Recap

Statistical model: What is ‘means model’ and what is ‘effects model’ for SINGLE factor EXP?

Single Factor
(multiple settings/levels)

Process/Product

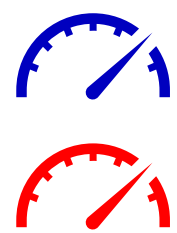
Response

Recap

Statistical model: What is ‘means model’ and what is ‘effects model’ for TWO factor EXP?

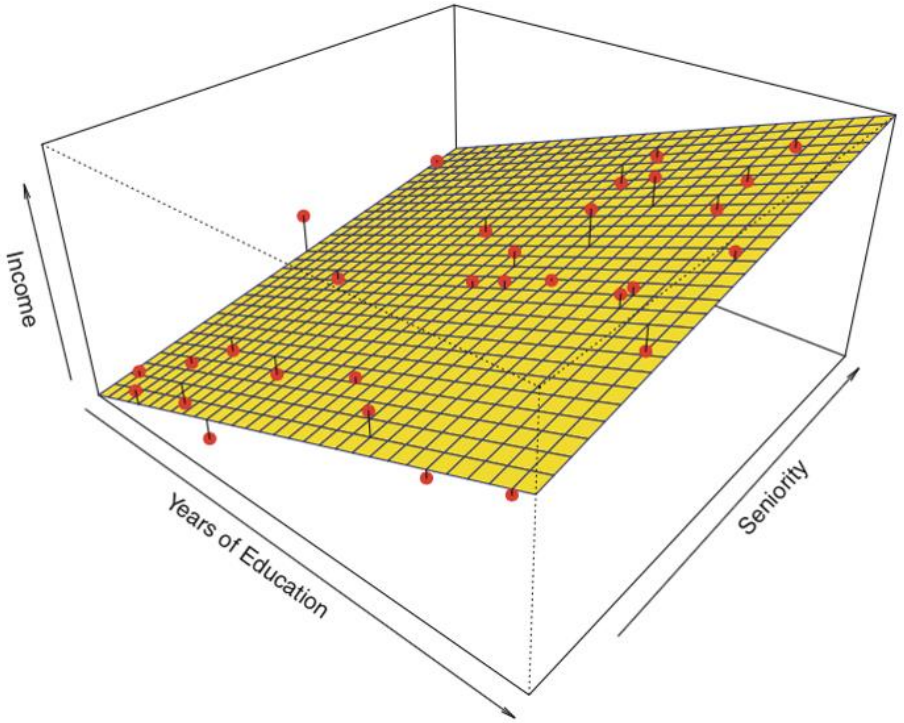
Factor 1
(multiple settings/levels)

Factor 2
(multiple settings/levels)



Process/Product

Response



Example: Two Factor ANOVA

Question is WHICH factor(s) have significant effect on the response?

Factor 1? Factor 2? Both? None?

		Process (Chemicals) (Factor 1: 4 levels)				
		1	2	3	4	Average
Batch (Factor 2: 5 levels)	A	89	88	97	94	92
	B	84	77	92	79	83
	C	81	87	87	85	85
	D	87	92	89	84	88
	E	79	81	80	88	82
	Average	84	85	89	86	

Two-Factor ANOVA: Model

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (\bar{y}_j - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{\bar{y}})$$

- y_{ij}: Obs value
- $\bar{\bar{y}}$: Grand mean
- $(\bar{y}_i - \bar{\bar{y}})$: Effect of first factor
- $(\bar{y}_j - \bar{\bar{y}})$: Effect of second factor
- $(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{\bar{y}})$: Residual Error or Intrinsic Variation

$$SS_T = SS_{mean} + SS_1 + SS_2 + SS_{error}$$

$N = ab$

DOF	N	1	$a-1$	$b-1$	$N-a-b+1 = (a-1)(b-1)$

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

- y_{ij}: Obs.
- τ_i: Factor 1 effect
- β_j: Factor 2 effect
- ε_{ij}: error

"No interaction effect assumed"

TWO Factor ANOVA

How do we statistically determine WHICH effect is significant?

When Factor 1 has ‘a’ levels and Factor 2 has ‘b’ levels and all possible combinations (N = ab) are tested.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor 1	$SS_{\text{Factor 1}}$	$a-1$	$\frac{SS_{\text{Factor 1}}}{a-1}$	$\frac{MS_{\text{Factor 1}}}{MS_E}$
Factor 2	$SS_{\text{Factor 2}}$	$b-1$	$\frac{SS_{\text{Factor 2}}}{b-1}$	$\frac{MS_{Factor\ 2}}{MS_E}$
Error	SS_E	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$N-1$		

Randomized Complete Block Design (RCBD)

- In any experiment, variability arising from a nuisance factor can affect the results. *Generally, we define a nuisance factor as a design factor that probably has an effect on the response, but we are not interested in that effect.*
- Sometimes a nuisance factor is unknown and uncontrolled; that is, we don't know that the factor exists, and it may even be changing levels while we are conducting the experiment. Randomization is the design technique used to guard against such a "lurking" nuisance factor.
- In other cases, the nuisance factor is known but uncontrollable. If we can at least observe the value that the nuisance factor takes on at each run of the experiment, we can compensate for it in the statistical analysis by using the ANOVA
- When the **nuisance source of variability is known and controllable**, a design technique called **blocking** can be used to systematically eliminate its effect on the statistical comparisons among treatments.

RCBD: Example

Does Virat Kohli's performance in IPL change based on which team he plays against?

Single Factor: Opposition team
(3 levels: MI vs CSK vs RR)

What kind of (exp) data do we need?

Opposition Team	Match 1	Match 2
MI		
CSK		
RR		

RCBD: Example

Do **Single Factor ANOVA** to find out

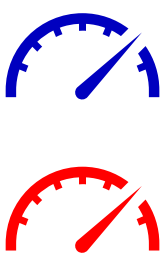
Opposition Team	Match 1	Match 2
MI		
CSK		
RR		

RCBD: Example

.. But wait …a sports analyst claims Virat's performance also depends on whether is he is playing at home (Bengaluru) or away! **Did you BLOCK the effect of that?**

Main Factor: Opposition team
(3 levels: MI vs CSK vs RR)

Block Factor: Location
(2 levels: Home vs Away)



Runs Scored

What kind of (exp) data do we need?

Opposition Team	Match 1	Match 2
MI		
CSK		
RR		