

Lecture 4: Two-Factor Studies with Equal Sample Sizes

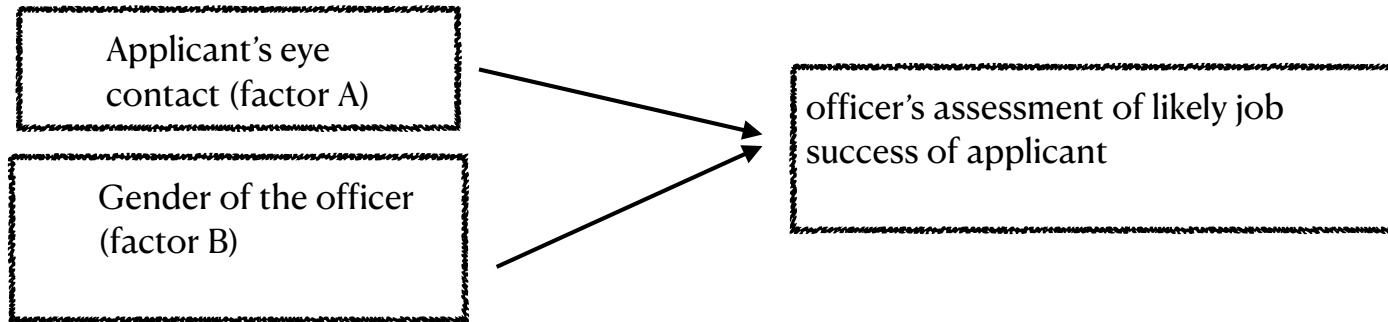
STA 106: Analysis of Variance

Suggested reading: ALSM Chapter 19

Example

(The Eye Contact Study)

The objective of the study:



The study setup:

10 male and 10 female officers were chosen

Half of the officers in each gender group were chosen at random to receive a photograph of the applicant in which the applicant made eye contact with the camera lens; the other half received a version in which there was no eye contact

Officers were asked to give a rating on a scale 0 (total failure) to 20 (outstanding success) on likely job success

		Factor B (gender of officer)	
		<i>j</i> = 1 Male	<i>j</i> = 2 Female
<i>i</i> = 1	Present	11	15
		7	12
	
		10	16
<i>i</i> = 2	Absent	12	14
		16	17
	
		14	18

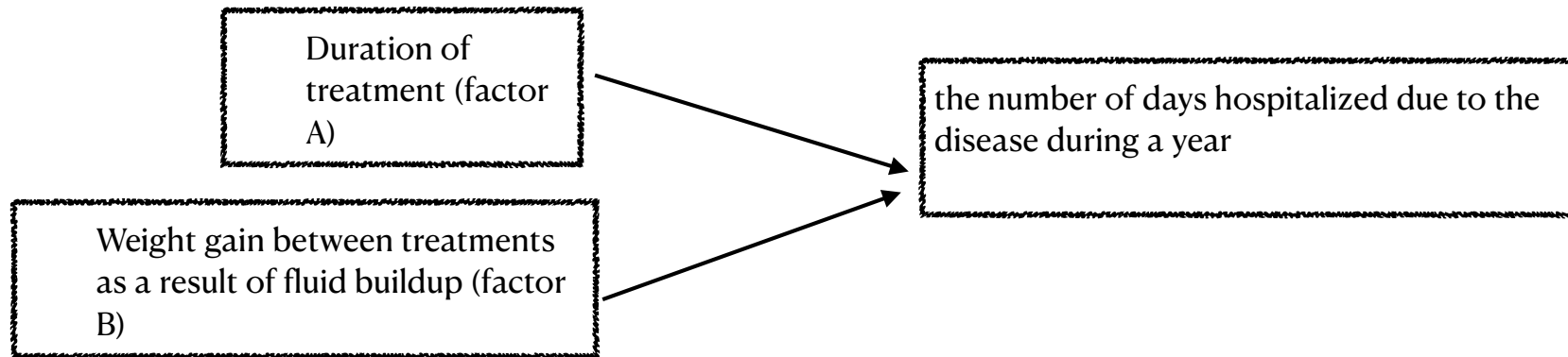
☒ Randomized Block Design

Example

(The Kidney Failure Hospitalization Study)

The objective of the study:

What is the appropriate “dose” for effective dialysis treatment ?



The study setup:

A random sample of 10 patients per treatment group at dialysis facility were chosen

Factor A (duration)		Factor B (weight gain)					
		$j = 1$ Mild		$j = 2$ Moderate		$j = 3$ Substantial	
$i = 1$	Short	0	2	2	4	15	16
		2	0	4	3	10	7
	
$i = 2$	Long	0	8	15	20	25	27
		0	2	5	1	10	15
		1	7	3	3	8	4
		
		4	3	1	9	7	1

- ✓ Observational study
- Without randomly assigning factor levels

Two-Factor Studies with Equal Sample Sizes



Two-Way ANOVA Model










Analysis of Variance

Tests for Interaction effects and
main effects

Analysis of Factor A and B Main
Effects (When Factors Do Not
Interact)

Analysis of Treatment Means
when Interactions Are Present

Two-Way ANOVA Model for Two-Factor Studies

		Factor B			
		$j=1$	$j=2$...	$j=b$
Factor A	$i=1$				
	$i=2$				
	\vdots				
	$i=a$				

Factor A is studied at a levels

Factor B is studied at b levels

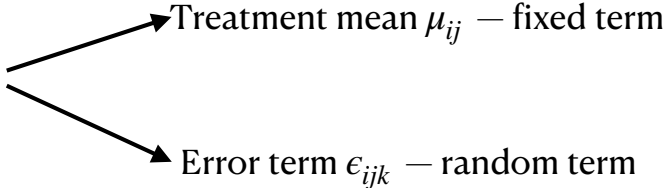
All treatment sample sizes are equal with $n > 1$

Total sample size $n_T = abn$

k th observation ($k = 1 \dots n$) for the treatment ($A = i, B = j$) is Y_{ijk}

Two-Way ANOVA Model for Two-Factor Studies

Assume: observed value of response variable is the sum of two components



```
graph LR; A[Assume: observed value of response variable is the sum of two components] --> B[Treatment mean  $\mu_{ij}$  — fixed term]; A --> C[Error term  $\epsilon_{ijk}$  — random term]
```

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

- ϵ_{ijk} error term $\epsilon_{ijk} \sim N(0, \sigma^2)$

- μ_{ij} Treatment means

$$E(Y_{ijk}) = E(\mu_{ij} + \epsilon_{ijk}) = \mu_{ij} + E(\epsilon_{ijk}) = \mu_{ij}$$

However, this way of parameterization in terms of treatment means is not enough, what's the complication for two-factor studies?

Two-Way ANOVA Model for Two-Factor Studies

We ultimately are interested in **simultaneous investigating the joint effects of factor A and factor B**

That is, we want to parse out

What is the effect due to factor A only?

What is the effect due to factor B only?

Whether there is some extra effects unique to certain combinations of factor A and factor B ?

It's not enough to just look at the treatment means μ_{ij} , as it tells only the difference between treatments, we care about parsing out difference due to factor A and factor B










We want to make statements such as:

“ Factor A has beneficial / detrimental effects in general”

“Factor B has beneficial / detrimental effects in general”

“ Factor A has beneficial / detrimental effects only when factor B is at certain levels”

Two-Way ANOVA Model for Two-Factor Studies

		Factor B			
		$j=1$	$j=2$...	$j=b$
Factor A	$i=1$	μ_{11} 	μ_{12} 		μ_{1b} 
	$i=2$	μ_{21} 	μ_{22} 		μ_{2b} 
	\vdots				
	$i=a$	μ_{a1} 	μ_{a2} 		μ_{ab} 
	$\mu_{.j}$	$\mu_{.1}$	$\mu_{.2}$		$\mu_{.b}$

Factor level means:

$$\mu_{.j} = \frac{\sum_{i=1}^a \mu_{ij}}{a} \quad \mu_{i.} = \frac{\sum_{j=1}^b \mu_{ij}}{b}$$

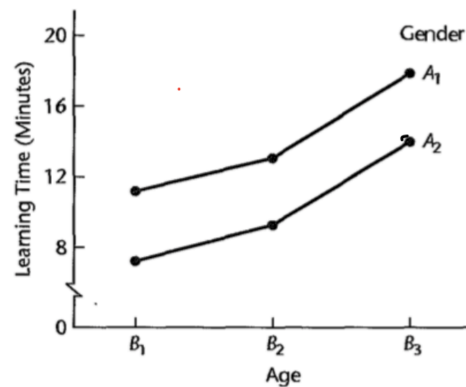
Overall mean:

$$\mu_{..} = \frac{\sum_{i=1}^a \sum_{j=1}^b \mu_{ij}}{ab} = \frac{\sum_{i=1}^a \mu_{i.}}{a} = \frac{\sum_{j=1}^b \mu_{.j}}{b}$$

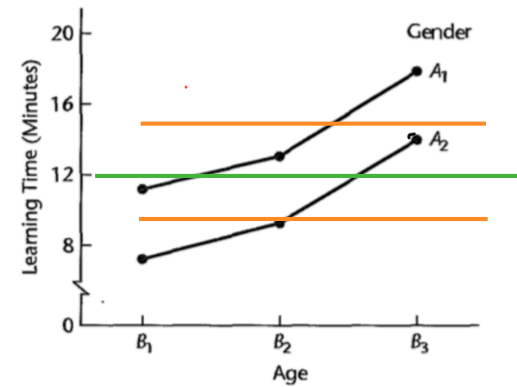
Case I: Additive Factor Effects

Suppose we know the true underlying model where μ_{ij} known

(a) Mean Learning Times (in minutes)				
		Factor B—Age		
Factor A—Gender		$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old
$i = 1$ Male		11 (μ_{11})	13 (μ_{12})	18 (μ_{13})
$i = 2$ Female		7 (μ_{21})	9 (μ_{22})	14 (μ_{23})
Column average		9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)
	Row Average			
		14 ($\mu_{1\cdot}$)	10 ($\mu_{2\cdot}$)	12 ($\mu_{\cdot\cdot}$)



Graphical representation: Treatment Means Plot or Interaction Plot



Graphical representation: Treatment Means Plot or Interaction Plot

What is the effect of factor A?

- how does the mean response change with different levels of factor A, regardless of other factors (here factor B)?

Define: Main effects

Main effect of factor A at ith level $\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}$

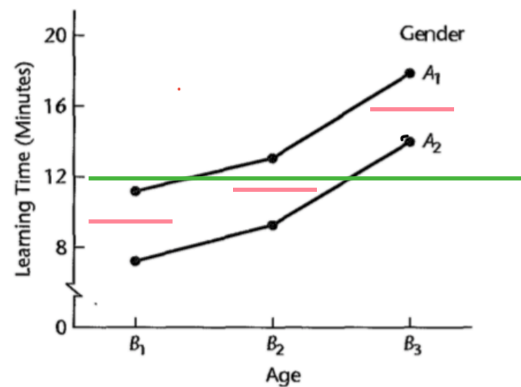
How much ith level of factor A shift the factor level mean away from overall mean

$$\sum_{i=1}^a \alpha_i = 0 \quad \text{Sum of main effects is 0}$$

Case I: Additive Factor Effects

Suppose we know the true underlying model where μ_{ij} known

(a) Mean Learning Times (in minutes)				
		Factor B—Age		
Factor A—Gender		$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old
$i = 1$ Male		11 (μ_{11})	13 (μ_{12})	18 (μ_{13})
$i = 2$ Female		7 (μ_{21})	9 (μ_{22})	14 (μ_{23})
Column average		9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)
	Row Average			
		14 ($\mu_{1\cdot}$)	10 ($\mu_{2\cdot}$)	12 ($\mu_{\cdot\cdot}$)



Graphical representation: **Treatment Means Plot or Interaction Plot**

What is the effect of factor B?

- how does the mean response change with different levels of factor B, regardless of other factors (here factor B)?

Define: **Main effects**

Main effect of factor B at j th level $\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$

How much j th level of factor B shift the factor level mean away from overall mean

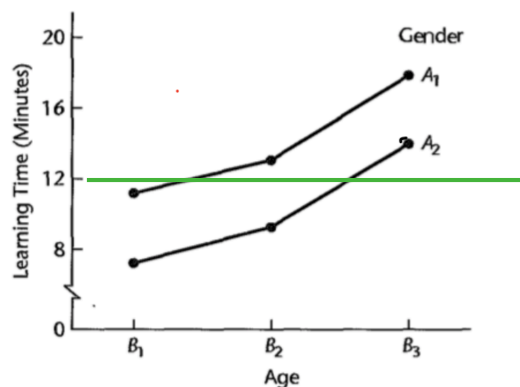
$$\sum_{j=1}^b \beta_j = 0 \quad \text{Sum of main effects is 0}$$

Case I: Additive Factor Effects

Suppose we know the true underlying model where μ_{ij} known

(a) Mean Learning Times (in minutes)				
Factor A—Gender	Factor B—Age			Row Average
	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old	
$i = 1$ Male	11 (μ_{11})	13 (μ_{12})	18 (μ_{13})	14 ($\mu_{1.}$)
$i = 2$ Female	7 (μ_{21})	9 (μ_{22})	14 (μ_{23})	10 ($\mu_{2.}$)
Column average	9 ($\mu_{.1}$)	11 ($\mu_{.2}$)	16 ($\mu_{.3}$)	12 ($\mu_{..}$)

(b) Main Gender Effects (in minutes)		(c) Main Age Effects (in minutes)	
$\alpha_1 = \mu_{1.} - \mu_{..} = 14 - 12 = 2$		$\beta_1 = \mu_{.1} - \mu_{..} = 9 - 12 = -3$	
$\alpha_2 = \mu_{2.} - \mu_{..} = 10 - 12 = -2$		$\beta_2 = \mu_{.2} - \mu_{..} = 11 - 12 = -1$	
		$\beta_3 = \mu_{.3} - \mu_{..} = 16 - 12 = 4$	



$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$$

[mean response] = [overall mean] + [Factor A main effect] + [Factor B main effect]

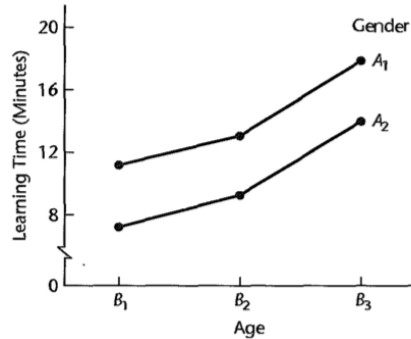
The joint effect of factor A and factor B, as measured by how much mean response deviate from overall mean is

Two main effects adding together : Additive Factor Effects or Factor Effects are Additive

$$\mu_{ij} - \mu_{..} = \alpha_i + \beta_j$$

Case I: Additive Factor Effects

The significance of additive factor effects:



Additive factor effects \Leftrightarrow Parallelism in interaction plot

For any factor B level j : $\mu_{1j} - \mu_{2j} = \alpha_1 - \alpha_2$
—> curves in the interaction plot are all parallel

- Ease of interpretation:

Effect of either factor (i.e. main effect of factor A or factor B) does not depend on the level of the other factor

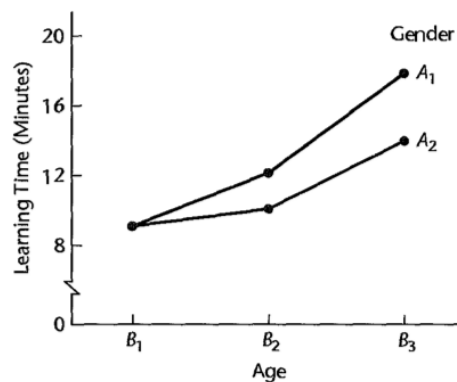
For example: No matter what is the level of factor A (gender), the effect of being young is reducing learning time by 3, no matter if we consider female and male separately or combined.

Case II: Interaction Effects

Case (II): Suppose we know the true underlying model where μ_{ij} known

(a) Mean Learning Times (in minutes)					
Factor A—Gender	Factor B—Age			Row Average	Main Gender Effect
	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old		
$i = 1$ Male	9 (μ_{11})	12 (μ_{12})	18 (μ_{13})	13 ($\mu_{1\cdot}$)	1 (α_1)
$i = 2$ Female	9 (μ_{21})	10 (μ_{22})	14 (μ_{23})	11 ($\mu_{2\cdot}$)	-1 (α_2)
Column average	9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)	12 ($\mu_{\cdot\cdot}$)	
Main age effect	-3 (β_1)	-1 (β_2)	4 (β_3)		

$$\mu_{ij} \neq \mu_{..} + \alpha_i + \beta_j$$



Factor A has no effect on Y when factor B=1, but has substantial effect when B=2 and B=3
 This differential influence of factor A, which depends on factor B, implies that factor A and factor B interact in their effect on Y

Case II: Interaction Effects

$$\mu_{ij} \neq \mu_{..} + \alpha_i + \beta_j$$

$$\Rightarrow \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j) \neq 0$$

There exists some effect unique to this combination of factor levels with $A = i, B = j$, due to the interaction of i th level of factor A and j th level of factor B

Define: **Interaction (or interaction effect)** of i th level of factor A with j th level of factor B

$$\begin{aligned}\gamma_{ij} &= \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j) \\ &= \mu_{ij} - (\mu_{..} + \mu_{i.} - \mu_{..} + \mu_{.j} - \mu_{..}) \\ &= \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}\end{aligned}$$

If two factors have no interaction = additive, then all interaction effects are 0

$$\gamma_{ij} = 0 \text{ for all } i, j$$

The interactions have two natural constraints by definition:

$$\sum_i \gamma_{ij} = 0 \quad j = 1, \dots, b$$

$$\sum_j \gamma_{ij} = 0 \quad i = 1, \dots, a$$

$$\rightarrow \text{therefore, } \sum_i \sum_j \gamma_{ij} = 0$$

$$\begin{aligned}\text{Proof: } \sum_i \gamma_{ij} &= \sum_{i=1}^a (\mu_{ij} - \mu_{..} - \alpha_i - \beta_j) \\ &= \sum_i \mu_{ij} - a\mu_{..} - \sum_i \alpha_i - a\beta_j = a\mu_{.j} - a\mu_{..} - a(\mu_{.j} - \mu_{..}) = 0\end{aligned}$$

Case II: Interaction Effects

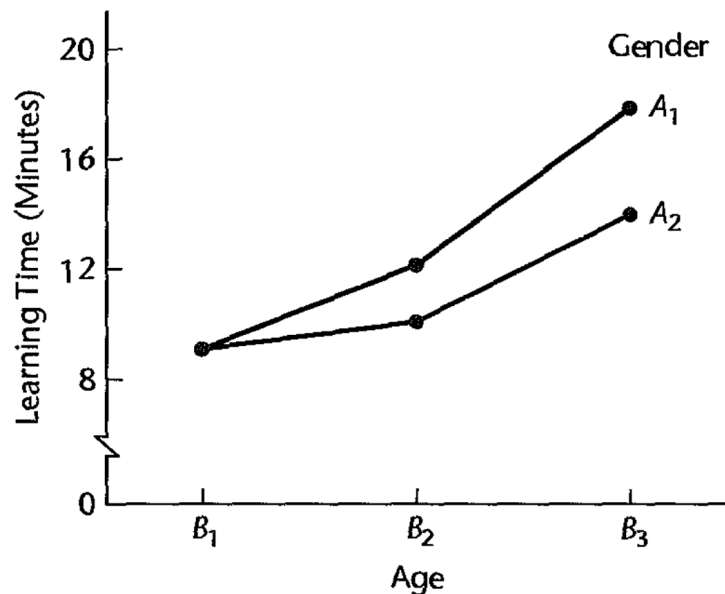
(a) Mean Learning Times (in minutes)					
Factor A—Gender	Factor B—Age			Row Average	Main Gender Effect
	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old		
$i = 1$ Male	9 (μ_{11})	12 (μ_{12})	18 (μ_{13})	13 ($\mu_{1\cdot}$)	1 (α_1)
$i = 2$ Female	9 (μ_{21})	10 (μ_{22})	14 (μ_{23})	11 ($\mu_{2\cdot}$)	-1 (α_2)
Column average	9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)	12 ($\mu_{\cdot\cdot}$)	
Main age effect	-3 (β_1)	-1 (β_2)	4 (β_3)		

(b) Interactions (in minutes)				
	$j = 1$	$j = 2$	$j = 3$	Row Average
$i = 1$	-1	0	1	0
$i = 2$	1	0	-1	0
Column average	0	0	0	0

-1: 1 unit smaller than what is expected by an additive effects model where only main effects are considered

Case II: Interaction Effects

The significance of Interaction effects, or the lack of additive factor effects:



Presence of Interaction effects \Leftrightarrow Nonparallel in interaction plot

For factor B level j:

$$\mu_{1j} - \mu_{2j} = (\mu_{..} + \alpha_1 + \beta_j + \gamma_{1j}) - (\mu_{..} + \alpha_2 + \beta_j + \gamma_{2j})$$

$$= \alpha_1 - \alpha_2 + \gamma_{1j} - \gamma_{2j}$$

Not same for different levels of j

\rightarrow curves in the interaction plot are not parallel

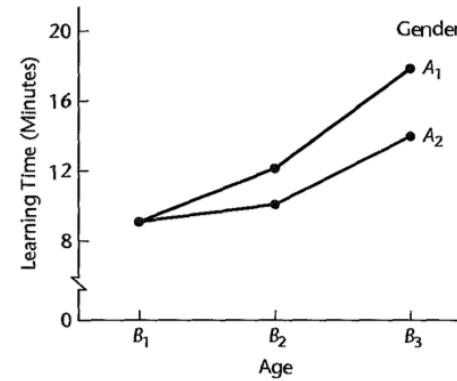
Distinction between Additive factor effects and Interaction effects:

By examining whether the curves in interaction plot are parallel, nonparallel occurs if and only if the interaction effects are present.

Case II: Interaction Effects

What happens when interaction effects are present?

(a) Mean Learning Times (in minutes)					
Factor A—Gender	Factor B—Age			Row Average	Main Gender Effect
	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old		
$i = 1$ Male	9 (μ_{11})	12 (μ_{12})	18 (μ_{13})	13 ($\mu_{1\cdot}$)	1 (α_1)
$i = 2$ Female	9 (μ_{21})	10 (μ_{22})	14 (μ_{23})	11 ($\mu_{2\cdot}$)	-1 (α_2)
Column average	9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)	12 ($\mu_{\cdot\cdot}$)	
Main age effect	-3 (β_1)	-1 (β_2)	4 (β_3)		



In additive factor effects case, the main effects is a meaningful measure of factor A (or B) effects.

interaction effects are not present, then it means factor A and B can be studied separately, since the effect of one factor does not depend on the other factor, i.e. they do not interact.



When interaction effects are present, the main effects alone no longer are meaningful measure of factor A (or B) effects, since they are intertwined.

more complicated interpretation

but if it is true, it means the factors A and B are in fact “interacting” in some way:

there are some combinations of A and B that may lead to better or worse outcome, while some combinations may have no effect at all.

If interaction effects indeed are present, then it's the unique strength of the two-factor studies

Case II: Interaction Effects

Important versus Unimportant Interactions:

Sometimes when two factors interact, the interaction effects are so small

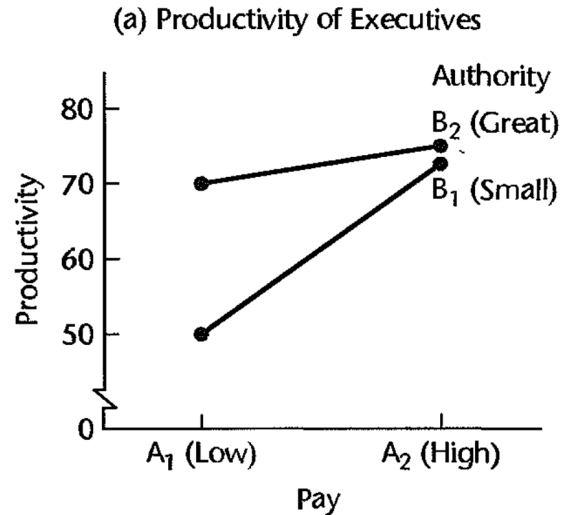
Subject area specialist or researcher decide that such a small interaction effect can be considered unimportant interactions.

Reduce to additive factor effects, where no interaction exists and each factor can be studied separately based on factor level means $\mu_{i.}$ and $\mu_{.j}$

Case II: Interaction Effects

Interaction Patterns

(a) Productivity of Executives		
Factor A—Pay	Factor B—Authority	
	Small	Great
Low	50	72
High	74	75



For low-paid executives with small authority, raising the pay only or increasing the authority of low-paid executives alone, leads to substantial increased productivity.

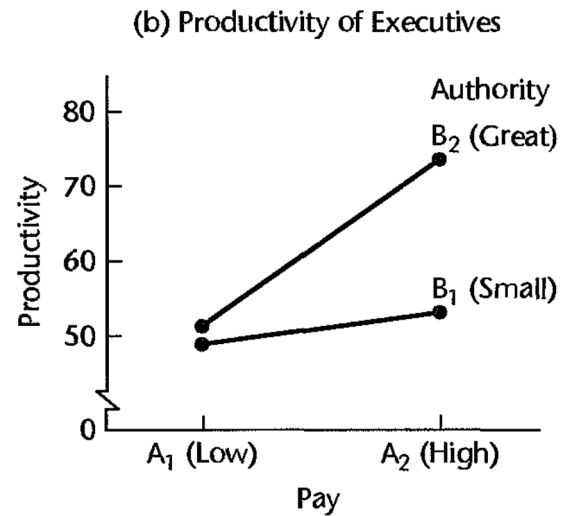
However, combining both high pay and great authority has a smaller beneficial effect than either one alone.

(Combined forces cancel each one out)

Case II: Interaction Effects

Interaction Patterns

(b) Productivity of Executives		
Factor A—Pay	Factor B—Authority	
	Small	Great
Low	50	52
High	53	75



For low-paid executives with small authority, raising the pay or increasing the authority leads to almost negligible increased productivity.

Only when combining both high pay and great authority has a substantial beneficial effect.

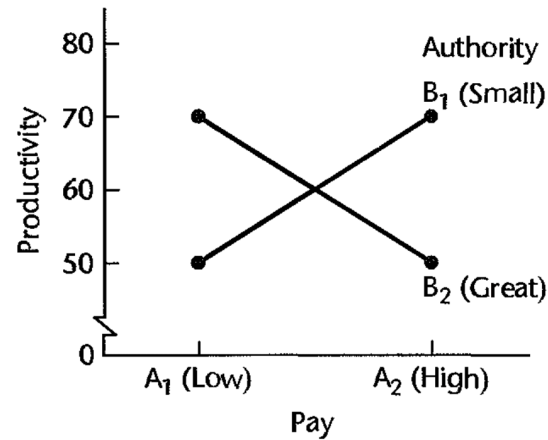
(Combined forces reinforce each other)

Case II: Interaction Effects

Interaction Patterns

(c) Productivity of Executives

(c) Productivity of Executives		
Factor A—Pay	Factor B—Authority	
	Small	Great
Low	50	72
High	72	50



The main effects for both factors are 0: misleading if we assume no interaction effects present or run separate one-factor studies

There are indeed factor effects of A and B, but they would not be seen by main effects, due to interactions in opposite directions that balance out.

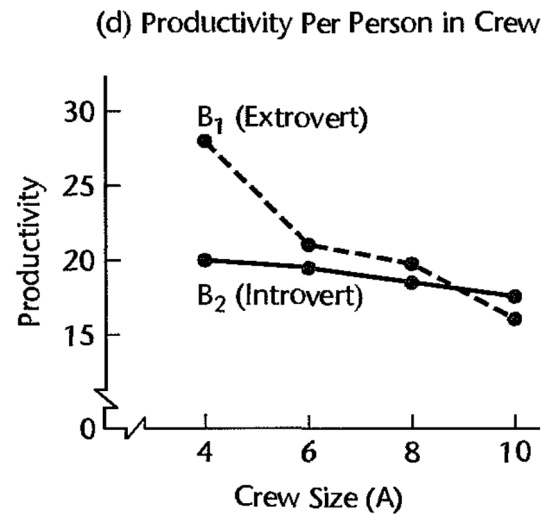
(Weird, rare, but possible situation)

Case II: Interaction Effects

Interaction Patterns

(d) Productivity per Person in Crew

Factor A—Crew Size	Factor B—Personality of Crew Chief	
	Extrovert	Introvert
4 persons	28	20
6 persons	22	20
8 persons	20	19
10 persons	17	18



size of crew and personality of crew chief interact in a complex way

Extrovert crew chief has a huge advantage over introvert, in a small crew.

However, this advantage become smaller and smaller when crew size become larger, crew size of 10 with an introvert crew chief can even lead to a slightly higher productivity.

Two-Way ANOVA Model for Two-Factor Studies

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \quad \text{Treatment means parameterization}$$

$$= \mu_{..} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad \text{Factor effects parameterization}$$

$$\bullet \mu_{..} = \frac{\sum_{i=1}^a \sum_{j=1}^b \mu_{ij}}{ab} : \text{overall mean}$$

$$\bullet \alpha_i = \mu_{i.} - \mu_{..} \text{ main effect of factor A at } i\text{th level}$$

$$\text{Subject to constraint } \sum \alpha_i = 0$$

$$\bullet \beta_j = \mu_{.j} - \mu_{..} \text{ main effect of factor B at } j\text{th level}$$

$$\text{Subject to constraint } \sum \beta_j = 0$$

$$\bullet \gamma_{ij} = \mu_{ij} - \alpha_i - \beta_j = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \text{ interaction effect of factor A at } i\text{th level with factor B at } j\text{th level}$$

$$\text{Subject to } a+b-1 \text{ constraints}$$

$$\sum_i \gamma_{ij} = 0 \quad j = 1, \dots, b$$


$$\sum_j \gamma_{ij} = 0 \quad i = 1, \dots, a$$

$$\bullet \varepsilon_{ijk} \text{ are independent } N(0, \sigma^2) \text{ for } i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, n$$

Fitting the Two-Way ANOVA Model

Least squares estimates for treatment means in treatment means parameterization:

$$Q = \sum_i \sum_j \sum_k (Y_{ijk} - \mu_{ij})^2$$


$$\hat{\mu}_{ij} = \bar{Y}_{ij} \text{ for } i = 1 \dots a, b = 1 \dots b$$



Least squares estimates for parameters in factor effects parameterization:

$$\hat{\mu}_{..} = \frac{\sum_i \sum_j \hat{\mu}_{ij}}{ab} = \frac{\sum_i \sum_j \bar{Y}_{ij}}{ab} = \bar{Y}_{...}$$

$$\hat{\alpha}_i = \hat{\mu}_{i.} - \hat{\mu}_{..} = \frac{\sum_j \bar{Y}_{ij}}{b} - \bar{Y}_{...} = \bar{Y}_{i..} - \bar{Y}_{...}$$

$$\hat{\beta}_j = \hat{\mu}_{.j} - \hat{\mu}_{..} = \frac{\sum_i \bar{Y}_{ij}}{a} - \bar{Y}_{...} = \bar{Y}_{.j.} - \bar{Y}_{...}$$

$$\hat{\gamma}_{ij} = \hat{\mu}_{ij} - \hat{\mu}_{i.} - \hat{\mu}_{.j} + \hat{\mu}_{..} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$$

Where:

$$\bar{Y}_{...} = \frac{\sum_i \sum_j \sum_k Y_{ijk}}{abn}$$

$$\bar{Y}_{i..} = \frac{\sum_j \sum_k Y_{ijk}}{bn}$$

$$\bar{Y}_{.j.} = \frac{\sum_i \sum_k Y_{ijk}}{an}$$

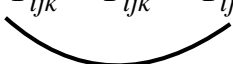
Fitting the Two-Way ANOVA Model

- fitted value for an observation Y_{ij}

ANOVA model's "best guess" or "best prediction" for Y_{ijk}

$$\hat{Y}_{ijk} = \hat{\mu}_{ij} = \bar{Y}_{ij}.$$

- residual e_{ij} corresponds to observation Y_{ij} is

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \bar{Y}_{ij}.$$


Difference between observed value and fitted value which is estimated factor level mean

Example

(Kidney failure hospitalization)

Kidney failure patients are commonly treated on dialysis machines that filter toxic substances from the blood.

The appropriate "dose" for effective treatment depends, among other things, on duration of treatment and weight gain between treatments as a result of fluid buildup.

To study the effects of these two factors on the number of days hospitalized (attributable to the disease) during a year, a random sample of 10 patients per group who had undergone treatment at a large dialysis facility was obtained.

Treatment duration (factor A) was categorized into two groups: short duration (average dialysis time for the year under four hours) and long duration (average dialysis time for the year equal to or greater than four hours).

Average weight gain between treatments (factor B) during the year was categorized into three groups: slight, moderate, and substantial.

Two-way ANOVA Model:

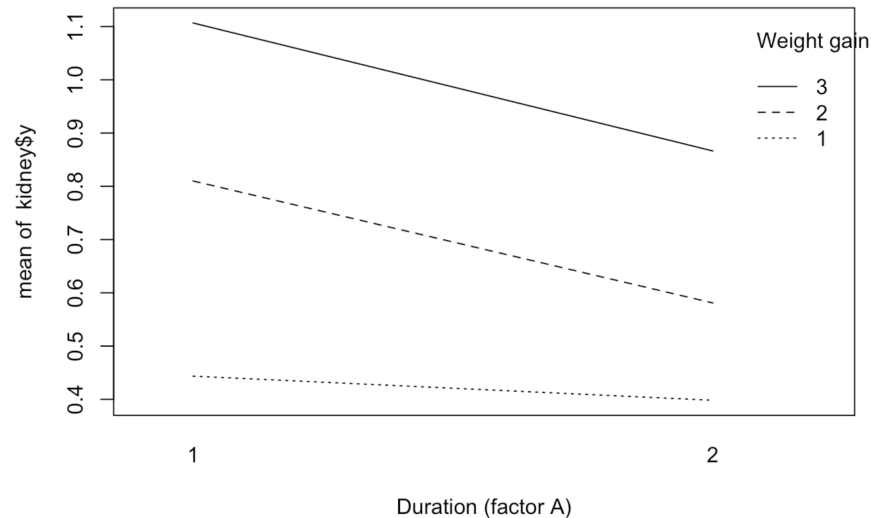
$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \quad \text{Treatment means parameterization}$$

$$= \mu_{..} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad \text{Factor effects parameterization}$$

$$i = 1, 2; \quad j = 1, 2, 3; \quad k = 1 \dots 10$$

Example

Draw an interaction plot based on the estimated treatment means. Comment on the plot in terms of interaction effects, factor A and B main effects.



The nonparallel lines suggests that the two factors interact, that is, the effect of factor B depends on the levels of factor A and vice versa.

Main effects:

Long treatment duration tends to reduce days of hospitalization, regardless of weight gain.

Larger weight gain tends to increase days of hospitalization, regardless of treatment duration.

The interaction pattern:

Larger weight gain prolongs the hospitalization days substantially for patients with short treatment duration, but this prolonged effect is smaller for patients with long treatment duration.

Longer treatment duration has negligible effect for patients with small weight gain, but has larger effect for patients with moderate or substantial weight gains.

Example

Least squares estimates for the parameters in the factor effects parameterization.

Overall mean:

$$\hat{\mu}_{..} = \frac{\sum_i \sum_j \hat{\mu}_{ij}}{ab} = \frac{\sum_i \sum_j \bar{Y}_{ij}}{ab} = \bar{Y}_{...} = 0.7$$

Factor A main effects:

$$\hat{\alpha}_1 = 0.09 \qquad \hat{\alpha}_2 = -0.09$$

Factor B main effects:

$$\hat{\beta}_1 = -0.28 \qquad \hat{\beta}_2 = -0.01 \qquad \hat{\beta}_3 = 0.29$$

Interaction effects $\hat{\gamma}_{ij}$:

	j=1 Mild	j=2 Moderate	j=3 Substantial
i=1 Short	-0.06	0.03	0.03
i=2 Long	0.06	-0.03	-0.03