ISE 315: Engineering Statistics

Lecture 25: Two-Factor Factorial Experiments

Instructor: Mansur M. Arief, PhD
Industrial and Systems Engineering, KFUPM

Office: 22-219 — Email: mansur.arief@kfupm.edu.sa

Based on Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed.

Lecture 25

Two-Factor Factorial Experiments (Chapter 9, §9.1–§9.7)

Lecture 25 Outline

- Why factorial? OFAT vs. factorial (§9.1)

Lecture 25 Outline

- Why factorial? OFAT vs. factorial (§9.1)
- Main effects, interactions, and interaction plots (§9.2)

Lecture 25 Outline

- Why factorial? OFAT vs. factorial (§9.1)
- Main effects, interactions, and interaction plots (§9.2)
- The two-factor effects model (§9.3, (9.1))

Lecture 25 Outline

- Why factorial? OFAT vs. factorial (§9.1)
- Main effects, interactions, and interaction plots (§9.2)
- The two-factor effects model (§9.3, (9.1))
- Partitioning variability and the ANOVA table (§9.4–§9.5, Procedure 9.1)

Lecture 25 Outline

- Why factorial? OFAT vs. factorial (§9.1)
- Main effects, interactions, and interaction plots (§9.2)
- The two-factor effects model (§9.3, (9.1))
- Partitioning variability and the ANOVA table (§9.4–§9.5, Procedure 9.1)
- Complete worked example: perception-model failures (2×3, $n = 2$)

Lecture 25 Outline

- Why factorial? OFAT vs. factorial (§9.1)
- Main effects, interactions, and interaction plots (§9.2)
- The two-factor effects model (§9.3, (9.1))
- Partitioning variability and the ANOVA table (§9.4–§9.5, Procedure 9.1)
- Complete worked example: perception-model failures (2×3, $n = 2$)
- When the interaction is significant: warehouse picking (3×3, $n = 4$)

Why Factorial Designs?

Section 9.1

One-factor-at-a-time (OFAT): fix all factors but one, vary it, pick the winner, move on. Sequential A/B testing at a startup is OFAT.

Why Factorial Designs?

Section 9.1

One-factor-at-a-time (OFAT): fix all factors but one, vary it, pick the winner, move on. Sequential A/B testing at a startup is OFAT.

What OFAT misses: *interactions*. The effect of A may depend on the level of B. OFAT cannot detect this — the sequential winners combined can lose to a combination never tested.

Why Factorial Designs?

Section 9.1

One-factor-at-a-time (OFAT): fix all factors but one, vary it, pick the winner, move on. Sequential A/B testing at a startup is OFAT.

What OFAT misses: *interactions*. The effect of A may depend on the level of B. OFAT cannot detect this — the sequential winners combined can lose to a combination never tested.

Factorial design: run *all combinations*. With A at a levels and B at b levels: ab cells, each replicated n times. Total runs $N = abn$.

Why Factorial Designs?

Section 9.1

One-factor-at-a-time (OFAT): fix all factors but one, vary it, pick the winner, move on. Sequential A/B testing at a startup is OFAT.

What OFAT misses: *interactions*. The effect of A may depend on the level of B. OFAT cannot detect this — the sequential winners combined can lose to a combination never tested.

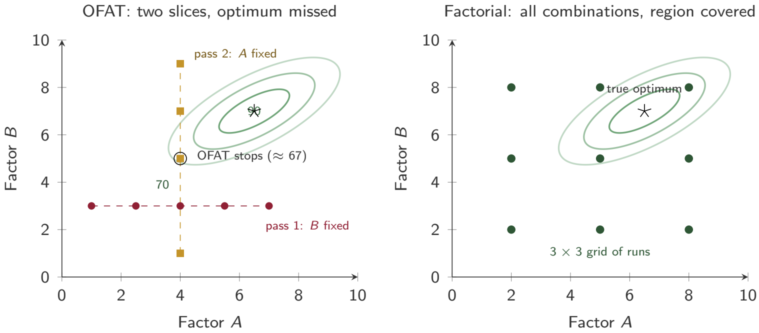
Factorial design: run *all combinations*. With A at a levels and B at b levels: ab cells, each replicated n times. Total runs $N = abn$.

Two advantages:

- Detects interactions, because every level of A is observed at every level of B.
- More efficient: every observation informs *both* main effects (hidden replication).

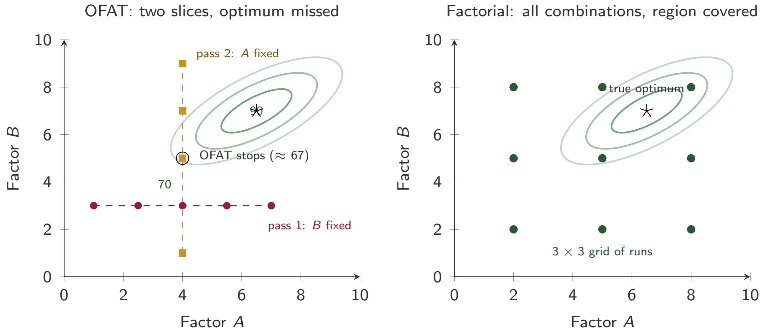
OFAT Misses the Optimum

Figure 9.1 in the textbook



OFAT Misses the Optimum

Figure 9.1 in the textbook



OFAT explores two perpendicular slices and settles near a response of 67, never seeing the peak near 90. The factorial grid covers the whole region and points straight at it (Figure 9.1).

Main Effects vs. Interactions

Section 9.2

Main effect of A: the average change in the response when A changes level, averaged over all levels of B. Answers “does this factor matter *on average*?”

Main Effects vs. Interactions

Section 9.2

Main effect of A: the average change in the response when A changes level, averaged over all levels of B. Answers "does this factor matter *on average*?"

Interaction AB: the effect of A is *not the same* at every level of B. Compare two 2×2 tables of cell means:

	B_1	B_2
A_1	20	30
A_2	40	52

versus

	B_1	B_2
A_1	20	30
A_2	40	12

Main Effects vs. Interactions

Section 9.2

Main effect of A: the average change in the response when A changes level, averaged over all levels of B. Answers "does this factor matter *on average*?"

Interaction AB: the effect of A is *not the same* at every level of B. Compare two 2×2 tables of cell means:

	B_1	B_2
A_1	20	30
A_2	40	52

versus

	B_1	B_2
A_1	20	30
A_2	40	12

Left: moving $A_1 \to A_2$ adds $+20$ at B_1 and $+22$ at B_2 — essentially no interaction; "A_2 beats A_1 by about 21" is a fair summary.

Main Effects vs. Interactions

Section 9.2

Main effect of A: the average change in the response when A changes level, averaged over all levels of B. Answers "does this factor matter *on average*?"

Interaction AB: the effect of A is *not the same* at every level of B. Compare two 2×2 tables of cell means:

	B_1	B_2
A_1	20	30
A_2	40	52

versus

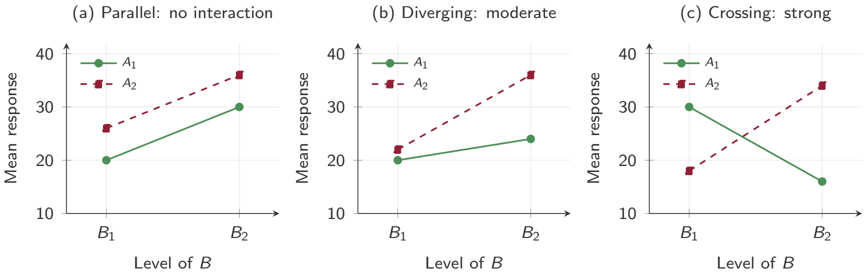
	B_1	B_2
A_1	20	30
A_2	40	12

Left: moving $A_1 \to A_2$ adds $+20$ at B_1 and $+22$ at B_2 — essentially no interaction; "A_2 beats A_1 by about 21" is a fair summary.

Right: the same move adds $+20$ at B_1 but *subtracts* 18 at B_2. No single sentence about A alone is honest — the effect of A depends entirely on B.

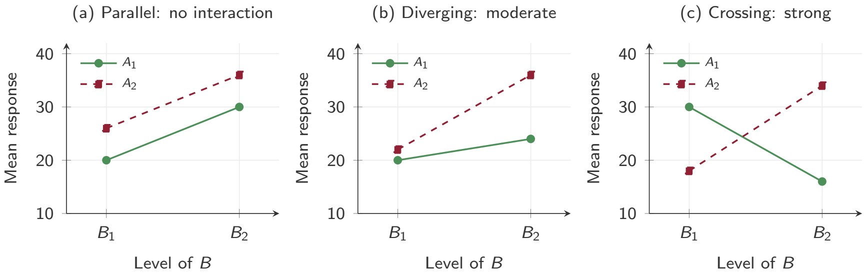
Reading an Interaction Plot: Three Patterns

Figure 9.2 in the textbook



Reading an Interaction Plot: Three Patterns

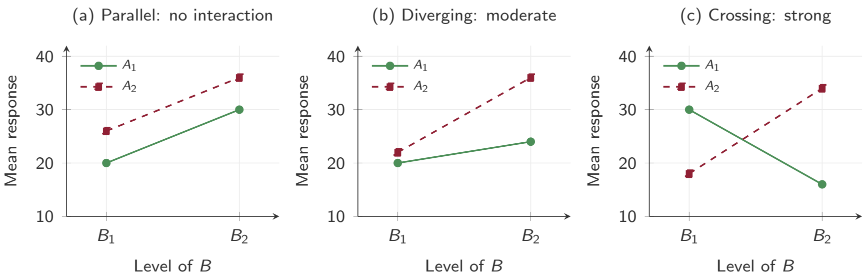
Figure 9.2 in the textbook



Read in this order: (1) Are the lines parallel? If yes, no visible interaction. (2) If not parallel, do they cross? Crossing means even the *direction* of the A effect reverses. (3) Only then read the main effects (average slope, average separation).

Reading an Interaction Plot: Three Patterns

Figure 9.2 in the textbook

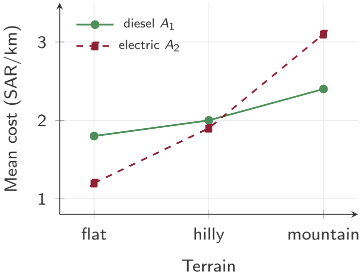


Read in this order: (1) Are the lines parallel? If yes, no visible interaction. (2) If not parallel, do they cross? Crossing means even the *direction* of the A effect reverses. (3) Only then read the main effects (average slope, average separation).

Note. The plot shows sample means, which contain noise. Real lines are never perfectly parallel — the ANOVA is the formal test of whether the departure is larger than chance.

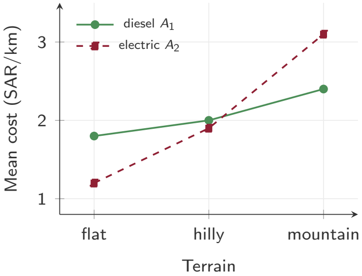
Example: Fleet Energy Cost (Truck Type $\times$ Terrain)

Example 9.1 in the textbook



Example: Fleet Energy Cost (Truck Type × Terrain)

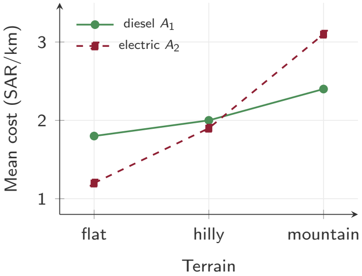
Example 9.1 in the textbook



Step 1 — Parallel? No. The gap goes from -0.6 (flat) to $+0.7$ (mountain): truck type and terrain interact.

Example: Fleet Energy Cost (Truck Type $\times$ Terrain)

Example 9.1 in the textbook

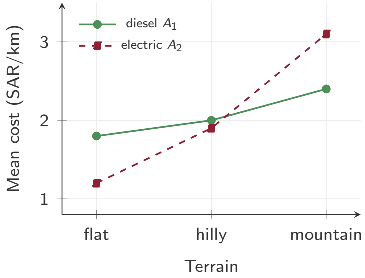


Step 1 — Parallel? No. The gap goes from -0.6 (flat) to $+0.7$ (mountain): truck type and terrain interact.

Step 2 — Crossing? Yes, between hilly and mountain. Electric is cheaper on flat routes; diesel is cheaper on mountain routes.

Example: Fleet Energy Cost (Truck Type $\times$ Terrain)

Example 9.1 in the textbook



Step 1 — Parallel? No. The gap goes from -0.6 (flat) to $+0.7$ (mountain): truck type and terrain interact.

Step 2 — Crossing? Yes, between hilly and mountain. Electric is cheaper on flat routes; diesel is cheaper on mountain routes.

Step 3 — Report. Both row means equal 2.067 SAR/km — yet "the types cost the same" is *false* on every terrain. Report terrain by terrain: electric on flat, diesel on mountain.

The Two-Factor Effects Model

Section 9.3

With A at a levels, B at b levels, n replicates per cell:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n \tag{9.1}$$

The Two-Factor Effects Model

Section 9.3

With A at a levels, B at b levels, n replicates per cell:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n \tag{9.1}$$

- μ: overall mean τ_i: main effect of A β_j: main effect of B
- $(\tau\beta)_{ij}$: interaction effect — one parameter, *not* a product
- $\varepsilon_{ijk} \sim N(0, \sigma^2)$, independent

The Two-Factor Effects Model

Section 9.3

With A at a levels, B at b levels, n replicates per cell:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \qquad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n \tag{9.1}$$

- μ: overall mean $\quad \tau_i$: main effect of A $\quad \beta_j$: main effect of B
- $(\tau\beta)_{ij}$: interaction effect — one parameter, *not* a product
- $\varepsilon_{ijk} \sim N(0, \sigma^2)$, independent

Constraints: $\sum_i \tau_i = 0, \quad \sum_j \beta_j = 0, \quad \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0.$

The Two-Factor Effects Model

Section 9.3

With A at a levels, B at b levels, n replicates per cell:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \qquad i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, n \tag{9.1}$$

- μ: overall mean $\quad \tau_i$: main effect of A $\quad \beta_j$: main effect of B
- $(\tau\beta)_{ij}$: interaction effect — one parameter, *not* a product
- $\varepsilon_{ijk} \sim N(0, \sigma^2)$, independent

Constraints: $\sum_i \tau_i = 0, \quad \sum_j \beta_j = 0, \quad \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0.$

Cell mean: $\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}$. The first three terms are the additive prediction; $(\tau\beta)_{ij}$ is how far that cell beats or misses it. Parallel lines in the interaction plot = every $(\tau\beta)_{ij}$ is zero.

Three Hypotheses, and the Dot Notation

The experiment tests three null hypotheses at once:

$$H_0^{AB}: (\tau\beta)_{ij} = 0\ \forall i, j, \qquad H_0^{A}: \tau_i = 0\ \forall i, \qquad H_0^{B}: \beta_j = 0\ \forall j \tag{9.2}$$

Three Hypotheses, and the Dot Notation

The experiment tests three null hypotheses at once:

$$H_0^{AB} : (\tau\beta)_{ij} = 0 \ \forall i, j, \qquad H_0^{A} : \tau_i = 0 \ \forall i, \qquad H_0^{B} : \beta_j = 0 \ \forall j \tag{9.2}$$

The interaction hypothesis is written *first* deliberately — it must also be *tested* first (why: coming up).

Three Hypotheses, and the Dot Notation

The experiment tests three null hypotheses at once:

$$H_0^{AB} : (\tau\beta)_{ij} = 0 \ \forall i, j, \qquad H_0^{A} : \tau_i = 0 \ \forall i, \qquad H_0^{B} : \beta_j = 0 \ \forall j \tag{9.2}$$

The interaction hypothesis is written *first* deliberately — it must also be *tested* first (why: coming up).

Dot notation (a dot means "summed over that index"):

- $y_{i\cdot\cdot}$: row total (bn observations) $y_{\cdot j\cdot}$: column total (an observations)
- $y_{ij\cdot}$: cell total (n replicates) $y_{\cdot\cdot\cdot}$: grand total ($N = abn$)
- Bars are means: $\bar{y}_{ij\cdot} = y_{ij\cdot}/n$ is a cell mean.

Partitioning the Variability

Section 9.4

Total variability splits into *four* components:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \qquad (9.3)$$

Partitioning the Variability

Section 9.4

Total variability splits into *four* components:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \tag{9.3}$$

- SS_A: variability between the *row* averages
- SS_B: variability between the *column* averages
- SS_{AB}: departure of the cell means from additivity
- SS_E: scatter of replicates inside their own cells

Partitioning the Variability

Section 9.4

Total variability splits into *four* components:

$$SS_T \;=\; SS_A + SS_B + SS_{AB} + SS_E \tag{9.3}$$

- SS_A: variability between the *row* averages
- SS_B: variability between the *column* averages
- SS_{AB}: departure of the cell means from additivity
- SS_E: scatter of replicates inside their own cells

The four pieces are the sample versions of τ_i, β_j, $(\tau\beta)_{ij}$, and ε_{ijk} from (9.1): each sum of squares grows when its set of effects is real and stays small when they are zero.

Hand-Computation Formulas

Equations 9.4–9.8

Work with *totals*, not means. Correction term: $C = y_{...}^2/abn$.

$$SS_T = \sum_{i,j,k} y_{ijk}^2 - C \quad (9.4), \qquad SS_A = \frac{\sum_i y_{i..}^2}{bn} - C \quad (9.5), \qquad SS_B = \frac{\sum_j y_{.j.}^2}{an} - C \quad (9.6)$$

Hand-Computation Formulas

Equations 9.4–9.8

Work with *totals*, not means. Correction term: $C = y_{\cdots}^2/abn$.

$$SS_T = \sum_{i,j,k} y_{ijk}^2 - C \quad (9.4), \qquad SS_A = \frac{\sum_i y_{i\cdot\cdot}^2}{bn} - C \quad (9.5), \qquad SS_B = \frac{\sum_j y_{\cdot j\cdot}^2}{an} - C \quad (9.6)$$

$$SS_{AB} = \frac{1}{n}\sum_{i,j} y_{ij\cdot}^2 - C - SS_A - SS_B \quad (9.7), \qquad SS_E = SS_T - SS_A - SS_B - SS_{AB} \quad (9.8)$$

Hand-Computation Formulas

Equations 9.4–9.8

Work with *totals*, not means. Correction term: $C = y_{\dots}^2/abn$.

$$SS_T = \sum_{i,j,k} y_{ijk}^2 - C \quad (9.4), \qquad SS_A = \frac{\sum_i y_{i\cdot\cdot}^2}{bn} - C \quad (9.5), \qquad SS_B = \frac{\sum_j y_{\cdot j\cdot}^2}{an} - C \quad (9.6)$$

$$SS_{AB} = \frac{1}{n}\sum_{i,j} y_{ij\cdot}^2 - C - SS_A - SS_B \quad (9.7), \qquad SS_E = SS_T - SS_A - SS_B - SS_{AB} \quad (9.8)$$

Pattern: square the totals, divide by the number of observations inside each total (bn per row, an per column, n per cell), subtract C.

Hand-Computation Formulas

Equations 9.4–9.8

Work with *totals*, not means. Correction term: $C = y_{\cdots}^2 / abn$.

$$SS_T = \sum_{i,j,k} y_{ijk}^2 - C \quad (9.4), \qquad SS_A = \frac{\sum_i y_{i\cdot\cdot}^2}{bn} - C \quad (9.5), \qquad SS_B = \frac{\sum_j y_{\cdot j\cdot}^2}{an} - C \quad (9.6)$$

$$SS_{AB} = \frac{1}{n} \sum_{i,j} y_{ij\cdot}^2 - C - SS_A - SS_B \quad (9.7), \qquad SS_E = SS_T - SS_A - SS_B - SS_{AB} \quad (9.8)$$

Pattern: square the totals, divide by the number of observations inside each total (bn per row, an per column, n per cell), subtract C.

In (9.7), subtracting SS_A and SS_B from the cell-total variability leaves pure interaction.

Degrees of Freedom

The degrees of freedom split the same way as the sums of squares:

$$\underbrace{(a-1)}_{A}+\underbrace{(b-1)}_{B}+\underbrace{(a-1)(b-1)}_{AB}+\underbrace{ab(n-1)}_{\text{Error}} = abn-1 \tag{9.9}$$

Degrees of Freedom

The degrees of freedom split the same way as the sums of squares:

$$\underbrace{(a-1)}_{A}+\underbrace{(b-1)}_{B}+\underbrace{(a-1)(b-1)}_{AB}+\underbrace{ab(n-1)}_{\text{Error}} = abn-1 \tag{9.9}$$

- Row averages consume $a-1$ free deviations; column averages $b-1$.
- The ab cell means, after the mean and main effects, leave $(a-1)(b-1)$ for interaction.
- Each cell contributes $n-1$ within-cell deviations to error.

Degrees of Freedom

The degrees of freedom split the same way as the sums of squares:

$$\underbrace{(a-1)}_{A}+\underbrace{(b-1)}_{B}+\underbrace{(a-1)(b-1)}_{AB}+\underbrace{ab(n-1)}_{\text{Error}} \;=\; abn-1 \tag{9.9}$$

- Row averages consume $a-1$ free deviations; column averages $b-1$.
- The ab cell means, after the mean and main effects, leave $(a-1)(b-1)$ for interaction.
- Each cell contributes $n-1$ within-cell deviations to error.

Check every time: the df column must add to $abn-1$. If it does not, there is an arithmetic slip upstream. Note also $df_E = ab(n-1)$ requires $n \geq 2$ — with $n = 1$ there are zero error df (Lecture 26).

The Two-Factor ANOVA Table

Section 9.5, Table 9.1

Source	**SS**	**df**	**MS**	$\mathbf{F_0}$
A (rows)	SS_A	$a-1$	MS_A	MS_A/MS_E
B (columns)	SS_B	$b-1$	MS_B	MS_B/MS_E
AB (interaction)	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	MS_{AB}/MS_E
Error	SS_E	$ab(n-1)$	MS_E	
Total	SS_T	$abn-1$		

The Two-Factor ANOVA Table

Section 9.5, Table 9.1

Source	**SS**	**df**	**MS**	$\mathbf{F_0}$
A (rows)	SS_A	$a-1$	MS_A	MS_A/MS_E
B (columns)	SS_B	$b-1$	MS_B	MS_B/MS_E
AB (interaction)	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	MS_{AB}/MS_E
Error	SS_E	$ab(n-1)$	MS_E	
Total	SS_T	$abn-1$		

Divide each SS by its df to get the mean squares. MS_E is always an unbiased estimator of σ^2; each other mean square estimates σ^2 *plus* a positive term from its own effects.

The Two-Factor ANOVA Table

Section 9.5, Table 9.1

Source	**SS**	**df**	**MS**	$\mathbf{F_0}$
A (rows)	SS_A	$a-1$	MS_A	MS_A/MS_E
B (columns)	SS_B	$b-1$	MS_B	MS_B/MS_E
AB (interaction)	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	MS_{AB}/MS_E
Error	SS_E	$ab(n-1)$	MS_E	
Total	SS_T	$abn-1$		

Divide each SS by its df to get the mean squares. MS_E is always an unbiased estimator of σ^2; each other mean square estimates σ^2 *plus* a positive term from its own effects.

So each test is a ratio with MS_E in the denominator — large when its effects are real. Three tests, *one* shared error term.

The Three F Statistics

Interaction statistic (test this one first):

$$F_0^{AB} = \frac{MS_{AB}}{MS_E} = \frac{SS_{AB}/[(a-1)(b-1)]}{SS_E/[ab(n-1)]} \tag{9.10}$$

The Three F Statistics

Interaction statistic (test this one first):

$$F_0^{AB} = \frac{MS_{AB}}{MS_E} = \frac{SS_{AB}/[(a-1)(b-1)]}{SS_E/[ab(n-1)]} \tag{9.10}$$

Under H_0^{AB} it follows the F distribution with $(a-1)(b-1)$ and $ab(n-1)$ df.

The Three F Statistics

Interaction statistic (test this one first):

$$F_0^{AB} = \frac{MS_{AB}}{MS_E} = \frac{SS_{AB}/[(a-1)(b-1)]}{SS_E/[ab(n-1)]} \tag{9.10}$$

Under H_0^{AB} it follows the F distribution with $(a-1)(b-1)$ and $ab(n-1)$ df.

Main-effect statistics:

$$F_0^A = \frac{MS_A}{MS_E} \text{ with } (a-1),\ ab(n-1) \text{ df}, \qquad F_0^B = \frac{MS_B}{MS_E} \text{ with } (b-1),\ ab(n-1) \text{ df} \tag{9.11}$$

The Three F Statistics

Interaction statistic (test this one first):

$$F_0^{AB} = \frac{MS_{AB}}{MS_E} = \frac{SS_{AB}/[(a-1)(b-1)]}{SS_E/[ab(n-1)]} \tag{9.10}$$

Under H_0^{AB} it follows the F distribution with $(a-1)(b-1)$ and $ab(n-1)$ df.

Main-effect statistics:

$$F_0^A = \frac{MS_A}{MS_E} \text{ with } (a-1),\ ab(n-1) \text{ df,} \qquad F_0^B = \frac{MS_B}{MS_E} \text{ with } (b-1),\ ab(n-1) \text{ df} \tag{9.11}$$

Each hypothesis in (9.2) is rejected at level α when its statistic exceeds f_{α,ν_1,ν_2} with the matching df.

Why Test the Interaction First?

The main effect of A is an *average* of the A effect over the levels of B.

Why Test the Interaction First?

The main effect of A is an *average* of the A effect over the levels of B.

If the interaction is significant, that average may combine a large gain at one level of B with a large loss at another — and can come out near *zero* while A matters strongly level-by-level. (The fleet example: equal row means, opposite winners by terrain.)

Why Test the Interaction First?

The main effect of A is an *average* of the A effect over the levels of B.

If the interaction is significant, that average may combine a large gain at one level of B with a large loss at another — and can come out near *zero* while A matters strongly level-by-level. (The fleet example: equal row means, opposite winners by terrain.)

So the interaction test decides how the other two tests may be read:

- H_0^{AB} rejected $\Rightarrow$ interpret main effects cautiously; work with *cell means*, level by level.
- H_0^{AB} not rejected $\Rightarrow$ the additive story holds; report main effects directly.

Why Test the Interaction First?

The main effect of A is an *average* of the A effect over the levels of B.

If the interaction is significant, that average may combine a large gain at one level of B with a large loss at another — and can come out near *zero* while A matters strongly level-by-level. (The fleet example: equal row means, opposite winners by terrain.)

So the interaction test decides how the other two tests may be read:

- H_0^{AB} rejected $\Rightarrow$ interpret main effects cautiously; work with *cell means*, level by level.
- H_0^{AB} not rejected $\Rightarrow$ the additive story holds; report main effects directly.

AI-safety reading. A leaderboard that averages an eval score over scenario families reports only the main effect of model version — and hides every version-by-scenario interaction, which is often where the safety-relevant behavior lives.

Procedure 9.1: Building the Two-Factor ANOVA Table

1. Organize the data in an $a \times b$ table. Compute cell totals $y_{ij\cdot}$, row totals $y_{i\cdot\cdot}$, column totals $y_{\cdot j\cdot}$, grand total $y_{\cdots}$. *Check: all must add to the grand total.*
2. Compute $C = y_{\cdots}^2/(abn)$ and $\sum y_{ijk}^2$; then SS_T by (9.4).
3. Compute SS_A from row totals (9.5) and SS_B from column totals (9.6).
4. Compute $\frac{1}{n}\sum y_{ij\cdot}^2 - C$, subtract SS_A and SS_B to get SS_{AB} (9.7).
5. Obtain SS_E by subtraction (9.8). *It must be nonnegative.*
6. Fill in df; verify they add to $abn - 1$ (9.9). Divide for the MS.
7. Compute F_0^{AB} (9.10); test against $f_{\alpha,(a-1)(b-1),\,ab(n-1)}$ **first**.
8. Compute F_0^A, F_0^B (9.11); state conclusions in problem language, respecting step 7.

Quick Check: Degrees of Freedom Before Running

Example 9.2 in the textbook

Q. A startup plans a factorial test of $a = 4$ landing-page designs against $b = 3$ traffic sources (search ads, social, email), with $n = 5$ replicate traffic batches per combination. The df for SS_{AB} and SS_E are, respectively:

(A) 6 and 48

(B) 7 and 59

(C) 12 and 48

(D) 6 and 59

Quick Check: Degrees of Freedom Before Running

Example 9.2 in the textbook

Q. A startup plans a factorial test of $a = 4$ landing-page designs against $b = 3$ traffic sources (search ads, social, email), with $n = 5$ replicate traffic batches per combination. The df for SS_{AB} and SS_E are, respectively:

(A) 6 and 48

(B) 7 and 59

(C) 12 and 48

(D) 6 and 59

Answer: (A). $df_{AB} = (a-1)(b-1) = 3\times2 = 6$; $df_E = ab(n-1) = 12\times4 = 48$. Check (9.9): $3 + 2 + 6 + 48 = 59 = abn - 1$. (B) and (D) confuse df_E with $df_T = 59$; (C) uses $ab = 12$ instead of $(a - 1)(b - 1)$.

Worked Example: Perception-Model Failures (2×3, $n = 2$)

Example 9.3, Table 9.2

Setup. An AI safety team validates two versions of a perception model (A: V1, V2) across three scenario types (B: clear, rain, night). Each cell: $n = 2$ test batches of 1,000 frames; response = missed detections per batch. $\alpha = 0.05$.

Worked Example: Perception-Model Failures (2×3, $n = 2$)

Example 9.3, Table 9.2

Setup. An AI safety team validates two versions of a perception model (A: V1, V2) across three scenario types (B: clear, rain, night). Each cell: $n = 2$ test batches of 1,000 frames; response = missed detections per batch. $\alpha = 0.05$.

Model	**Clear**	**Rain**	**Night**	$y_{i..}$
V1	8, 10	14, 16	21, 23	92
V2	7, 9	9, 11	20, 22	78
$y_{.j.}$	34	50	86	$y_{...} = 170$

Worked Example: Perception-Model Failures (2×3, $n = 2$)

Example 9.3, Table 9.2

Setup. An AI safety team validates two versions of a perception model (A: V1, V2) across three scenario types (B: clear, rain, night). Each cell: $n = 2$ test batches of 1,000 frames; response = missed detections per batch. $\alpha = 0.05$.

Model	**Clear**	**Rain**	**Night**	$y_{i\cdot\cdot}$
V1	8, 10	14, 16	21, 23	92
V2	7, 9	9, 11	20, 22	78
$y_{\cdot j \cdot}$	34	50	86	$y_{\cdots} = 170$

Step 1 (Procedure 9.1) — Totals. $a = 2$, $b = 3$, $n = 2$, so $N = 12$. Cell totals: $y_{11\cdot} = 18$, $y_{12\cdot} = 30$, $y_{13\cdot} = 44$, $y_{21\cdot} = 16$, $y_{22\cdot} = 20$, $y_{23\cdot} = 42$. Check: rows $92 + 78 = 170$; columns $34 + 50 + 86 = 170$. ✓

Worked Example: Correction Term and SS_T

Step 2 of Procedure 9.1

Correction term:

$$C = \frac{y_{\cdots}^2}{abn} = \frac{170^2}{12} = \frac{28900}{12} = 2408.33$$

Worked Example: Correction Term and SS_T

Step 2 of Procedure 9.1

Correction term:

$$C = \frac{y_{\cdots}^2}{abn} = \frac{170^2}{12} = \frac{28900}{12} = 2408.33$$

Sum of squared observations:

$$\sum y_{ijk}^2 = 8^2 + 10^2 + 14^2 + \cdots + 20^2 + 22^2 = 2802$$

Worked Example: Correction Term and SS_T

Step 2 of Procedure 9.1

Correction term:

$$C = \frac{y_{\cdots}^2}{abn} = \frac{170^2}{12} = \frac{28900}{12} = 2408.33$$

Sum of squared observations:

$$\sum y_{ijk}^2 = 8^2 + 10^2 + 14^2 + \cdots + 20^2 + 22^2 = 2802$$

Total sum of squares by (9.4):

$$SS_T = \sum y_{ijk}^2 - C = 2802 - 2408.33 = 393.67$$

Worked Example: Main-Effect Sums of Squares

Step 3 of Procedure 9.1

Each row total contains $bn = 3 \times 2 = 6$ observations; each column total $an = 2 \times 2 = 4$.

Worked Example: Main-Effect Sums of Squares

Step 3 of Procedure 9.1

Each row total contains $bn = 3 \times 2 = 6$ observations; each column total $an = 2 \times 2 = 4$.

Model version (*A*) by (9.5):

$$SS_A = \frac{92^2 + 78^2}{6} - C = \frac{14548}{6} - 2408.33$$

Worked Example: Main-Effect Sums of Squares

Step 3 of Procedure 9.1

Each row total contains $bn = 3 \times 2 = 6$ observations; each column total $an = 2 \times 2 = 4$.

Model version (A) by (9.5):

$$SS_A = \frac{92^2 + 78^2}{6} - C = \frac{14548}{6} - 2408.33 = 2424.67 - 2408.33 = 16.33$$

Worked Example: Main-Effect Sums of Squares

Step 3 of Procedure 9.1

Each row total contains $bn = 3 \times 2 = 6$ observations; each column total $an = 2 \times 2 = 4$.

Model version (A) by (9.5):

$$SS_A = \frac{92^2 + 78^2}{6} - C = \frac{14548}{6} - 2408.33 = 2424.67 - 2408.33 = 16.33$$

Scenario type (B) by (9.6):

$$SS_B = \frac{34^2 + 50^2 + 86^2}{4} - C = \frac{11052}{4} - 2408.33$$

Worked Example: Main-Effect Sums of Squares

Step 3 of Procedure 9.1

Each row total contains $bn = 3 \times 2 = 6$ observations; each column total $an = 2 \times 2 = 4$.

Model version (A) by (9.5):

$$SS_A = \frac{92^2 + 78^2}{6} - C = \frac{14548}{6} - 2408.33 = 2424.67 - 2408.33 = 16.33$$

Scenario type (B) by (9.6):

$$SS_B = \frac{34^2 + 50^2 + 86^2}{4} - C = \frac{11052}{4} - 2408.33 = 2763.00 - 2408.33 = 354.67$$

Worked Example: Interaction and Error

Steps 4–5 of Procedure 9.1

Step 4 — Interaction by (9.7). The six cell totals give

$$\frac{1}{n}\sum_{i,j} y_{ij\cdot}^2 - C = \frac{18^2 + 30^2 + 44^2 + 16^2 + 20^2 + 42^2}{2} - 2408.33 = 381.67$$

Worked Example: Interaction and Error

Steps 4–5 of Procedure 9.1

Step 4 — Interaction by (9.7). The six cell totals give

$$\frac{1}{n}\sum_{i,j} y_{ij\cdot}^2 - C = \frac{18^2 + 30^2 + 44^2 + 16^2 + 20^2 + 42^2}{2} - 2408.33 = 381.67$$

$$SS_{AB} = 381.67 - SS_A - SS_B = 381.67 - 16.33 - 354.67 = 10.67$$

Worked Example: Interaction and Error

Steps 4–5 of Procedure 9.1

Step 4 — Interaction by (9.7). The six cell totals give

$$\frac{1}{n}\sum_{i,j} y_{ij\cdot}^2 - C = \frac{18^2 + 30^2 + 44^2 + 16^2 + 20^2 + 42^2}{2} - 2408.33 = 381.67$$

$$SS_{AB} = 381.67 - SS_A - SS_B = 381.67 - 16.33 - 354.67 = 10.67$$

Step 5 — Error by (9.8) (nonnegative ✓— a negative SS_E signals an arithmetic error):

$$SS_E = SS_T - SS_A - SS_B - SS_{AB} = 393.67 - 16.33 - 354.67 - 10.67 = 12.00$$

Worked Example: Degrees of Freedom and Mean Squares

Step 6 of Procedure 9.1

Degrees of freedom:

$$df_A = a - 1 = 1, \qquad df_B = b - 1 = 2,$$

$$df_{AB} = (a-1)(b-1) = 2, \qquad df_E = ab(n-1) = 6$$

Worked Example: Degrees of Freedom and Mean Squares

Step 6 of Procedure 9.1

Degrees of freedom:

$$df_A = a - 1 = 1, \qquad df_B = b - 1 = 2,$$

$$df_{AB} = (a-1)(b-1) = 2, \qquad df_E = ab(n-1) = 6$$

Check (9.9): $1 + 2 + 2 + 6 = 11 = abn - 1$. ✓

Worked Example: Degrees of Freedom and Mean Squares

Step 6 of Procedure 9.1

Degrees of freedom:

$$df_A = a - 1 = 1, \qquad df_B = b - 1 = 2,$$

$$df_{AB} = (a-1)(b-1) = 2, \qquad df_E = ab(n-1) = 6$$

Check (9.9): $1 + 2 + 2 + 6 = 11 = abn - 1.$ ✓

Mean squares (SS divided by df):

$$MS_A = \frac{16.33}{1} = 16.33, \qquad MS_B = \frac{354.67}{2} = 177.33,$$

$$MS_{AB} = \frac{10.67}{2} = 5.333, \qquad MS_E = \frac{12.00}{6} = 2.000$$

Worked Example: F Tests — Interaction First

Steps 7–8, Table 9.3

$f_{0.05,\nu_1,\nu_2}$ excerpt (Table 9.3); error df $\nu_2 = 6$:

	ν_1 (numerator df)				
ν_2	1	2	3	4	5
5	6.61	5.79	5.41	5.19	5.05
6	5.99	5.14	4.76	4.53	4.39
7	5.59	4.74	4.35	4.12	3.97

Worked Example: F Tests — Interaction First

Steps 7–8, Table 9.3

$f_{0.05,\nu_1,\nu_2}$ excerpt (Table 9.3); error df $\nu_2 = 6$:

	ν_1 (numerator df)				
ν_2	1	2	3	4	5
5	6.61	5.79	5.41	5.19	5.05
6	5.99	5.14	4.76	4.53	4.39
7	5.59	4.74	4.35	4.12	3.97

Step 7 — Interaction first (9.10): $F_0^{AB} = 5.333/2.000 = 2.67 < f_{0.05,2,6} = 5.14$
$\Rightarrow$ do **not** reject H_0^{AB}: no significant interaction; main effects may be read directly.

Worked Example: F Tests — Interaction First

Steps 7–8, Table 9.3

$f_{0.05,\nu_1,\nu_2}$ excerpt (Table 9.3); error df $\nu_2 = 6$:

	ν_1 (numerator df)				
ν_2	1	2	3	4	5
5	6.61	5.79	5.41	5.19	5.05
6	5.99	5.14	4.76	4.53	4.39
7	5.59	4.74	4.35	4.12	3.97

Step 7 — Interaction first (9.10): $F_0^{AB} = 5.333/2.000 = 2.67 < f_{0.05,2,6} = 5.14$ $\Rightarrow$ do **not** reject H_0^{AB}: no significant interaction; main effects may be read directly.

Step 8 — Main effects (9.11): $F_0^{A} = 16.33/2.000 = 8.17 > f_{0.05,1,6} = 5.99$ $\Rightarrow$ reject H_0^{A}.

Worked Example: F Tests — Interaction First

Steps 7–8, Table 9.3

$f_{0.05,\nu_1,\nu_2}$ excerpt (Table 9.3); error df $\nu_2 = 6$:

	ν_1 (numerator df)				
ν_2	1	2	3	4	5
5	6.61	5.79	5.41	5.19	5.05
6	5.99	5.14	4.76	4.53	4.39
7	5.59	4.74	4.35	4.12	3.97

Step 7 — Interaction first (9.10): $F_0^{AB} = 5.333/2.000 = 2.67 < f_{0.05,2,6} = 5.14$ $\Rightarrow$ do **not** reject H_0^{AB}: no significant interaction; main effects may be read directly.

Step 8 — Main effects (9.11): $F_0^{A} = 16.33/2.000 = 8.17 > f_{0.05,1,6} = 5.99 \quad \Rightarrow$ reject H_0^{A}.

$$F_0^{B} = 177.33/2.000 = 88.67 > f_{0.05,2,6} = 5.14 \quad \Rightarrow \text{reject } H_0^{B}.$$

Worked Example: ANOVA Table and Conclusion

Table 9.4

Source	**SS**	**df**	**MS**	F_0	$f_{0.05}$
Model version (A)	16.33	1	16.33	**8.17**	5.99
Scenario type (B)	354.67	2	177.33	**88.67**	5.14
Interaction (AB)	10.67	2	5.333	2.67	5.14
Error	12.00	6	2.000		
Total	393.67	11			

Worked Example: ANOVA Table and Conclusion

Table 9.4

Source	**SS**	**df**	**MS**	F_0	$f_{0.05}$
Model version (A)	16.33	1	16.33	**8.17**	5.99
Scenario type (B)	354.67	2	177.33	**88.67**	5.14
Interaction (AB)	10.67	2	5.333	2.67	5.14
Error	12.00	6	2.000		
Total	393.67	11			

In problem language. Scenario type dominates: night scenes produce far more missed detections than clear scenes for both versions ($\bar{y}_{\cdot 1 \cdot} = 8.5$ vs. $\bar{y}_{\cdot 3 \cdot} = 21.5$ misses per batch).

Worked Example: ANOVA Table and Conclusion

Table 9.4

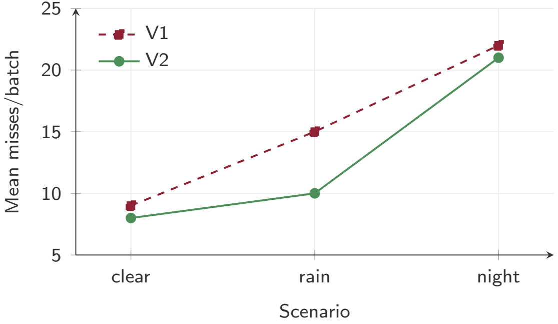
Source	**SS**	**df**	**MS**	F_0	$f_{0.05}$
Model version (A)	16.33	1	16.33	**8.17**	5.99
Scenario type (B)	354.67	2	177.33	**88.67**	5.14
Interaction (AB)	10.67	2	5.333	2.67	5.14
Error	12.00	6	2.000		
Total	393.67	11			

In problem language. Scenario type dominates: night scenes produce far more missed detections than clear scenes for both versions ($\bar{y}_{\cdot 1 \cdot} = 8.5$ vs. $\bar{y}_{\cdot 3 \cdot} = 21.5$ misses per batch).

Averaged over scenarios, V2 misses significantly fewer detections than V1 ($\bar{y}_{2 \cdot \cdot} = 13.0$ vs. $\bar{y}_{1 \cdot \cdot} = 15.33$) — and since the interaction is *not* significant, that advantage is honestly quoted as a single number.

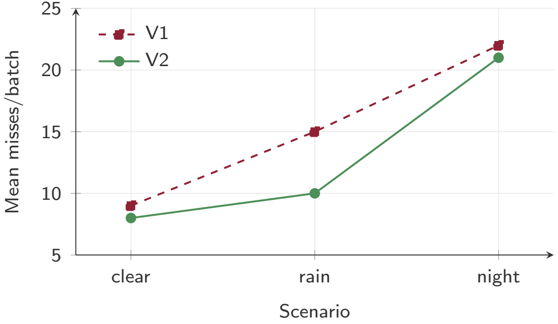
Worked Example: The Interaction Plot Agrees

Figure 9.4



Worked Example: The Interaction Plot Agrees

Figure 9.4



The lines are roughly parallel; the rain gap is within noise ($F_0^{AB} = 2.67$, not significant). Main effects summarize the study fairly: V2 is better everywhere, and night is hard for both.

When the Interaction IS Significant: Warehouse Picking

Section 9.7, Example 9.4

Setup. A fulfillment engineer studies picking time (s per order line) under three warehouse layouts (A) and three shifts (B), $n = 4$ picks per cell, $N = 36$ (data: Table 9.5). Row totals 998, 1300, 1501; column totals 1738, 1291, 770; grand 3799.

When the Interaction IS Significant: Warehouse Picking

Section 9.7, Example 9.4

Setup. A fulfillment engineer studies picking time (s per order line) under three warehouse layouts (A) and three shifts (B), $n = 4$ picks per cell, $N = 36$ (data: Table 9.5). Row totals 998, 1300, 1501; column totals 1738, 1291, 770; grand 3799.

Running Procedure 9.1 ($C = 3799^2/36 = 400900.03$) gives Table 9.6:

Source	**SS**	**df**	**MS**	F_0	$f_{0.05}$	**Decision**
Layout (A)	10683.72	2	5341.86	**7.91**	3.35	reject H_0^A
Shift (B)	39118.72	2	19559.36	**28.97**	3.35	reject H_0^B
Interaction (AB)	9613.78	4	2403.44	**3.56**	2.73	reject H_0^{AB}
Error	18230.75	27	675.21			
Total	77646.97	35				

When the Interaction IS Significant: Warehouse Picking

Section 9.7, Example 9.4

Setup. A fulfillment engineer studies picking time (s per order line) under three warehouse layouts (A) and three shifts (B), $n = 4$ picks per cell, $N = 36$ (data: Table 9.5). Row totals 998, 1300, 1501; column totals 1738, 1291, 770; grand 3799.

Running Procedure 9.1 ($C = 3799^2/36 = 400900.03$) gives Table 9.6:

Source	SS	df	MS	F_0	$f_{0.05}$	Decision
Layout (A)	10683.72	2	5341.86	**7.91**	3.35	reject H_0^A
Shift (B)	39118.72	2	19559.36	**28.97**	3.35	reject H_0^B
Interaction (AB)	9613.78	4	2403.44	**3.56**	2.73	reject H_0^{AB}
Error	18230.75	27	675.21			
Total	77646.97	35				

All three rejected: layout matters, shift matters strongly, and *the fastest layout depends on the shift*.

Picking Study: Read the Cell Means, Not the Row Means

Table 9.7, Figure 9.5

Cell means $\bar{y}_{ij\cdot} = y_{ij\cdot}/4$:

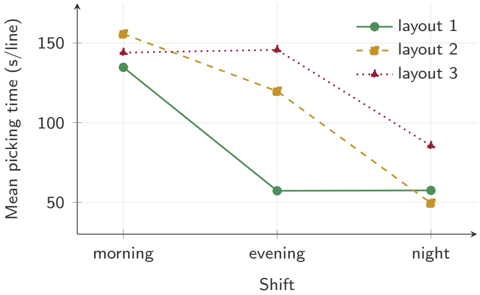
Layout	**Morn.**	**Even.**	**Night**
1	134.75	57.25	57.50
2	155.75	119.75	49.50
3	144.00	145.75	85.50

Picking Study: Read the Cell Means, Not the Row Means

Table 9.7, Figure 9.5

Cell means $\bar{y}_{ij\cdot} = y_{ij\cdot}/4$:

Layout	Morn.	Even.	Night
1	134.75	57.25	57.50
2	155.75	119.75	49.50
3	144.00	145.75	85.50

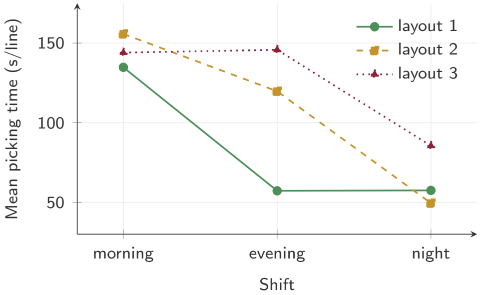


Picking Study: Read the Cell Means, Not the Row Means

Table 9.7, Figure 9.5

Cell means $\bar{y}_{ij\cdot} = y_{ij\cdot}/4$:

Layout	Morn.	Even.	Night
1	134.75	57.25	57.50
2	155.75	119.75	49.50
3	144.00	145.75	85.50



The lines are far from parallel and the layout ranking changes across shifts — layout 1 leads in the evening, layout 2 at night. This is the picture behind $F_0^{AB} = 3.56$.

Picking Study: Recommendation and Follow-Up

Column by column: morning — all layouts congested and comparable; evening — layout 1 much faster (57.25 s vs. 145.75 s for layout 3); night — layout 2 fastest (49.50 s), layout 1 close (57.50 s).

Picking Study: Recommendation and Follow-Up

Column by column: morning — all layouts congested and comparable; evening — layout 1 much faster (57.25 s vs. 145.75 s for layout 3); night — layout 2 fastest (49.50 s), layout 1 close (57.50 s).

No layout wins every shift — exactly what the significant interaction means. If one must be chosen, the robust choice is **layout 1**: fastest or within about 8 s of fastest in every shift, lowest average (83.17 s vs. 108.33 and 125.08). Row means alone would hide that layout 2 beats it at night.

Picking Study: Recommendation and Follow-Up

Column by column: morning — all layouts congested and comparable; evening — layout 1 much faster (57.25 s vs. 145.75 s for layout 3); night — layout 2 fastest (49.50 s), layout 1 close (57.50 s).

No layout wins every shift — exactly what the significant interaction means. If one must be chosen, the robust choice is **layout 1**: fastest or within about 8 s of fastest in every shift, lowest average (83.17 s vs. 108.33 and 125.08). Row means alone would hide that layout 2 beats it at night.

Follow-up comparisons (Chapter 8 tools):

- Interaction *not* significant: LSD or Tukey on row/column means, with MS_E and $ab(n-1)$ df.
- Interaction *significant*: compare within a fixed level of the other factor (layouts within the night shift) — averaged comparisons describe no real operating condition.

Picking Study: Recommendation and Follow-Up

Column by column: morning — all layouts congested and comparable; evening — layout 1 much faster (57.25 s vs. 145.75 s for layout 3); night — layout 2 fastest (49.50 s), layout 1 close (57.50 s).

No layout wins every shift — exactly what the significant interaction means. If one must be chosen, the robust choice is **layout 1**: fastest or within about 8 s of fastest in every shift, lowest average (83.17 s vs. 108.33 and 125.08). Row means alone would hide that layout 2 beats it at night.

Follow-up comparisons (Chapter 8 tools):

- Interaction *not* significant: LSD or Tukey on row/column means, with MS_E and $ab(n-1)$ df.
- Interaction *significant*: compare within a fixed level of the other factor (layouts within the night shift) — averaged comparisons describe no real operating condition.

Engineering note. "Choose the robust cell" is an engineering rule, not a statistical one. If night throughput is the bottleneck, layout 2 may be right — statistics shows the trade-off is real; context decides it.

Lecture 25 Summary

- Factorial designs vary all factors together: they catch interactions OFAT cannot see, and every run informs every effect.

Lecture 25 Summary

- Factorial designs vary all factors together: they catch interactions OFAT cannot see, and every run informs every effect.
- Two-factor model (9.1): $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$; three hypotheses (9.2).

Lecture 25 Summary

- Factorial designs vary all factors together: they catch interactions OFAT cannot see, and every run informs every effect.
- Two-factor model (9.1): $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$; three hypotheses (9.2).
- Decomposition (9.3): $SS_T = SS_A + SS_B + SS_{AB} + SS_E$, computed from row, column, and cell totals ((9.4)–(9.8)); df add to $abn - 1$ (9.9).

Lecture 25 Summary

- Factorial designs vary all factors together: they catch interactions OFAT cannot see, and every run informs every effect.
- Two-factor model (9.1): $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$; three hypotheses (9.2).
- Decomposition (9.3): $SS_T = SS_A + SS_B + SS_{AB} + SS_E$, computed from row, column, and cell totals ((9.4)–(9.8)); df add to $abn - 1$ (9.9).
- Three F tests share one MS_E: interaction (9.10) is tested *before* main effects (9.11) — Procedure 9.1.

Lecture 25 Summary

- Factorial designs vary all factors together: they catch interactions OFAT cannot see, and every run informs every effect.
- Two-factor model (9.1): $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$; three hypotheses (9.2).
- Decomposition (9.3): $SS_T = SS_A + SS_B + SS_{AB} + SS_E$, computed from row, column, and cell totals ((9.4)–(9.8)); df add to $abn - 1$ (9.9).
- Three F tests share one MS_E: interaction (9.10) is tested *before* main effects (9.11) — Procedure 9.1.
- Interaction not significant $\Rightarrow$ report main effects (perception-model example). Significant $\Rightarrow$ work from cell means, level by level (picking example).

Lecture 25 Summary

- Factorial designs vary all factors together: they catch interactions OFAT cannot see, and every run informs every effect.
- Two-factor model (9.1): $Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$; three hypotheses (9.2).
- Decomposition (9.3): $SS_T = SS_A + SS_B + SS_{AB} + SS_E$, computed from row, column, and cell totals ((9.4)–(9.8)); df add to $abn - 1$ (9.9).
- Three F tests share one MS_E: interaction (9.10) is tested *before* main effects (9.11) — Procedure 9.1.
- Interaction not significant $\Rightarrow$ report main effects (perception-model example). Significant $\Rightarrow$ work from cell means, level by level (picking example).
- *Next lecture:* residual diagnostics (Procedure 9.2), one observation per cell, general factorials, and the 2^k designs.

Reminders

- Quiz on Chapter 9 basics (model, df counting, reading interaction plots) is due **before class**.
- Practice: redo the perception-model ANOVA on a blank sheet using Procedure 9.1 — every step, every book equation number.
- Read §9.8–§9.11 to prepare for Lecture 26.
- Office hours: Tuesdays 9–10, room 22-219, or on Zoom.