ISE 315: Engineering Statistics

Lecture 22: Introduction to ANOVA — Comparing More Than Two Means

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 22

Introduction to ANOVA: Comparing More Than Two Means (Chapter 8)

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)
- The vocabulary of designed experiments

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)
- The vocabulary of designed experiments
- The effects model (8.2) and its assumptions

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)
- The vocabulary of designed experiments
- The effects model (8.2) and its assumptions
- Partitioning total variability: $SS_T = SS_{\text{Treatments}} + SS_E$ (8.5)

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)
- The vocabulary of designed experiments
- The effects model (8.2) and its assumptions
- Partitioning total variability: $SS_T = SS_{\text{Treatments}} + SS_E$ (8.5)
- Mean squares, the F-test, and the ANOVA table (Procedure 8.1)

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)
- The vocabulary of designed experiments
- The effects model (8.2) and its assumptions
- Partitioning total variability: $SS_T = SS_{\text{Treatments}} + SS_E$ (8.5)
- Mean squares, the F-test, and the ANOVA table (Procedure 8.1)
- Worked example: three red-team strategies, by hand

Lecture 22 Outline

- Why not repeated t-tests? Family-wise error (8.1)
- The vocabulary of designed experiments
- The effects model (8.2) and its assumptions
- Partitioning total variability: $SS_T = SS_{\text{Treatments}} + SS_E$ (8.5)
- Mean squares, the F-test, and the ANOVA table (Procedure 8.1)
- Worked example: three red-team strategies, by hand
- Confidence intervals from MS_E

Beyond Two Samples

Why not just repeat two-sample t-tests?

Scenario: an AI safety group probes a deployed language model with **three** red-team strategies: prompt *mutation*, *role-play* jailbreaks, and multi-turn *escalation*. Response: defects discovered per session.

Beyond Two Samples

Why not just repeat two-sample t-tests?

Scenario: an AI safety group probes a deployed language model with **three** red-team strategies: prompt *mutation*, *role-play* jailbreaks, and multi-turn *escalation*. Response: defects discovered per session.

Idea: run all pairwise two-sample t-tests, each at $\alpha = 0.05$.

Comparison	Test
Mutation vs. Role-play	t-test #1
Mutation vs. Escalation	t-test #2
Role-play vs. Escalation	t-test #3

Beyond Two Samples

Why not just repeat two-sample t-tests?

Scenario: an AI safety group probes a deployed language model with **three** red-team strategies: prompt *mutation*, *role-play* jailbreaks, and multi-turn *escalation*. Response: defects discovered per session.

Idea: run all pairwise two-sample t-tests, each at $\alpha = 0.05$.

Comparison	Test
Mutation vs. Role-play	t-test #1
Mutation vs. Escalation	t-test #2
Role-play vs. Escalation	t-test #3

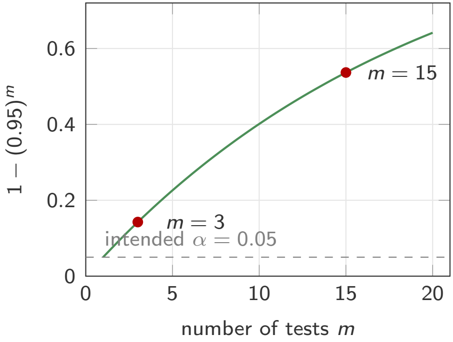
Problem — the family-wise error rate. With m independent tests, each at level α:

$$P(\text{at least one false rejection}) = 1 - (1 - \alpha)^m \tag{8.1}$$

$$m = 3 : \quad 1 - (0.95)^3 = 0.1426 \neq 0.05$$

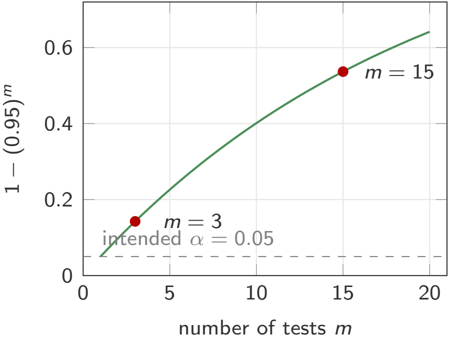
How Bad Does It Get?

Six strategies, fifteen pairs



How Bad Does It Get?

Six strategies, fifteen pairs

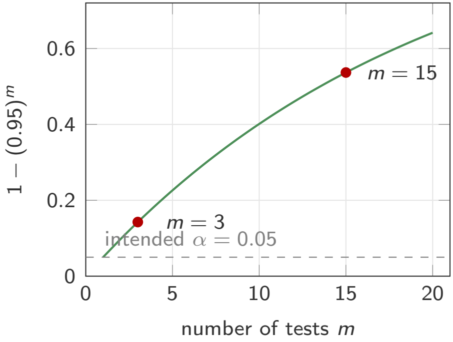


Step 1 — count the pairs. With $a = 6$ strategies:

$$m = \binom{6}{2} = 15$$

How Bad Does It Get?

Six strategies, fifteen pairs



Step 1 — count the pairs. With $a=6$ strategies:

$$m = \binom{6}{2} = 15$$

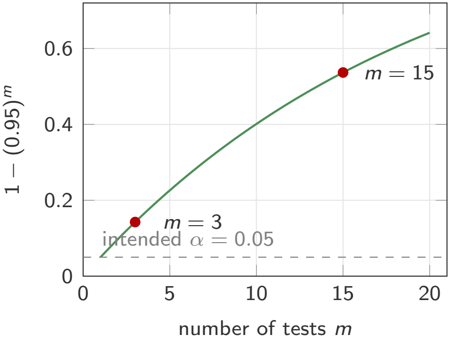
Step 2 — apply (8.1):

$$1-(0.95)^{15} = 1-0.4633 = 0.5367$$

A 54% chance of at least one false discovery.

How Bad Does It Get?

Six strategies, fifteen pairs



Step 1 — count the pairs. With $a = 6$ strategies:

$$m = \binom{6}{2} = 15$$

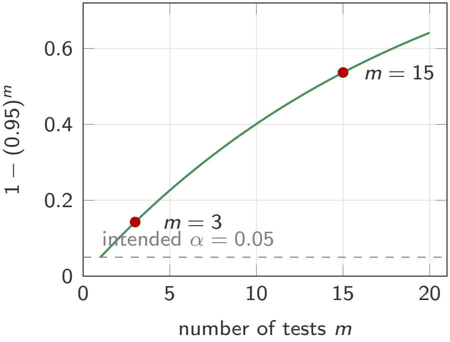
Step 2 — apply (8.1):

$$1 - (0.95)^{15} = 1 - 0.4633 = 0.5367$$

A 54% chance of at least one false discovery. **Step 3 — could we shrink α?** Keeping the family at 0.05 needs $\alpha^* \approx 0.0034$ per test — which destroys

How Bad Does It Get?

Six strategies, fifteen pairs



Step 1 — count the pairs. With $a = 6$ strategies:

$$m = \binom{6}{2} = 15$$

Step 2 — apply (8.1):

$$1 - (0.95)^{15} = 1 - 0.4633 = 0.5367$$

A 54% chance of at least one false discovery. **Step 3 — could we shrink α?** Keeping the family at 0.05 needs $\alpha^* \approx 0.0034$ per test — which destroys

The Vocabulary of Designed Experiments

§8.2

Four words index every symbol in this chapter:

- **Factor** — the variable we deliberately change (red-team strategy).
- **Levels (treatments)** — the a specific settings compared.
- **Replicates** — n independent runs per level; $N = an$ total.
- **Response** — the measured outcome (defects per session).

The Vocabulary of Designed Experiments

§8.2

Four words index every symbol in this chapter:

- **Factor** — the variable we deliberately change (red-team strategy).
- **Levels (treatments)** — the a specific settings compared.
- **Replicates** — n independent runs per level; $N = an$ total.
- **Response** — the measured outcome (defects per session).

Two principles make the comparison trustworthy:

- **Replication** — without it, no estimate of σ^2, no test.
- **Randomization** — random run order protects against unknown drift.

The Vocabulary of Designed Experiments

§8.2

Four words index every symbol in this chapter:

- **Factor** — the variable we deliberately change (red-team strategy).
- **Levels (treatments)** — the a specific settings compared.
- **Replicates** — n independent runs per level; $N = an$ total.
- **Response** — the measured outcome (defects per session).

Two principles make the comparison trustworthy:

- **Replication** — without it, no estimate of σ^2, no test.
- **Randomization** — random run order protects against unknown drift.

Replication + randomization, no blocking = the **completely randomized design (CRD)** — today's design. Blocking arrives in Lecture 24.

The Effects Model

§8.3

Setup: a treatments, n observations each ($N = an$). Each observation is an overall level, plus a treatment shift, plus noise:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, n \tag{8.2}$$

The Effects Model

§8.3

Setup: a treatments, n observations each ($N = an$). Each observation is an overall level, plus a treatment shift, plus noise:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, n \tag{8.2}$$

Symbol	Meaning
μ	Overall (grand) mean
τ_i	Effect of treatment i; $\sum_{i=1}^{a} \tau_i = 0$
ε_{ij}	Random error: $\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$

The Effects Model

§8.3

Setup: a treatments, n observations each $(N = an)$. Each observation is an overall level, plus a treatment shift, plus noise:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, n \tag{8.2}$$

Symbol	Meaning
μ	Overall (grand) mean
τ_i	Effect of treatment i; $\sum_{i=1}^{a} \tau_i = 0$
ε_{ij}	Random error: $\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$

Hypotheses: $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$ vs. $H_1 : \tau_i \neq 0$ for at least one i

The Effects Model

§8.3

Setup: a treatments, n observations each ($N = an$). Each observation is an overall level, plus a treatment shift, plus noise:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \quad j = 1, \ldots, n \tag{8.2}$$

Symbol	Meaning
μ	Overall (grand) mean
τ_i	Effect of treatment i; $\sum_{i=1}^{a} \tau_i = 0$
ε_{ij}	Random error: $\varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$

Hypotheses: $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \quad \text{vs.} \quad H_1 : \tau_i \neq 0$ for at least one i

Rejecting H_0 says *some* means differ — not *which* ones, and not that *all* differ. *Which* is

Model Assumptions

Three conditions carry the whole analysis — the same three as in regression:

Model Assumptions

Three conditions carry the whole analysis — the same three as in regression:

1. **Normality** — each treatment group follows a normal distribution.

Model Assumptions

Three conditions carry the whole analysis — the same three as in regression:

1. **Normality** — each treatment group follows a normal distribution.
2. **Equal variances** — $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 = \sigma^2$.

Model Assumptions

Three conditions carry the whole analysis — the same three as in regression:

1. **Normality** — each treatment group follows a normal distribution.
2. **Equal variances** — $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 = \sigma^2$.
3. **Independence** — earned by randomizing, not checked afterward.

Model Assumptions

Three conditions carry the whole analysis — the same three as in regression:

1. **Normality** — each treatment group follows a normal distribution.
2. **Equal variances** — $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 = \sigma^2$.
3. **Independence** — earned by randomizing, not checked afterward.

Natural estimates (8.3):

$$\hat{\mu} = \bar{y}_{\cdot\cdot}, \qquad \hat{\tau}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$$

Model Assumptions

Three conditions carry the whole analysis — the same three as in regression:

1. **Normality** — each treatment group follows a normal distribution.
2. **Equal variances** — $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2 = \sigma^2$.
3. **Independence** — earned by randomizing, not checked afterward.

Natural estimates (8.3):

$$\hat{\mu} = \bar{y}_{..}, \qquad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

Startup view: in an A/B/n test the treatments are product variants and the units are users. Random assignment is what makes independence believable; self-selected variants break it before any data arrive.

Notation and Totals

Dot notation (Table 8.1)

A dot replaces a subscript that has been summed over; a bar denotes an average.

Symbol	Formula	Meaning
y_{ij}	—	j-th obs. in treatment i
$y_{i\cdot}$	$\sum_{j=1}^{n} y_{ij}$	Total for treatment i
$\bar{y}_{i\cdot}$	$y_{i\cdot}/n$	Mean of treatment i
$y_{\cdot\cdot}$	$\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$	Grand total
$\bar{y}_{\cdot\cdot}$	$y_{\cdot\cdot}/N$	Grand mean

Notation and Totals

Dot notation (Table 8.1)

A dot replaces a subscript that has been summed over; a bar denotes an average.

Symbol	Formula	Meaning
y_{ij}	—	j-th obs. in treatment i
$y_{i\cdot}$	$\sum_{j=1}^{n} y_{ij}$	Total for treatment i
$\bar{y}_{i\cdot}$	$y_{i\cdot}/n$	Mean of treatment i
$y_{\cdot\cdot}$	$\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$	Grand total
$\bar{y}_{\cdot\cdot}$	$y_{\cdot\cdot}/N$	Grand mean

This notation does the bookkeeping — the formulas that follow are compact because of it. Give it one careful minute now.

Partitioning Total Variability

§8.4

The key identity (8.4) — every deviation splits in two:

$$\underbrace{y_{ij} - \bar{y}_{\cdot\cdot}}_{\text{total deviation}} = \underbrace{(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})}_{\text{treatment deviation}} + \underbrace{(y_{ij} - \bar{y}_{i\cdot})}_{\text{error deviation}}$$

Partitioning Total Variability

§8.4

The key identity (8.4) — every deviation splits in two:

$$\underbrace{y_{ij} - \bar{y}_{\cdot\cdot}}_{\text{total deviation}} = \underbrace{(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})}_{\text{treatment deviation}} + \underbrace{(y_{ij} - \bar{y}_{i\cdot})}_{\text{error deviation}}$$

Square, sum over all N observations; the cross-term vanishes:

$$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2}_{SS_T} = \underbrace{n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}_{SS_{\text{Treatments}}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2}_{SS_E} \qquad (8.5)$$

Partitioning Total Variability

§8.4

The key identity (8.4) — every deviation splits in two:

$$\underbrace{y_{ij} - \bar{y}_{\cdot\cdot}}_{\text{total deviation}} \quad = \quad \underbrace{(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})}_{\text{treatment deviation}} \quad + \quad \underbrace{(y_{ij} - \bar{y}_{i\cdot})}_{\text{error deviation}}$$

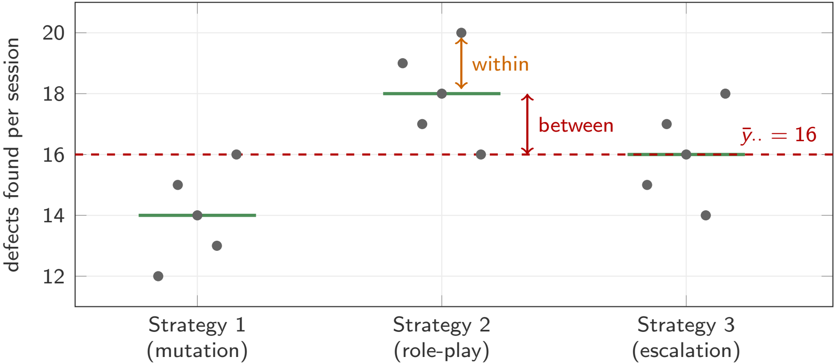
Square, sum over all N observations; the cross-term vanishes:

$$\underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2}_{SS_T} = \underbrace{n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}_{SS_{\text{Treatments}}} + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2}_{SS_E} \qquad (8.5)$$

Sum of Squares	Measures	df

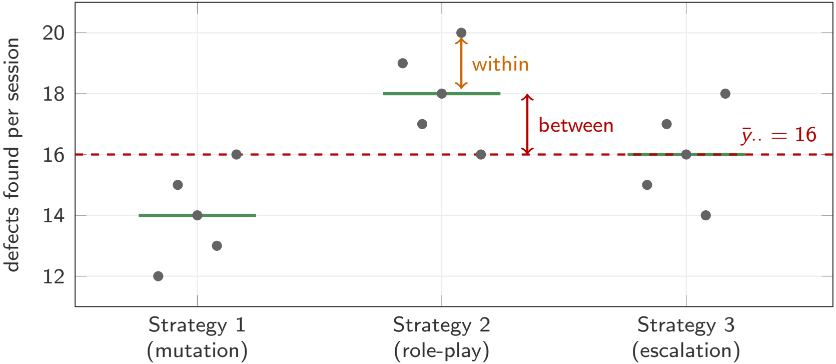
Between vs. Within, in One Picture

Figure 8.2



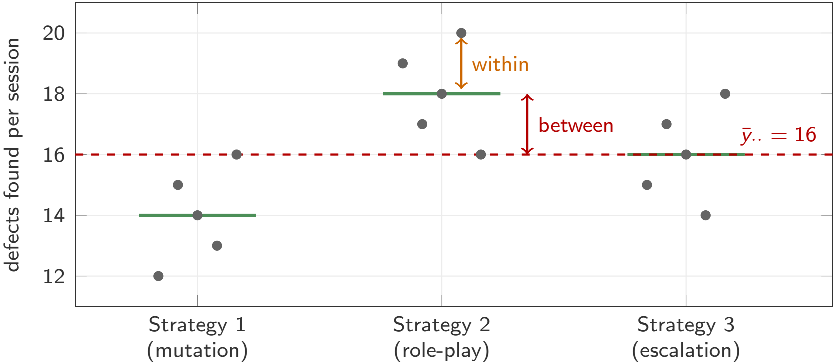
Between vs. Within, in One Picture

Figure 8.2



Between vs. Within, in One Picture

Figure 8.2



$SS_{\text{Treatments}}$ measures the green segments' spread around the dashed grand mean; SS_E measures the points' spread around their own green segment.

Shortcut Forms for Hand Computation

On the formula sheet

Expanding the squares gives forms that use only totals — no individual deviations:

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{8.6}$$

Shortcut Forms for Hand Computation

On the formula sheet

Expanding the squares gives forms that use only totals — no individual deviations:

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \tag{8.6}$$

$$SS_{\text{Treatments}} = \frac{1}{n} \sum_{i=1}^{a} y_{i\cdot}^2 - \frac{y_{\cdot\cdot}^2}{N}, \qquad SS_E = SS_T - SS_{\text{Treatments}} \tag{8.7}$$

Shortcut Forms for Hand Computation

On the formula sheet

Expanding the squares gives forms that use only totals — no individual deviations:

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{8.6}$$

$$SS_{\text{Treatments}} = \frac{1}{n} \sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}, \qquad SS_E = SS_T - SS_{\text{Treatments}} \tag{8.7}$$

Workflow: compute SS_T and $SS_{\text{Treatments}}$ from totals, then get SS_E *by subtraction*.

Shortcut Forms for Hand Computation

On the formula sheet

Expanding the squares gives forms that use only totals — no individual deviations:

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{8.6}$$

$$SS_{\text{Treatments}} = \frac{1}{n} \sum_{i=1}^{a} y_{i\cdot}^2 - \frac{y_{..}^2}{N}, \qquad SS_E = SS_T - SS_{\text{Treatments}} \tag{8.7}$$

Workflow: compute SS_T and $SS_{\text{Treatments}}$ from totals, then get SS_E *by subtraction*.

The correction term $y_{..}^2/N$ appears in both formulas — compute it once, use it twice.

Mean Squares

$SS \div df$

Sums of squares become comparable only after dividing by degrees of freedom:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1} \tag{8.8}$$

Mean Squares

$SS \div df$

Sums of squares become comparable only after dividing by degrees of freedom:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1} \tag{8.8}$$

$$MS_E = \frac{SS_E}{N-a} \tag{8.9}$$

Mean Squares

$SS \div df$

Sums of squares become comparable only after dividing by degrees of freedom:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1} \tag{8.8}$$

$$MS_E = \frac{SS_E}{N-a} \tag{8.9}$$

MS_E is the pooled average of the a group variances — the direct generalization of s_p^2 from the two-sample t-test. It estimates σ^2 whether or not H_0 is true.

Mean Squares

$SS \div df$

Sums of squares become comparable only after dividing by degrees of freedom:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1} \tag{8.8}$$

$$MS_E = \frac{SS_E}{N-a} \tag{8.9}$$

MS_E is the pooled average of the a group variances — the direct generalization of s_p^2 from the two-sample t-test. It estimates σ^2 whether or not H_0 is true.

Expected mean squares (8.10):

$$E[MS_E] = \sigma^2, \qquad E[MS_{\text{Treatments}}] = \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}$$

The F-Test

§8.5

Under H_0 both mean squares estimate the same σ^2, so their ratio follows an F distribution:

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}, \qquad F_0 \sim F_{a-1,\, N-a} \text{ under } H_0 \tag{8.11}$$

The F-Test

§8.5

Under H_0 both mean squares estimate the same σ^2, so their ratio follows an F distribution:

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}, \qquad F_0 \sim F_{a-1,\, N-a} \text{ under } H_0 \tag{8.11}$$

Intuition:

- Under H_0: F_0 hovers around 1.
- Real effects push F_0 above 1 — between-group variation exceeds within-group variation.

The F-Test

§8.5

Under H_0 both mean squares estimate the same σ^2, so their ratio follows an F distribution:

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}, \qquad F_0 \sim F_{a-1,\, N-a} \text{ under } H_0 \tag{8.11}$$

Intuition:

- Under H_0: F_0 hovers around 1.
- Real effects push F_0 above 1 — between-group variation exceeds within-group variation.

One-tailed to the right, even though H_1 is "means differ in any direction": effects of any sign inflate the numerator, so only the upper tail signals them.

The F-Test

§8.5

Under H_0 both mean squares estimate the same σ^2, so their ratio follows an F distribution:

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}, \qquad F_0 \sim F_{a-1,\, N-a} \text{ under } H_0 \tag{8.11}$$

Intuition:

- Under H_0: F_0 hovers around 1.
- Real effects push F_0 above 1 — between-group variation exceeds within-group variation.

One-tailed to the right, even though H_1 is "means differ in any direction": effects of any sign inflate the numerator, so only the upper tail signals them.

Reject H_0 if $F_0 > f_{\alpha, a-1, N-a}$, equivalently if p-value $< \alpha$

The ANOVA Table

Table 8.2 and Procedure 8.1

Source	**SS**	**df**	**MS**	F_0
Treatments	SS_{Trt}	$a-1$	$MS_{\text{Trt}} = \dfrac{SS_{\text{Trt}}}{a-1}$	$\dfrac{MS_{\text{Trt}}}{MS_E}$
Error	SS_E	$N-a$	$MS_E = \dfrac{SS_E}{N-a}$	
Total	SS_T	$N-1$		

The ANOVA Table

Table 8.2 and Procedure 8.1

Source	SS	df	MS	F_0
Treatments	SS_{Trt}	$a-1$	$MS_{\text{Trt}} = \dfrac{SS_{\text{Trt}}}{a-1}$	$\dfrac{MS_{\text{Trt}}}{MS_E}$
Error	SS_E	$N-a$	$MS_E = \dfrac{SS_E}{N-a}$	
Total	SS_T	$N-1$		

Procedure 8.1 builds it in six steps: totals $\to$ SS_T (8.6) $\to$ SS_{Trt}, SS_E (8.7) $\to$ df $\to$ MS (8.8), (8.9) and F_0 (8.11) $\to$ decision.

The ANOVA Table

Table 8.2 and Procedure 8.1

Source	SS	df	MS	F_0
Treatments	SS_{Trt}	$a-1$	$MS_{\text{Trt}} = \dfrac{SS_{\text{Trt}}}{a-1}$	$\dfrac{MS_{\text{Trt}}}{MS_E}$
Error	SS_E	$N-a$	$MS_E = \dfrac{SS_E}{N-a}$	
Total	SS_T	$N-1$		

Procedure 8.1 builds it in six steps: totals $\to$ SS_T (8.6) $\to$ SS_{Trt}, SS_E (8.7) $\to$ df $\to$ MS (8.8), (8.9) and F_0 (8.11) $\to$ decision.

Built-in checks: both the SS column and the df column must add down to the Total row. Verify both before reading off F_0.

Worked Example: Three Red-Team Strategies

Table 8.3

Each strategy runs in $n = 5$ independent four-hour sessions, in randomized order. Response: distinct safety defects discovered per session.

Worked Example: Three Red-Team Strategies

Table 8.3

Each strategy runs in $n = 5$ independent four-hour sessions, in randomized order. Response: distinct safety defects discovered per session.

Strategy	**Defects per session**					$y_{i\cdot}$	$\bar{y}_{i\cdot}$
1. Mutation	12	15	14	13	16	70	14.0
2. Role-play	19	17	18	20	16	90	18.0
3. Escalation	15	17	16	14	18	80	16.0
					Grand total $y_{\cdot\cdot}$	240	
					Grand mean $\bar{y}_{\cdot\cdot}$		16.0

Worked Example: Three Red-Team Strategies

Table 8.3

Each strategy runs in $n = 5$ independent four-hour sessions, in randomized order. Response: distinct safety defects discovered per session.

Strategy	**Defects per session**					$y_{i\cdot}$	$\bar{y}_{i\cdot}$
1. Mutation	12	15	14	13	16	70	14.0
2. Role-play	19	17	18	20	16	90	18.0
3. Escalation	15	17	16	14	18	80	16.0
					Grand total $y_{\cdot\cdot}$	240	
					Grand mean $\bar{y}_{\cdot\cdot}$		16.0

$a = 3$, $n = 5$, $N = 15$. Test at $\alpha = 0.05$: **do the strategies differ in mean defect discovery?**

Worked Example: Step 1–2, Setup and SS_T

Following Procedure 8.1

Step 1 — hypotheses and totals. $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_1: at least one pair differs. Totals: $y_{1.} = 70$, $y_{2.} = 90$, $y_{3.} = 80$, $y_{..} = 240$.

Worked Example: Step 1–2, Setup and SS_T

Following Procedure 8.1

Step 1 — hypotheses and totals. $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_1: at least one pair differs. Totals: $y_{1\cdot} = 70$, $y_{2\cdot} = 90$, $y_{3\cdot} = 80$, $y_{\cdot\cdot} = 240$.

Step 2 — SS_T by the shortcut (8.6). First the two ingredients:

$$\sum_i \sum_j y_{ij}^2 = 990 + 1630 + 1290 = 3910$$

Worked Example: Step 1–2, Setup and SS_T

Following Procedure 8.1

Step 1 — hypotheses and totals. $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_1: at least one pair differs. Totals: $y_{1\cdot} = 70$, $y_{2\cdot} = 90$, $y_{3\cdot} = 80$, $y_{\cdot\cdot} = 240$.

Step 2 — SS_T by the shortcut (8.6). First the two ingredients:

$$\sum_i \sum_j y_{ij}^2 = 990 + 1630 + 1290 = 3910$$

$$\frac{y_{\cdot\cdot}^2}{N} = \frac{240^2}{15} = \frac{57600}{15} = 3840$$

Worked Example: Step 1–2, Setup and SS_T

Following Procedure 8.1

Step 1 — hypotheses and totals. $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_1: at least one pair differs. Totals: $y_{1\cdot} = 70$, $y_{2\cdot} = 90$, $y_{3\cdot} = 80$, $y_{\cdot\cdot} = 240$.

Step 2 — SS_T by the shortcut (8.6). First the two ingredients:

$$\sum_i \sum_j y_{ij}^2 = 990 + 1630 + 1290 = 3910$$

$$\frac{y_{\cdot\cdot}^2}{N} = \frac{240^2}{15} = \frac{57600}{15} = 3840$$

Then subtract:

$$SS_T = 3910 - 3840 = \boxed{70.00}$$

Worked Example: Step 3, $SS_{\text{Treatments}}$ and SS_E

Treatment SS by the shortcut (8.7):

$$SS_{\text{Trt}} = \frac{70^2 + 90^2 + 80^2}{5} - 3840 = \frac{4900 + 8100 + 6400}{5} - 3840$$

Worked Example: Step 3, $SS_{\text{Treatments}}$ and SS_E

Treatment SS by the shortcut (8.7):

$$SS_{\text{Trt}} = \frac{70^2 + 90^2 + 80^2}{5} - 3840 = \frac{4900 + 8100 + 6400}{5} - 3840$$

$$= \frac{19400}{5} - 3840 = 3880 - 3840 = \boxed{40.00}$$

Worked Example: Step 3, $SS_{\text{Treatments}}$ and SS_E

Treatment SS by the shortcut (8.7):

$$SS_{\text{Trt}} = \frac{70^2 + 90^2 + 80^2}{5} - 3840 = \frac{4900 + 8100 + 6400}{5} - 3840$$

$$= \frac{19400}{5} - 3840 = 3880 - 3840 = \boxed{40.00}$$

Check with the definition form:

$$5\left[(14-16)^2 + (18-16)^2 + (16-16)^2\right] = 5(4+4+0) = 40 \checkmark$$

Worked Example: Step 3, $SS_{\text{Treatments}}$ and SS_E

Treatment SS by the shortcut (8.7):

$$SS_{\text{Trt}} = \frac{70^2 + 90^2 + 80^2}{5} - 3840 = \frac{4900 + 8100 + 6400}{5} - 3840$$

$$= \frac{19400}{5} - 3840 = 3880 - 3840 = \boxed{40.00}$$

Check with the definition form:

$$5\left[(14-16)^2 + (18-16)^2 + (16-16)^2\right] = 5(4+4+0) = 40 \checkmark$$

Error SS by subtraction (8.5):

Worked Example: Step 4–5, Mean Squares and F_0

Step 4 — degrees of freedom:

$$a - 1 = 2, \qquad N - a = 12, \qquad N - 1 = 14 \qquad (2 + 12 = 14\ \checkmark)$$

Worked Example: Step 4–5, Mean Squares and F_0

Step 4 — degrees of freedom:

$$a - 1 = 2, \qquad N - a = 12, \qquad N - 1 = 14 \qquad (2 + 12 = 14 \checkmark)$$

Mean squares from (8.8) and (8.9):

$$MS_{\text{Trt}} = \frac{40}{2} = 20.00, \qquad MS_E = \frac{30}{12} = 2.500$$

Worked Example: Step 4–5, Mean Squares and F_0

Step 4 — degrees of freedom:

$$a - 1 = 2, \qquad N - a = 12, \qquad N - 1 = 14 \qquad (2 + 12 = 14\ \checkmark)$$

Mean squares from (8.8) and (8.9):

$$MS_{\mathrm{Trt}} = \frac{40}{2} = 20.00, \qquad MS_E = \frac{30}{12} = 2.500$$

Step 5 — test statistic from (8.11):

$$F_0 = \frac{20.00}{2.500} = \boxed{8.00}$$

Worked Example: Step 4–5, Mean Squares and F_0

Step 4 — degrees of freedom:

$$a - 1 = 2, \qquad N - a = 12, \qquad N - 1 = 14 \qquad (2 + 12 = 14\ \checkmark)$$

Mean squares from (8.8) and (8.9):

$$MS_{\mathrm{Trt}} = \frac{40}{2} = 20.00, \qquad MS_E = \frac{30}{12} = 2.500$$

Step 5 — test statistic from (8.11):

$$F_0 = \frac{20.00}{2.500} = \boxed{8.00}$$

Worked Example: Step 6, Decision and Conclusion

Using Table 8.4

F table, upper 5% points:

$\nu_2 \backslash \nu_1$	1	2	3
11	4.84	3.98	3.59
12	4.75	3.89	3.49
13	4.67	3.81	3.41

Completed table (Table 8.5):

Source	SS	df	MS	F_0
Strategies	40.00	2	20.00	8.00
Error	30.00	12	2.500	
Total	70.00	14		

Worked Example: Step 6, Decision and Conclusion

Using Table 8.4

F table, upper 5% points:

$\nu_2 \backslash \nu_1$	1	2	3
11	4.84	3.98	3.59
12	4.75	3.89	3.49
13	4.67	3.81	3.41

Completed table (Table 8.5):

Source	SS	df	MS	F_0
Strategies	40.00	2	20.00	8.00
Error	30.00	12	2.500	
Total	70.00	14		

Decision: $f_{0.05,\,2,\,12} = 3.89$; since $F_0 = 8.00 > 3.89$, **reject** H_0 $\quad (p \approx 0.006)$.

Worked Example: Step 6, Decision and Conclusion

Using Table 8.4

F table, upper 5% points:

$\nu_2 \backslash \nu_1$	1	2	3
11	4.84	3.98	3.59
12	4.75	3.89	3.49
13	4.67	3.81	3.41

Completed table (Table 8.5):

Source	SS	df	MS	F_0
Strategies	40.00	2	20.00	8.00
Error	30.00	12	2.500	
Total	70.00	14		

Decision: $f_{0.05,2,12} = 3.89$; since $F_0 = 8.00 > 3.89$, **reject** H_0 $(p \approx 0.006)$.

Conclusion: mean defect discovery differs across strategies. Role-play found the most (18.0 per session).

Worked Example: Step 6, Decision and Conclusion

Using Table 8.4

F table, upper 5% points:

$\nu_2 \backslash \nu_1$	1	2	3
11	4.84	3.98	3.59
12	4.75	3.89	3.49
13	4.67	3.81	3.41

Completed table (Table 8.5):

Source	SS	df	MS	F_0
Strategies	40.00	2	20.00	8.00
Error	30.00	12	2.500	
Total	70.00	14		

Decision: $f_{0.05,2,12} = 3.89$; since $F_0 = 8.00 > 3.89$, **reject** H_0 $(p \approx 0.006)$.

Conclusion: mean defect discovery differs across strategies. Role-play found the most (18.0 per session).

What next? The F-test says *some* means differ, not *which* ones. Multiple comparisons (Lecture 23) answer that.

After Rejecting: Effects and Intervals from MS_E

§8.6

Point estimates (8.3): $\hat{\mu} = 16.0$ and

$$\hat{\tau}_1 = 14 - 16 = -2.0, \quad \hat{\tau}_2 = 18 - 16 = +2.0, \quad \hat{\tau}_3 = 0.0 \quad (\text{sum} = 0 \checkmark)$$

After Rejecting: Effects and Intervals from MS_E

§8.6

Point estimates (8.3): $\hat{\mu} = 16.0$ and

$$\hat{\tau}_1 = 14 - 16 = -2.0, \quad \hat{\tau}_2 = 18 - 16 = +2.0, \quad \hat{\tau}_3 = 0.0 \quad (\text{sum} = 0 \checkmark)$$

CI for one mean (8.12) — MS_E plays the role of s^2, with $N - a = 12$ df. For role-play, with $t_{0.025,12} = 2.179$:

$$18.0 \pm 2.179\sqrt{\tfrac{2.5}{5}} = 18.0 \pm 2.179(0.7071) = 18.0 \pm 1.541 \;\Rightarrow\; (16.46,\ 19.54)$$

After Rejecting: Effects and Intervals from MS_E

§8.6

Point estimates (8.3): $\hat{\mu} = 16.0$ and

$$\hat{\tau}_1 = 14 - 16 = -2.0, \quad \hat{\tau}_2 = 18 - 16 = +2.0, \quad \hat{\tau}_3 = 0.0 \quad (\text{sum} = 0\ \checkmark)$$

CI for one mean (8.12) — MS_E plays the role of s^2, with $N - a = 12$ df. For role-play, with $t_{0.025,12} = 2.179$:

$$18.0 \pm 2.179\sqrt{\frac{2.5}{5}} = 18.0 \pm 2.179(0.7071) = 18.0 \pm 1.541 \ \Rightarrow\ (16.46,\ 19.54)$$

CI for a difference (8.13) — role-play minus mutation:

$$(18 - 14) \pm 2.179\sqrt{\frac{2(2.5)}{5}} = 4.0 \pm 2.179(1.000) \ \Rightarrow\ (1.82,\ 6.18)$$

After Rejecting: Effects and Intervals from MS_E

§8.6

Point estimates (8.3): $\hat{\mu} = 16.0$ and

$$\hat{\tau}_1 = 14 - 16 = -2.0, \quad \hat{\tau}_2 = 18 - 16 = +2.0, \quad \hat{\tau}_3 = 0.0 \quad (\text{sum} = 0 \checkmark)$$

CI for one mean (8.12) — MS_E plays the role of s^2, with $N - a = 12$ df. For role-play, with $t_{0.025,12} = 2.179$:

$$18.0 \pm 2.179\sqrt{\frac{2.5}{5}} = 18.0 \pm 2.179(0.7071) = 18.0 \pm 1.541 \;\Rightarrow\; (16.46,\ 19.54)$$

CI for a difference (8.13) — role-play minus mutation:

$$(18 - 14) \pm 2.179\sqrt{\frac{2(2.5)}{5}} = 4.0 \pm 2.179(1.000) \;\Rightarrow\; (1.82,\ 6.18)$$

Quick Check: Reconstructing an ANOVA Table

Supply chain

Q. A freight broker compares $a = 4$ carriers with $n = 6$ observed deliveries each. The analyst reports $SS_T = 280$ and $SS_{\text{Treatments}} = 120$, but the rest of the output is lost. Complete the test at $\alpha = 0.05$, given $f_{0.05,\,3,\,20} = 3.10$.

Quick Check: Reconstructing an ANOVA Table

Supply chain

Q. A freight broker compares $a = 4$ carriers with $n = 6$ observed deliveries each. The analyst reports $SS_T = 280$ and $SS_{\text{Treatments}} = 120$, but the rest of the output is lost. Complete the test at $\alpha = 0.05$, given $f_{0.05,\,3,\,20} = 3.10$.

df: $N = 24$, so $a - 1 = 3, \quad N - a = 20, \quad N - 1 = 23.$

Quick Check: Reconstructing an ANOVA Table

Supply chain

Q. A freight broker compares $a = 4$ carriers with $n = 6$ observed deliveries each. The analyst reports $SS_T = 280$ and $SS_{\text{Treatments}} = 120$, but the rest of the output is lost. Complete the test at $\alpha = 0.05$, given $f_{0.05,3,20} = 3.10$.

df: $N = 24$, so $a - 1 = 3, \quad N - a = 20, \quad N - 1 = 23.$

SS: by (8.5), $SS_E = 280 - 120 = 160.$

Quick Check: Reconstructing an ANOVA Table

Supply chain

Q. A freight broker compares $a = 4$ carriers with $n = 6$ observed deliveries each. The analyst reports $SS_T = 280$ and $SS_{\text{Treatments}} = 120$, but the rest of the output is lost. Complete the test at $\alpha = 0.05$, given $f_{0.05,3,20} = 3.10$.

df: $N = 24$, so $a - 1 = 3, \quad N - a = 20, \quad N - 1 = 23.$

SS: by (8.5), $SS_E = 280 - 120 = 160.$

MS and F_0: $MS_{\text{Trt}} = 120/3 = 40.00, \quad MS_E = 160/20 = 8.000,$

$$F_0 = \frac{40.00}{8.000} = 5.000$$

Quick Check: Reconstructing an ANOVA Table

Supply chain

Q. A freight broker compares $a = 4$ carriers with $n = 6$ observed deliveries each. The analyst reports $SS_T = 280$ and $SS_{\text{Treatments}} = 120$, but the rest of the output is lost. Complete the test at $\alpha = 0.05$, given $f_{0.05,3,20} = 3.10$.

df: $N = 24$, so $a - 1 = 3, \quad N - a = 20, \quad N - 1 = 23.$

SS: by (8.5), $SS_E = 280 - 120 = 160.$

MS and F_0: $MS_{\text{Trt}} = 120/3 = 40.00, \quad MS_E = 160/20 = 8.000,$

$$F_0 = \frac{40.00}{8.000} = 5.000$$

$F_0 = 5.000 > 3.10 \Rightarrow$ **reject** H_0: the carriers' mean delivery performances differ. The whole table was recoverable from two numbers plus the decomposition.

Quick Check: Family-Wise Error

Startup A/B/n testing

Q. Your startup's dashboard compares $a = 4$ pricing-page variants by flagging every pairwise gap with a separate test at $\alpha = 0.05$. If all four variants convert identically, the chance the dashboard flags at least one "winner" is closest to:

(A) 5%, because each test runs at $\alpha = 0.05$.

(B) 10%, because there are two independent comparisons.

(C) 26%, because there are $\binom{4}{2} = 6$ pairs and $1 - (0.95)^6 \approx 0.26$.

(D) 100%, because some pair always differs numerically.

Quick Check: Family-Wise Error

Startup A/B/n testing

Q. Your startup's dashboard compares $a = 4$ pricing-page variants by flagging every pairwise gap with a separate test at $\alpha = 0.05$. If all four variants convert identically, the chance the dashboard flags at least one "winner" is closest to:

(A) 5%, because each test runs at $\alpha = 0.05$.

(B) 10%, because there are two independent comparisons.

(C) 26%, because there are $\binom{4}{2} = 6$ pairs and $1 - (0.95)^6 \approx 0.26$.

(D) 100%, because some pair always differs numerically.

Answer: (C). By (8.1) with $m = 6$: $1 - (0.95)^6 = 0.2649$ — roughly one chance in four, not one in twenty. (A) is the per-test rate, not the family rate; (D) confuses a numerical gap with a significant one. One ANOVA F-test of all four variants keeps the level at an honest 5%.

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1 - (1 - \alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1-(1-\alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.
- The effects model (8.2): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, errors iid $N(0, \sigma^2)$.

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1 - (1 - \alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.
- The effects model (8.2): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, errors iid $N(0, \sigma^2)$.
- Total variability partitions (8.5): $SS_T = SS_{\text{Treatments}} + SS_E$.

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1 - (1 - \alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.
- The effects model (8.2): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, errors iid $N(0, \sigma^2)$.
- Total variability partitions (8.5): $SS_T = SS_{\text{Treatments}} + SS_E$.
- Mean squares (8.8), (8.9); $F_0 = MS_{\text{Trt}}/MS_E \sim F_{a-1, N-a}$ under H_0 (8.11).

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1-(1-\alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.
- The effects model (8.2): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, errors iid $N(0, \sigma^2)$.
- Total variability partitions (8.5): $SS_T = SS_{\text{Treatments}} + SS_E$.
- Mean squares (8.8), (8.9); $F_0 = MS_{\text{Trt}}/MS_E \sim F_{a-1,\,N-a}$ under H_0 (8.11).
- Procedure 8.1 builds the table by hand; SS and df columns must both add to the Total row.

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1-(1-\alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.
- The effects model (8.2): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, errors iid $N(0, \sigma^2)$.
- Total variability partitions (8.5): $SS_T = SS_{\text{Treatments}} + SS_E$.
- Mean squares (8.8), (8.9); $F_0 = MS_{\text{Trt}}/MS_E \sim F_{a-1,\,N-a}$ under H_0 (8.11).
- Procedure 8.1 builds the table by hand; SS and df columns must both add to the Total row.
- MS_E then powers CIs for means (8.12) and differences (8.13).

Lecture 22 Summary

- Repeated pairwise t-tests inflate the family-wise error to $1 - (1 - \alpha)^m$ (8.1); ANOVA tests all a means at once at an honest α.
- The effects model (8.2): $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, errors iid $N(0, \sigma^2)$.
- Total variability partitions (8.5): $SS_T = SS_{\text{Treatments}} + SS_E$.
- Mean squares (8.8), (8.9); $F_0 = MS_{\text{Trt}}/MS_E \sim F_{a-1, N-a}$ under H_0 (8.11).
- Procedure 8.1 builds the table by hand; SS and df columns must both add to the Total row.
- MS_E then powers CIs for means (8.12) and differences (8.13).
- **Next lecture:** which means differ — Fisher's LSD and Tukey's HSD.

Reminders

- **Quiz on Chapter 8 basics** is due before class — it covers today's ANOVA table construction.

Reminders

- **Quiz on Chapter 8 basics** is due before class — it covers today's ANOVA table construction.
- Homework 7 (single-factor experiments) is posted on Blackboard.

Reminders

- **Quiz on Chapter 8 basics** is due before class — it covers today's ANOVA table construction.
- Homework 7 (single-factor experiments) is posted on Blackboard.
- Office hours: Tuesdays 9–10 AM, Room 22-219 or Zoom.

Reminders

- **Quiz on Chapter 8 basics** is due before class — it covers today's ANOVA table construction.
- Homework 7 (single-factor experiments) is posted on Blackboard.
- Office hours: Tuesdays 9–10 AM, Room 22-219 or Zoom.
- Read §8.7 (multiple comparisons) to prepare for the next lecture.