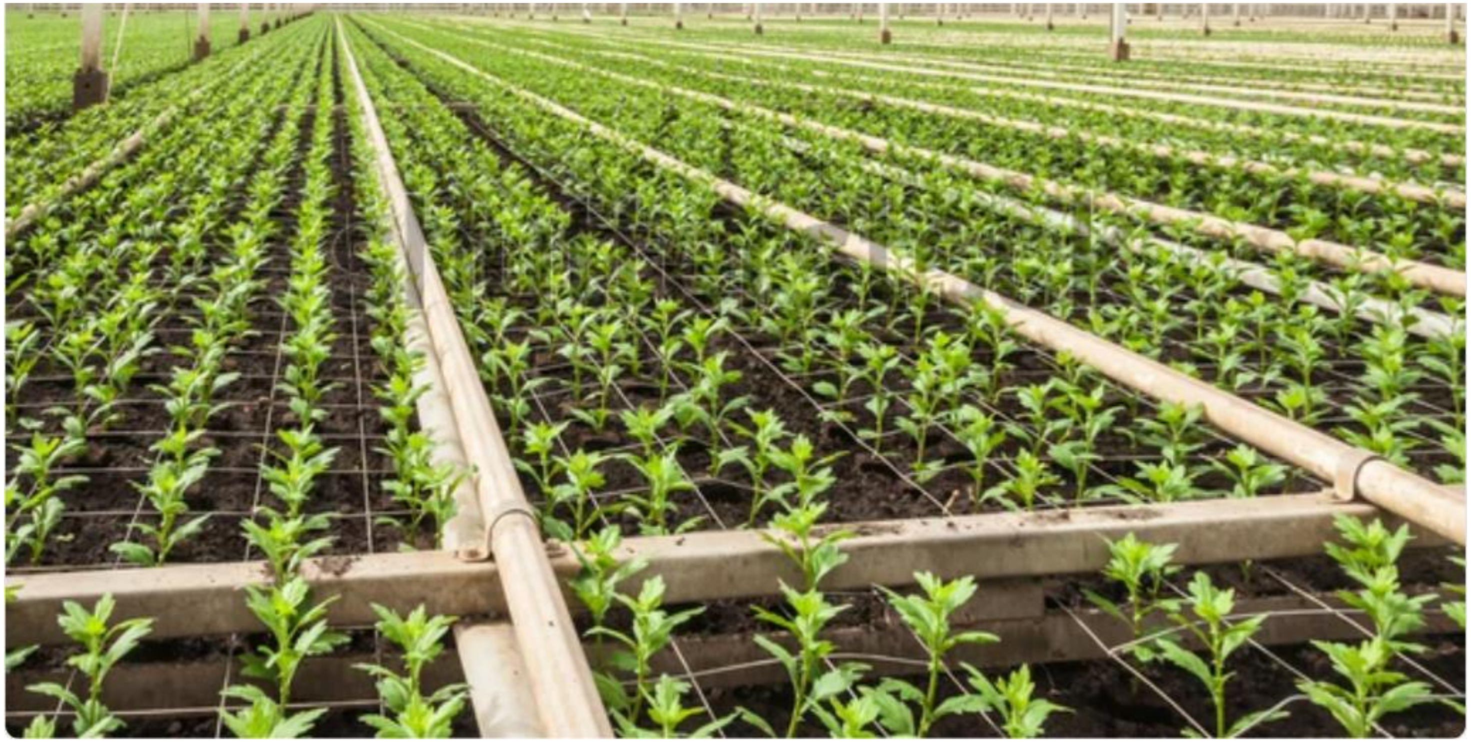
DOE 2.1
TWO FACTOR CROSS CLASSIFICATION DESIGNS



Division of Interdisciplinary Studies
Semester IV

DESIGNS WITH REPLICATION & INTERACTION

ACTIVITY 2 : Let's revisit the Case Study 1

CASE STUDY 3 : battery lifetime with two factors

Continuing with the study of battery lifetime, the dependent variable, we now use two independent variables (factors): battery brand and the device in which the battery is used.

Suppose we study three devices and four brands, and decide to run each combination of levels of factors for two batteries; that is, each combination of levels of factors, or "treatment combination," is replicated twice. (In the real world, battery testing often involves 16 batteries, sometimes more, for each combination of device and brand).

Table 5 : Battery lifetime (in hours) by brand and device

	Brand								
Factor 1	Device	1		2		3		4	
	1	17.9	18.1	17.8	17.8	18.1	18.2	17.8	17.9
	2	18.2	18.0	18.0	18.3	18.4	18.1	18.1	18.5
	3	18.0	17.8	17.8	18.0	18.1	18.3	18.1	17.9

Factor 2

Replicates

Identify variables, levels and replicates

THE STATISTICAL MODEL

$$\boldsymbol{Y_{ijk} = \mu + \rho_i + \tau_j + I_{ij} + \varepsilon_{ijk}} \qquad (9)$$

where

$\boldsymbol{i = 1, 2, 3, \ldots, R}$; that is, $\boldsymbol{i}$ indexes rows

$\boldsymbol{j = 1, 2, 3, \ldots, C}$; $\boldsymbol{j}$ indexes columns

$\boldsymbol{k = 1, 2, 3, \ldots, n}$; $\boldsymbol{k}$ indexes replication per cell

$\boldsymbol{Y_{ijk}}$ = the data point that corresponds with the $\boldsymbol{k^{th}}$ replicate in the cell at the intersection of the $\boldsymbol{i^{th}}$ row and $\boldsymbol{j^{th}}$ column;

$\boldsymbol{\mu}$ = overall average (mean)

$\boldsymbol{\rho_i}$ = the difference between the $\boldsymbol{i^{th}}$ row mean and the grand mean

$\boldsymbol{\tau_j}$ = the difference between the $\boldsymbol{j^{th}}$ column mean and the grand mean

$\boldsymbol{I_{ij}}$ = a measure of the interaction associated with the $\boldsymbol{i^{th}}$ row and the $\boldsymbol{j^{th}}$ column

$\boldsymbol{\varepsilon_{ijk}}$ = noise or error associated with the particular $\boldsymbol{ijk^{th}}$ data value; the difference between the data value and the true mean of that cell

PARAMETER ESTIMATION OF THE MODEL

Parameter estimation is similar to the one factor model (1).
Replacing each parameter in the model by its corresponding estimate,

$$Y_{ijk} - \bar{Y}_{\dots} = (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\dots}) + (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\dots}) + (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\dots}) + (Y_{ijk} - \bar{Y}_{ij\cdot}) \qquad (10)$$

By squaring and summing up it can be shown that,

$$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\dots})^2 = nC \sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\dots})^2 + nR \sum_j (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\dots})^2 + n \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\dots})^2 + \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2 \qquad (11)$$

$$TSS \quad = \quad SSB_r \quad + SSB_c \quad + SSI_{r,c} \quad + \quad SSW$$

and allocate the degrees of freedom as follows:

$$nRC - 1 = (R - 1) + (C - 1) + (R - 1)(C - 1) + RC(n - 1)$$

Identify variables, levels and replicates from Table 5.

INTERACTION

- Suppose we have two factors, A and B, each at two levels, high (H) and low (L). suppose that the cell means are as follows:

	B_L	B_H
A_L	5	8
A_H	10	?

- Increase in factor A level keeping factor B level as a constant : 5 →10 by 5
- Increase in factor B level keeping factor A level as a constant : 5 →8 by 3
- Increase in the yield when both factor levels change ? How many possibilities are there?

If $(A_H, B_H) = 13$, there is no interaction.
If $(A_H, B_H) > 13$, there is positive interaction.
If $(A_H, B_H) < 13$, there is negative interaction.

Interaction is the degree of difference from the sum of the separate effects

CALCULATING SUMS OF SQUARES

Table 6 : Battery lifetime (in hours) by brand and device with mean values

	Brand								$\bar{Y}_{i\cdot\cdot}$
Device	1		2		3		4		
1	17.9	18.1	17.8	17.8	18.1	18.2	17.8	17.9	
2	18.2	18.0	18.0	18.3	18.4	18.1	18.1	18.5	
3	18.0	17.8	17.8	18.0	18.1	18.3	18.1	17.9	
$\bar{Y}_{\cdot j\cdot}$									

Cell means

Row means

Column means

Grand mean

CALCULATING SUMS OF SQUARES

Table 6 : Battery lifetime (in hours) by brand and device with mean values

	Brand								$\bar{Y}_{i\cdot\cdot}$
Device	1		2		3		4		
1	17.9	18.1	17.8	17.8	18.1	18.2	17.8	17.9	
Cell means	18.0		17.8		18.15		17.88		**17.95**
2	18.2	18.0	18.0	18.3	18.4	18.1	18.1	18.5	
	18.1		18.15		18.25		18.3		**18.20**
3	18.0	17.8	17.8	18.0	18.1	18.3	18.1	17.9	
	17.9		17.9		18.2		18.0		**18.00**
$\bar{Y}_{\cdot j\cdot}$	**18.00**		**17.95**		**18.20**		**18.05**		**18.05**

Row means

Column means

Grand mean

$$SSB_r = (2)(4)\left[(17.95 - 18.05)^2 + (18.20 - 18.05)^2 + (18.00 - 18.05)^2\right]$$
$$= 8[.01 + .0225 + .0025]$$
$$= .28$$

$$SSB_c = (2)(3)\left[(18.00 - 18.05)^2 + (17.95 - 18.05)^2 + (18.20 - 18.05)^2 + (18.05 - 18.05)^2\right]$$
$$= 6[.0025 + .001 + .0225 + 0]$$
$$= .21$$

$$SSI_{r,c} = (2)\left[(18.00 - 17.95 - 18.00 + 18.05)^2 + (17.80 - 17.95 - 17.95 + 18.05)^2 + \ldots + (18.00 - 18.00 - 18.05 + 18.05)^2 +\right]$$
$$= 2[.055]$$
$$= .11$$

$$SSW = \left[(17.90 - 18.00)^2 + (18.10 - 18.00)^2 + (17.80 - 17.80)^2 + \cdots + (17.90 - 18.00)^2\right]$$
$$= .30$$

and, as a consequence,

$$TSS = .28 + .21 + .11 + .30$$
$$= .90$$

If doing the calculations by hand, it may be simpler to independently compute SSB_c, SSB_r, SSW, and TSS, and then derive $SSI_{r,c}$ by subtraction.

Table 7 : ANOVA Table for two-factor study of battery life

Source of variability	SSQ	df	$Mean\ Square\ (MS)$	$F\ calc$
Rows (due to device)	SSB_r	$R-1$	$MSB_r = SSB_r/(R-1)$	MSB_r/MSW
Columns (due to brand)	SSB_c	$C-1$	$MSB_c = SSB_c/(C-1)$	$MSB_c/\ MSW$
Interaction	$SSI_{r,c}$	$(R-1)(C-1)$	$MSI_{r,c} = SSI_{r,c}/(R-1)(C-1)$	$MSI_{r,c}/MSW$
Error	SSW	$RC(n-1)$	$MSW = SSW/RC(n-1)$	
Total	TSS	$nRC-1$		

Exercise 5:

Test the hypotheses for

1. Row factor
2. Column factor
3. Interaction effect

Ref: Experimental Design with Applications in Management, Engineering and Sciences (Berger et.al,2018)

Table 7 : ANOVA Table for two-factor study of battery life

Source of variability	*SSQ*	*df*	*Mean Square (MS)*	*F calc*
Rows (due to device)	$SSB_r = 0.28$	$R - 1 = 2$	$MSB_r = \frac{SSB_r}{R - 1} = 0.14$	$\frac{MSB_r}{MSW} = 5.6$
Columns (due to brand)	$SSB_c = 0.21$	$C - 1 = 3$	$MSB_c = \frac{SSB_c}{C - 1} = 0.07$	$\frac{MSB_c}{MSW} = 2.8$
Interaction	$SSI_{r,c} = 0.11$	$(R - 1)(C - 1) = 6$	$MSI_{r,c} = \frac{SSI_{r,c}}{(R-1)(C-1)} = 0.0183$	$\frac{MSI_{r,c}}{MSW} = 0.73$
Error	$SSW = 0.30$	$RC(n - 1) = 12$	$MSW = \frac{SSW}{RC(n - 1)} = 0.025$	
Total	$TSS = 0.90$	$nRC - 1 = 23$		

Hypotheses for row factor

H_0: All row means are equal
H_1: Not all row means are equal

Hypotheses for column factor

H_0: All column means are equal
H_1: Not all column means are equal

Hypotheses for interaction effect

H_0: There is no interaction between row and column factors
H_1: There is interaction between row and column factors

TWO FACTOR DESIGNS IN EXCEL

Exercise 6:

1. Conduct two factor ANOVA with replications in EXCEL for battery lifetime data
2. Discuss the results

DOE 2.2
DESIGNS WITH NO REPLICATION & NO INTERACTION

THE STATISTICAL MODEL

Recall the statistical model for two factor design with replication and interaction

$$\boldsymbol{Y_{ijk} = \mu + \rho_i + \tau_j + I_{ij} + \varepsilon_{ijk}} \qquad (9)$$

If we consider the design with no replication reduces as follows and the interaction term becomes error term.

$$\boldsymbol{Y_{ij} = \mu + \rho_i + \tau_j + \varepsilon_{ij}} \qquad (12)$$

$\boldsymbol{i = 1, 2, 3, \ldots, R}$; that is, $\boldsymbol{i}$ indexes rows

$\boldsymbol{j = 1, 2, 3, \ldots, C}$; $\boldsymbol{j}$ indexes columns

$\boldsymbol{Y_{ij}}$ = the data point that corresponds with the $\boldsymbol{i^{th}}$ row and $\boldsymbol{j^{th}}$ column

$\boldsymbol{\mu}$ = overall average (mean)

$\boldsymbol{\rho_i}$ = the difference between the $\boldsymbol{i^{th}}$ row mean and the grand mean

$\boldsymbol{\tau_j}$ = the difference between the $\boldsymbol{j^{th}}$ column mean and the grand mean

$\boldsymbol{\varepsilon_{ij}}$ = noise or error associated with the particular $\boldsymbol{ij^{th}}$ data value

PARAMETER ESTIMATION OF THE MODEL

- By following the procedure as in previous designs, sums of squares will be calculated as;

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = C \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 + R \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 \qquad (13)$$

$$TSS = \quad SSB_r \quad + \quad SSB_c \quad + \quad SSW$$

with the degrees of freedom as follows:

$$RC - 1 = (R - 1) + (C - 1) + (R - 1)(C - 1)$$

- $TSS - Total\ sum\ of\ squares = \text{sum of squares (Observation} - \text{Grand Mean)}$
- $SSB_r - Sum\ of\ squares\ between\ rows = C \times sum\ of\ squares(Row\ mean - Grand\ Mean)$
- $SSB_c - Sum\ of\ squares\ between\ columns = R \times sum\ of\ squares(Column\ Mean - Grand\ Mean)$
- $SSW - Sum\ of\ squares\ within\ columns = sum\ of\ squares(Observation - RM - CM + GM)$

An Un-replicated Numerical Example

- Data array of un-replicated two-factor experiment are shown in Table 8. Based on historical evidence it is assumed that there is no interaction between the two factors.

Table 8 : Data array for numerical example

Level of factor B	Level of factor A			
	1	**2**	**3**	$\bar{Y}_{i\cdot}$
1	7	3	4	
2	10	6	8	
3	6	2	5	
4	9	5	7	
$\bar{Y}_{\cdot j}$				$\bar{Y}_{\cdot\cdot} =$

Dependent variable :

Independent variables or factors:

No. of levels of factor A :

No. of levels of factor B :

Total no. of data points :

An Un-replicated Numerical Example

- Data array of un-replicated two-factor experiment are shown in Table 8. Based on historical evidence it is assumed that there is no interaction between the two factors.

Table 8 : Data array for numerical example

Level of factor B	Level of factor A			
	1	**2**	**3**	$\bar{Y}_{i.}$
1	7	3	4	4.7
2	10	6	8	8
3	6	2	5	4.3
4	9	5	7	7
$\bar{Y}_{.j}$	8	4	6	$\bar{Y}_{..} = 6$

Dependent variable : Y

Independent variables or factors: factor A and factor B

No. of levels of factor A : $C = 3$

No. of levels of factor B : $R = 4$

Total no. of data points : $R \times C = 12$

- $SSB_r = C\sum_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = 3\{(4.7-6)^2 + \cdots \ldots \ldots \ldots \ldots (7-6)^2\} = \mathbf{28.67}$
- $SSB_c = R\sum_j(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 = 4\{(8-6)^2 + (4-6)^2 + (6-6)^2\} = \mathbf{32.00}$
- $TSS = \sum_i\sum_j(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = (7-6)^2 + (10-6)^2 + \cdots + (7-6)^2 = \mathbf{62.00}$
- $SSW = TSS - (SSB_r + SSB_c) = 62.00 - (28.67 + 32.00) = \mathbf{1.33}$

Table 9 : ANOVA Table for un-replicated numerical example

Source of variability	SSQ	*df*	MS	F_{calc}
Row	28.67	3	9.55	43
Column	32.00	2	16.00	72
Error	1.33	6	.22	
Total	62.00	11		

Ref: Experimental Design with Applications in Management, Engineering and Sciences (Berger et.al, 2018)

Exercise 7 : Test the hypotheses for row factor and column factor at α = 0.05 and α = 0.01 levels.

Hypotheses for row factor

H_0: All row means are equal
H_1: Not all row means are equal

When α = 0.05 table F value, $F_{0.05,3,6} = 4.76 < \ F_{calc} = 43.$
When α = 0.01 table F value, $F_{0.01,3,6} = 9.78 < \ F_{calc} = 43.$
Reject $\boldsymbol{H_0}$ and row factor is highly significant at both levels.

Hypotheses for column factor

H_0: All column means are equal
H_1: Not all column means are equal

It is assumed that there is no interaction between row and column factors.
If our assumption of no interaction is incorrect, conclusions are not valid.

TWO FACTOR DESIGNS NO REPLICATION IN EXCEL

Exercise 8:

1. Conduct two factor ANOVA with no replication in EXCEL for un-replicated numerical data
2. Discuss the results

Tips & Take aways

THANK YOU