_1, X_2, \ldots, X_m) + \epsilon$$
$$= \hat{f}(X_1, X_2, \ldots, X_m) + \hat{\epsilon}$$

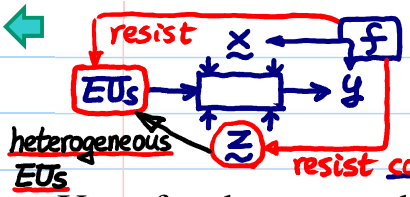
lack of fit ; over-fitting

$\hat{\beta}$ / s.e.($\hat{\beta}$) $\propto \frac{1}{\sqrt{n}} \hat{\sigma}$ — sample size ; var est'or

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

We need to know Var(ε) so that we can judge whether an effect is (statistically) significant relative to the error.





Randomization

for uncontrollable / unknown / nonmeasurable z, EUs

Use of a chance mechanism (e.g., random number generators) to assign treatments to units or to run order. It has the following advantages. ↑ LNp.1-17

independent btw x & z ⟷ cf. orthogonality btw x & z.

- Protect against latent variables or "lurking" variables (give an example).
- Reduce influence of subjective bias in treatment assignments (e.g., clinical trials).
  - ↳ single-blind, double-blind, ...
- 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.)

population of EUs → without replacement → EUs
population of EUs → ss with → i.i.d. → ε

true model: $Y = X\beta + Z\gamma + \varepsilon$

fitted model: $Y = X\beta + \varepsilon'$

$Z\gamma = HZ\gamma + (I-H)Z\gamma$

bias $\hat{\beta} \Rightarrow E(\hat{\beta}) = \beta + (X^TX)^{-1}X^TZ\gamma$

$X(X^TX)^{-1}X^T$ ← hat matrix

If $X \perp Z$ ($X^TZ = 0$) $\Rightarrow HZ\gamma = 0 \Rightarrow E(\hat{\beta}) = \beta$

design matrix / planning matrix

Effect Aliasing/Confounding

混淆(污染)

cf. ⟷ collinearity

confounded

	A	B	C	Operator
1	low	low	low	Peter
2	low	low	high	Peter
3	low	high	low	Peter
4	low	high	high	Peter
5	high	low	low	John
6	high	low	high	John
7	high	high	low	John
8	high	high	high	John

run order

aliased

	A	B	C	AB
2	low−	low−	high+	high+
3	low−	high+	low−	low−
5	high+	low−	low−	low−
8	high+	high+	high+	high+

= A×B

Q: what if operators have an effect on response?

- Q: Is aliasing/confounding always a bad thing?
  - pros & cons

Randomization

But we don't know

Q: what if operators have an effect on response?

randomize run order

run order

confounded

slightly confounded

unknown Z

	A	B	C	operator		Z	A	B	C	operator
1	low	low	low	Peter	EU_{11}	5	high	low	low	Peter
2	low	low	high	Peter	EU_{12}	2	low	low	high	Peter
3	low	high	low	Peter	EU_{13}	8	high	high	high	Peter
4	low	high	high	Peter	EU_{14}	4	low	high	high	Peter
5	high	low	low	John	EU_{21}	3	low	high	low	John
6	high	low	high	John	EU_{22}	1	low	low	low	John
7	high	high	low	John	EU_{23}	6	high	low	high	John
8	high	high	high	John	EU_{24}	7	high	high	low	John

EUs are heterogeneous, but unknown to the exp'ters. Exp'ters still assume EUs are homogeneous.

Randomization provides protection against *extraneous factors* that are unknown to the experimenter, but may impact the response

like firewall, immune system

- what should be randomized?
  - ➢ allocation of exp'tal materials to treatments; the order of applying treatments; the order of measuring responses; ...