Chapter 11: Analysis of Variance

Lecture slides: PDF pages 1–2 (open)

Probability and Statistics for the Life Sciences **November 2026**

Suppose we compare a measurement across three treatment groups, rather than just two. The scientific question still concerns *population means*: are their long-run average responses the same? **Analysis of variance**, abbreviated **ANOVA**, answers that question by comparing two kinds of variability. The name describes the method, not the parameter being tested.

These notes can be read without the slides open. They retain the lecture examples, explain their arithmetic, and add intuition, assumptions, checks, and supplementary examples. Read a section, work through its calculations, then try its check before reading the reasoning.

The central idea

Sample means will almost never be exactly equal, even if the population means are equal. We ask whether their separation is large *relative to the ordinary variation among individual observations*. A difference of 5 units is compelling when individuals vary by about 1 unit; it may be unremarkable when they vary by about 50 units.

$$\boxed{\text{ANOVA test statistic} = \frac{\text{between-group variation per degree of freedom}}{\text{within-group variation per degree of freedom}}}$$

We will define every part of this ratio before using it. The second part of the chapter introduces **blocking**: accounting for a known source of background variation before comparing treatments.

Reading map

Topic	These notes	Slide PDF
Question, design vocabulary, multiple tests	2–3	2–9
Notation, assumptions, sums of squares	4–7	10–21, 25
Effects model and scenario comparison	8–9	22–24
Filling an incomplete ANOVA table	10	26–27
Calcium-intake example, every calculation	11–13	28–32
Individual and pairwise confidence intervals	14–15	34–36
Unequal group sizes and software workflow	16–17	12–21, 33
Randomized blocks: ideas, model and formulas	18–20	37–41
Cell-phone cost example, fully worked	21–24	42–47
Interaction, design cautions, final checklist	25–26	48–49

Headings link to the 49-page lecture handout. Numbers refer to PDF viewer pages. If a viewer ignores a page link, enter that page manually or use the section index in the footer. These are Chapter 11 notes, not a solution key to a current homework assignment.

What exactly are we comparing?

Lecture slides: PDF pages 2, 5–9 (open)

Population, sample, mean, and response

A **population** is the full collection or process we want to learn about. A **sample** is the set of observations we actually obtain. A **mean** is an average. A population mean is a fixed but usually unknown number; a sample mean is a number calculated from the observed data.

A **response variable** is the measured outcome. In the lecture's breakfast study, the response is an aptitude-test score. The participants are **experimental units**: the entities to which a treatment is assigned. A **factor** is an explanatory characteristic whose levels we compare. A **level** is a particular setting or category of that factor.

Term	Breakfast study
Experimental unit	One participant
Factor	Breakfast amount
Levels	No breakfast, partial breakfast, full breakfast
Response	Aptitude-test score

With one factor, a **treatment** is one factor level. With several factors, a treatment is a specified combination of levels. If age has two categories and breakfast has three, there are $2 \times 3 = 6$ combinations, such as younger participants receiving a full breakfast. The six combinations are:

	No breakfast	Partial breakfast	Full breakfast
Younger	Younger/no	Younger/partial	Younger/full
Older	Older/no	Older/partial	Older/full

We first study **one-way ANOVA**, meaning one factor, not one group.

State the target before the hypothesis

Let k be the number of groups. For group i, where $i = 1, \ldots, k$, let μ_i be the population mean response. For breakfast, $k = 3$, and μ_1, μ_2, μ_3 are the population mean scores for the three breakfast conditions.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k, \qquad H_a : \text{at least one pair of population means differs.}$$

Here H_0 is the **null hypothesis**, the common-mean claim being tested. H_a is the **alternative hypothesis**, which permits any departure from that claim. It does *not* require every mean to differ from every other mean.

What a significant result tells us. It supplies evidence against the all-equal claim. It does not identify the differing pair, establish the size of a meaningful effect, or automatically prove a causal effect. Causal interpretation depends on the study design, especially random treatment assignment and control of other differences.

Check your understanding. If $\mu_1 = \mu_2$ but μ_3 differs, is the alternative true?

Reasoning. Yes. One unequal pair is enough to contradict the claim that all means are equal.

Why not run every possible two-sample test?

Lecture slides: PDF pages 3–4 (open)

A **Type I error** is rejection of a true null hypothesis. The **significance level**, written α, is the probability limit chosen in advance for this error in a test. For example, $\alpha = 0.05$ means a 5% Type I error rate under the specified null model.

For three groups, the three possible pairs are $(1, 2)$, $(1, 3)$, and $(2, 3)$. Suppose we test each pair at level 0.05 and declare an overall difference if *any* test rejects. This gives us several opportunities to obtain a false positive. A 5% error rate for *each* test is not automatically a 5% error rate for the *whole collection*.

A calculation that shows the problem

Write $P(\cdot)$ for the probability of the event in parentheses. For the moment, imagine three *independent* tests, all with true null hypotheses and exact Type I error probability α. The probability that one test does not reject is $1 - \alpha$. Independence allows us to multiply probabilities:

$$P(\text{no false positives}) = (1-\alpha)(1-\alpha)(1-\alpha) = (1-\alpha)^3,$$
$$P(\text{at least one false positive}) = 1 - (1-\alpha)^3.$$

At $\alpha = 0.05$, this gives

$$1 - (0.95)^3 = 1 - 0.857375 = 0.142625,$$

about 14.3%, not 5%. With k groups, the number of pairs is

$$m = \binom{k}{2} = \frac{k(k-1)}{2}.$$

The symbol m denotes the number of comparisons. Under independent tests, the corresponding error probability is $1 - (1-\alpha)^m$.

Important qualification to the lecture calculation. Actual pairwise tests reuse the same groups, so they are generally *dependent*. The independent-test formula is an illustration, not an exact formula for the overlapping pairwise tests. The underlying warning remains: repeated unadjusted tests do not control the overall error rate at the desired level.

What ANOVA fixes, and what it does not

ANOVA performs *one overall test* of the all-equal hypothesis at the chosen level. If we subsequently examine many pairs, that is another multiple-comparison problem. Later we distinguish an individual confidence interval from a family of intervals and explain a simple adjustment.

Check your understanding. Does a significant ANOVA make every subsequent 95% pairwise interval simultaneously 95% reliable?

Reasoning. No. Overall testing and simultaneous coverage of many intervals are different guarantees. Multiple pairwise claims need an appropriate adjustment.

Organize the observations before calculating

Lecture slides: PDF pages 12–13, 25 (open)

Define each symbol

We have k independent groups. Group i contains n_i observations. The index i identifies a group; the index j identifies an observation within that group.

Symbol	Meaning
x_{ij}	Observed value for observation j in group i.
n_i	Number of observations in group i.
$N = \sum_{i=1}^{k} n_i$	Total number of observations across all groups.
$\bar{x}_i$	Sample mean within group i.
s_i^2	Sample variance within group i; squared response units.
$\bar{x}$	Grand mean: average of all N observations.
μ_i	Unknown population mean for group i.
σ^2	Common population variance assumed for all groups.

The summation symbol $\sum$ means "add the specified terms." For example, $\sum_{j=1}^{n_i} x_{ij}$ means add all n_i observations in group i.

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}, \qquad s_i^2 = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{n_i - 1}.$$

Compute a group mean first. Subtract it from each observation, square the differences, add them, then divide by $n_i - 1$ to obtain its sample variance.

The grand mean is a *weighted* average of the group means:

$$\bar{x} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}}{N} = \frac{n_1 \bar{x}_1 + \cdots + n_k \bar{x}_k}{N}.$$

A group with twice as many observations contributes twice as much information to this average. Only when all group sizes are equal can we simply average the group means.

Assumptions for the exact ANOVA test

1. Observations are independent within and across groups, supported by the sampling or experimental design.
2. Each group's population response distribution is normal, with its own possible mean μ_i.
3. The populations share a common variance σ^2.

Balanced sample sizes and roughly bell-shaped distributions often make the procedure reasonably tolerant of mild departures. They do not guarantee validity under strong skew, extreme outliers, dependence, or serious variance differences. Check the design and distributions; do not treat "ANOVA" as permission to ignore them.

Between groups versus within groups

Lecture slides: PDF pages 10–13, 15, 17 (open)

A **deviation** is a difference from a reference value. A **sum of squares**, abbreviated SS, adds squared deviations. Squaring prevents positive and negative deviations from canceling and gives larger discrepancies more weight.

Between-group variation: treatments sum of squares

The lecture uses SST for the *treatments* sum of squares:

$$\boxed{SST = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2.}$$

For each group: subtract the grand mean from its group mean, square that difference, multiply by the group's sample size, and then add across groups. The multiplier n_i counts how many observations share that group mean. A group mean based on many observations should matter more than one based on very few.

Notation warning. Some books use SST for *total* sum of squares. Here, to match the lecture, SST always means *treatments*; we spell out *Total SS* for the total.

Within-group variation: error sum of squares

The error sum of squares, written SSE, measures how individual observations scatter around their *own* group means:

$$\boxed{SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2.}$$

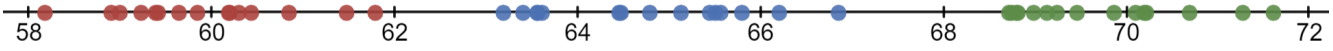
The second formula follows directly from the sample-variance definition: multiplying s_i^2 by $n_i - 1$ recovers the sum of squared deviations for group i. Add those quantities across groups. "Error" means unexplained individual variation, not necessarily a data-entry mistake.

Why both quantities are necessary

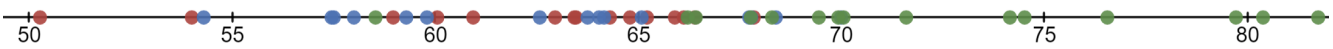
Suppose three group means are 10, 15, and 20. If observations within each group sit very close to their own mean, the separation is striking. If every group spreads broadly from 0 to 30, the same means are much less convincing evidence. The between-group differences have meaning only relative to the within-group variation.

The lecture's Scenario A has tightly clustered, well-separated groups. Scenario B has much more within-group spread. ANOVA quantifies that visual comparison by using a ratio, after adjusting both sums of squares for their degrees of freedom.

Scenario A:



Scenario B:



Each dot is one observation; color identifies its group. These are the lecture's dot plots.

Why total variation splits into two pieces

Lecture slides: PDF pages 18 (open)

The **total sum of squares** measures scatter of every observation about the grand mean:

$$\text{Total SS} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2.$$

It obeys the identity

$$\boxed{\text{Total SS} = SST + SSE.}$$

This is an exact algebraic identity for the data, not an extra statistical assumption. It says that overall scatter can be separated into movement of the group means and scatter within the groups.

Derive it one step at a time

For a single observation, add and subtract its group mean:

$$x_{ij} - \bar{x} = (x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x}).$$

Use the algebra rule $(a+b)^2 = a^2 + 2ab + b^2$ to square both sides:

$$\begin{aligned}(x_{ij} - \bar{x})^2 = (x_{ij} - \bar{x}_i)^2 \\ + 2(x_{ij} - \bar{x}_i)(\bar{x}_i - \bar{x}) + (\bar{x}_i - \bar{x})^2.\end{aligned}$$

Now add over all j in a fixed group i. The quantity $\bar{x}_i - \bar{x}$ does not change within that group, so the middle sum becomes

$$2(\bar{x}_i - \bar{x}) \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) = 0.$$

Why zero? The deviations from a group's own mean sum to zero:

$$\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) = \sum_{j=1}^{n_i} x_{ij} - n_i \bar{x}_i = n_i \bar{x}_i - n_i \bar{x}_i = 0.$$

The last squared term repeats n_i times. Thus, within group i,

$$\sum_{j} (x_{ij} - \bar{x})^2 = \sum_{j} (x_{ij} - \bar{x}_i)^2 + n_i (\bar{x}_i - \bar{x})^2.$$

Adding over all groups gives Total SS $= SSE + SST$.

How to use the identity

If a problem supplies total SS and one component, find the missing component by subtraction. As a check, all sums of squares must be nonnegative, apart from negligible numerical rounding. A substantial negative answer indicates a calculation or transcription error.

Degrees of freedom, mean squares, and the F test

Lecture slides: PDF pages 14–21 (open)

Degrees of freedom, abbreviated *df*, count the independent pieces of variation left after fitting the required means. Total variation has $N-1$ degrees of freedom because we estimate one grand mean. Within group i, estimating its mean leaves $n_i - 1$ degrees of freedom. Adding these gives

$$\sum_{i=1}^{k}(n_i - 1) = \sum_i n_i - k = N - k.$$

Between-group variation has $k-1$ degrees of freedom: k group means are compared with one grand mean. The counts add correctly: $(k-1) + (N-k) = N-1$.

A **mean square** is a sum of squares divided by its degrees of freedom. Define the treatment mean square MST and error mean square MSE by

$$MST = \frac{SST}{k-1}, \qquad MSE = \frac{SSE}{N-k}, \qquad \boxed{F = \frac{MST}{MSE}.}$$

Source	*df*	*SS*	*MS*	*F*
Treatments	$k-1$	SST	$SST/(k-1)$	MST/MSE
Error	$N-k$	SSE	$SSE/(N-k)$	
Total	$N-1$	Total SS		

Under the normal, independent, equal-variance model, MSE estimates the common variance whether or not the means are equal. When the null is true, MST also estimates that variance. Their ratio tends to be of order 1; it is not required to equal 1 exactly.

The Chapter 10 connection

A **chi-square distribution** is the distribution of a sum of squares of independent standard normal variables. Under H_0, SST/σ^2 has a chi-square distribution with $k-1$ degrees of freedom; SSE/σ^2 has one with $N-k$ degrees of freedom, and the two are independent. Dividing each by its degrees of freedom gives

$$\frac{(SST/\sigma^2)/(k-1)}{(SSE/\sigma^2)/(N-k)} = \frac{MST}{MSE} \sim F_{k-1,N-k}.$$

The common σ^2 cancels. F_{d_1,d_2} denotes the F distribution with numerator degrees of freedom d_1 and denominator degrees of freedom d_2.

Decision. Reject for a *large* observed F, in the upper tail. Let $F_{\alpha;d_1,d_2}$ mean the cutoff with probability α to its right. Reject if $F > F_{\alpha;k-1,N-k}$. The p-value is the null probability of an F at least as large as observed. Although the alternative includes differences in either direction, *ANOVA's F test is upper-tailed* because it uses squared differences.

What the effects model says about means

Lecture slides: PDF pages 22 (open)

The same hypothesis in model language

A **statistical model** expresses an observation as a systematic part plus random variation. Let X_{ij} be the random response before observation, μ a baseline mean, a_i the effect of treatment i, and ε_{ij} a random error. Write $N(0, \sigma^2)$ for a normal distribution with mean zero and variance σ^2. The model is

$$X_{ij} = \mu + a_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \text{ independently.}$$

Thus group i has population mean $\mu_i = \mu + a_i$. We use a_i rather than the lecture's α_i to avoid confusing a treatment effect with the test significance level α.

To make the representation unique, choose the baseline and a constraint, for example $\sum_i n_i a_i = 0$. Then μ is the sample-size-weighted population mean. Under that convention, all population means are equal exactly when all $a_i = 0$.

MSE estimates noise variance. MST reflects noise *plus differences in means*; it is not itself an estimate of an individual treatment effect. Write E for a repeated-sampling average. Under unequal means,

$$\mathrm{E}(MST) = \sigma^2 + \frac{\sum_i n_i a_i^2}{k-1}, \qquad \mathrm{E}(MSE) = \sigma^2,$$

This explains why separation among true means tends to make F large.

Read the expected-value formula in words

The notation a_i^2 squares each treatment effect, making its contribution nonnegative. Multiplying by n_i weights that effect by its group's sample size. If every treatment effect is zero, then the extra sum is zero and both mean squares estimate the same noise variance. If at least one treatment effect is nonzero, the numerator's repeated-sampling average increases while the denominator's does not.

This does not guarantee that every individual sample produces a large F: sampling noise remains. It explains why a large F is evidence against equal means, and why an F close to 1 is not strong evidence against them.

Do not confuse an effect with a mean square

For breakfast, a_i is a shift in *score units*. In contrast, MST has *squared score units* and combines noise with differences among several treatment means. To estimate a specific population-mean difference, compare the relevant sample means and use an appropriate standard error, as on pages 14–15.

Scenario A and Scenario B: all table calculations

Lecture slides: PDF pages 23–24 (open)

Lecture Example 1. Both dot plots have $k = 3$ groups, each containing 15 observations. The slide supplies the sums of squares. We use those supplied values rather than trying to read raw measurements from a graph.

Step 1: Count observations and degrees of freedom

Both scenarios have $N = 15 + 15 + 15 = 45$ observations. Treatment $df = k - 1 = 3 - 1 = 2$, error $df = N - k = 45 - 3 = 42$, and total $df = N - 1 = 45 - 1 = 44$. Check: $2 + 42 = 44$.

Step 2: Complete Scenario A

The supplied treatment and error sums are $SST = 720.82$ and $SSE = 43.75$.

$$\begin{aligned}\text{Total SS} &= SST + SSE = 720.82 + 43.75 = 764.57,\\ MST &= SST/2 = 720.82/2 = 360.41,\\ MSE &= SSE/42 = 43.75/42 = 1.0416667,\\ F &= MST/MSE = 360.41/1.0416667 = 345.9936 \approx 345.99.\end{aligned}$$

The slide's 345.97 differs slightly because of rounding. The displayed sums imply 345.99 to two decimal places.

Step 3: Complete Scenario B

Now $SST = 849.18$ and $SSE = 1190.92$, with the same degrees of freedom.

$$\begin{aligned}\text{Total SS} &= 849.18 + 1190.92 = 2040.10,\\ MST &= 849.18/2 = 424.59,\\ MSE &= 1190.92/42 = 28.3552381,\\ F &= 424.59/28.3552381 = 14.97395 \approx 14.97.\end{aligned}$$

Step 4: Compare evidence, not only mean separation

B has a slightly larger treatment sum of squares than A. But B's within-group scatter is far larger, so its denominator MSE is much larger and its F ratio is much smaller. This agrees with the plots: A's tight, separated groups give stronger evidence of unequal population means.

If we choose $\alpha = 0.05$, the upper-tail cutoff for $F_{2,42}$ is about 3.21994. Both 345.99 and 14.97 exceed it. Both reject the all-equal hypothesis under the model; A provides more extreme evidence. The upper-tail p-values are approximately 8.11×10^{-27} and 1.23×10^{-5}, respectively. B's more mixed appearance does *not* mean it has no evidence of a difference.

Complete an ANOVA table without raw data

Lecture slides: PDF pages 26–27 (open)

The lecture supplies the following incomplete table. The symbols and relationships are the same as on page 7; missing entries can be recovered algebraically.

Source	*df*	*SS*	*MS*	*F*
Treatments	?	?	?	?
Error	?	47.4	2.9625	
Total	19	123.2		

Step 1: Recover the error degrees of freedom

By definition, error mean square is error sum of squares divided by error degrees of freedom. Write d_E for the unknown error degrees of freedom:

$$2.9625 = \frac{47.4}{d_E} \implies 2.9625 d_E = 47.4 \implies d_E = \frac{47.4}{2.9625} = 16.$$

Step 2: Recover group and sample counts

Total $df = N - 1 = 19$ gives $N = 20$. Since treatment and error degrees of freedom add to total, treatment $df = 19 - 16 = 3$. Treatment $df = k - 1 = 3$ therefore gives $k = 4$ groups. **Step 3: Recover the treatment sum of squares**

$$SST = \text{Total SS} - SSE = 123.2 - 47.4 = 75.8.$$

Step 4: Divide by degrees of freedom, then form the ratio

$$MST = \frac{75.8}{3} = 25.26667, \qquad F = \frac{25.26667}{2.9625} = 8.52883.$$

Keep unrounded values in calculations and round only the displayed answer.

Source	*df*	*SS*	*MS*	*F*
Treatments	3	75.8	25.26667	8.52883
Error	16	47.4	2.96250	
Total	19	123.2		

Check the table. $3 + 16 = 19$ and $75.8 + 47.4 = 123.2$. Do not add the mean squares or the F column. Blank total-MS and total-F cells are intentional. If testing at 5%, this F has p-value about 0.00130 with $(3, 16)$ degrees of freedom, giving evidence against equal means under the model.

Calcium intake: set up the study and summaries

Lecture slides: PDF pages 28–30 (open)

The lecture compares daily calcium intake among adults with normal bone density, osteopenia (low bone density), and osteoporosis. We use the numerical intake units in the supplied data; squared sums and variances use squared intake units.

Normal bone density	Osteopenia	Osteoporosis
1200	1000	890
1000	1100	650
980	700	1100
900	800	900
750	500	400
800	700	350

Step 1: State the populations, hypotheses, and model

Let μ_1, μ_2, μ_3 be population mean daily calcium intake for the three categories in the order shown. We test

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad \text{against} \quad H_a : \text{at least one mean differs.}$$

For this worked example choose $\alpha = 0.05$ before examining the result. We apply the lecture's independent, normal, common-variance model as an assumption, not a fact established merely by this small table. The groups have six observations each, so $k = 3$ and $N = 6 + 6 + 6 = 18$.

Step 2: Compute each mean from its total

For group 1, add $1200 + 1000 + 980 + 900 + 750 + 800 = 5630$. The other totals are

$$1000 + 1100 + 700 + 800 + 500 + 700 = 4800,$$

$$890 + 650 + 1100 + 900 + 400 + 350 = 4290.$$

Divide each total by 6:

$$\bar{x}_1 = 5630/6 = 938.33333, \quad \bar{x}_2 = 4800/6 = 800, \quad \bar{x}_3 = 4290/6 = 715.$$

Step 3: Compute the grand mean

Add all group totals and divide by the number of individual observations:

$$\bar{x} = \frac{5630 + 4800 + 4290}{18} = \frac{14720}{18} = 817.77778.$$

Equivalently, $(6\bar{x}_1 + 6\bar{x}_2 + 6\bar{x}_3)/18$ gives the same value. The lecture's intermediate expression has a typographical 1938.3; the correct first group mean is 938.33333. Always check that the grand mean lies between the smallest and largest group means.

Calcium intake: compute every source of variation

Lecture slides: PDF pages 30–31 (open)

Step 4: Compute within-group squared deviations

For the first group, subtract its mean, square, and add:

$$\begin{aligned}\sum_{j=1}^{6}(x_{1j}-\bar{x}_1)^2 &= (1200-938.33333)^2+(1000-938.33333)^2\\ &\quad+(980-938.33333)^2+(900-938.33333)^2\\ &\quad+(750-938.33333)^2+(800-938.33333)^2\\ &= 68469.44444+3802.77778+1736.11111\\ &\quad+1469.44444+35469.44444+19136.11111\\ &= 130083.33333.\end{aligned}$$

Dividing by $6-1=5$ gives $s_1^2=26016.66667$. The same operation for the other groups gives

$$\begin{aligned}\sum_j(x_{2j}-800)^2 &= 200^2+300^2+(-100)^2+0^2+(-300)^2+(-100)^2\\ &= 240000, \qquad s_2^2=240000/5=48000;\\ \sum_j(x_{3j}-715)^2 &= 175^2+(-65)^2+385^2+185^2+(-315)^2+(-365)^2\\ &= 449750, \qquad s_3^2=449750/5=89950.\end{aligned}$$

Thus the error sum of squares is

$$SSE = 130083.33333+240000+449750 = 819833.33333.$$

Using variances gives the same answer: $SSE = 5(26016.66667+48000+89950)$.

Step 5: Compute between-group squared deviations

Subtract the grand mean from each group mean and weight each squared difference by its group size:

$$\begin{aligned}SST &= 6(938.33333-817.77778)^2\\ &\quad+6(800-817.77778)^2+6(715-817.77778)^2\\ &= 6(120.55556)^2+6(-17.77778)^2+6(-102.77778)^2\\ &= 152477.77778.\end{aligned}$$

The displayed decimals are rounded, but this answer uses the unrounded fractions. To avoid accumulated rounding error, store the full calculator values until the final step.

Step 6: Check total variation

$$\text{Total SS} = SST+SSE = 152477.77778+819833.33333 = 972311.11111.$$

This equals the sum of squared deviations of all 18 observations from 817.77778. Within-group variation is considerably larger than the between-group sum here, but we must still divide by degrees of freedom before comparing them.

Calcium intake: test and interpret

Lecture slides: PDF pages 32 (open)

Step 7: Convert sums of squares to mean squares

There are $k = 3$ groups and $N = 18$ observations, so treatment $df = 3 - 1 = 2$ and error $df = 18 - 3 = 15$.

$$MST = \frac{152477.77778}{2} = 76238.88889, \qquad MSE = \frac{819833.33333}{15} = 54655.55556.$$

Step 8: Form the F ratio

$$F = \frac{76238.88889}{54655.55556} = 1.39490.$$

The ratio has no units because numerator and denominator both use squared intake units.

Source	df	SS	MS	F
Treatments	2	152477.78	76238.89	1.39490
Error	15	819833.33	54655.56	
Total	17	972311.11		

Step 9: Compare with the null distribution

The correct reference is F with numerator $df = 2$ and denominator $df = 15$. The upper 5% cutoff is

$$F_{0.05;2,15} = 3.68232.$$

Because $1.39490 < 3.68232$, the statistic does not enter the rejection region. Equivalently,

$$p = P(F_{2,15} \geq 1.39490) = 0.27823 > 0.05.$$

This probability means: if the population means were equal and the model assumptions held, an F at least this large would occur in about 27.8% of repeated samples.

Step 10: Write a conclusion about the scientific question

At the 5% significance level, these data do not provide sufficient evidence of a difference in population mean daily calcium intake among the three bone-density categories.

Do not write: "The population means are equal," "calcium intake has no relationship with bone density," or "the null has probability 0.278." Failure to reject is not proof of equality; small samples and substantial within-group variation may make differences difficult to detect. A p-value is a probability about possible data under a model, not the probability that the null is true. Because these are patient categories rather than randomized treatment assignments, the analysis alone does not establish causation either.

Estimate a mean or a difference after ANOVA

Lecture slides: PDF pages 34–36 (open)

A **confidence interval** is an estimate plus and minus a margin of error. A 95% procedure has 95% coverage across repeated samples under its assumptions. It does not assign a 95% probability to a fixed population parameter after the interval is observed.

The common-variance model lets all groups contribute to the estimate of noise. Define the pooled variance estimate $s^2 = MSE$ and its square root $s = \sqrt{MSE}$. This s is not one group's sample standard deviation s_i. Define $\nu = N - k$, the error degrees of freedom. Let $t_{\alpha/2;\nu}$ denote the positive Student's t cutoff with upper-tail probability $\alpha/2$ and ν degrees of freedom.

One prespecified population mean

For group i, the estimate is $\bar{x}_i$. Its **standard error**, the estimated standard deviation of its sampling distribution, is $s/\sqrt{n_i}$. The interval is

$$\boxed{\mu_i : \quad \bar{x}_i \pm t_{\alpha/2;N-k}\sqrt{\frac{MSE}{n_i}}.}$$

Why divide by $\sqrt{n_i}$? Averaging n_i independent measurements reduces the variance of their mean to σ^2/n_i; replace σ^2 by MSE and take a square root.

A prespecified difference of population means

For distinct groups i and j, the target is $\mu_i - \mu_j$ and the estimate is $\bar{x}_i - \bar{x}_j$. Independent group means have variances that *add* when we subtract them. Thus

$$\text{estimated SE of the difference} = \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)},$$

so the interval is

$$\boxed{\mu_i - \mu_j : \quad (\bar{x}_i - \bar{x}_j) \pm t_{\alpha/2;N-k}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}.}$$

The common-variance and independence assumptions are essential. The t reference is exact under the normal model because the mean estimate and pooled variance estimate are independent and the latter has $N - k$ degrees of freedom.

Why use MSE rather than a single s_i^2? It combines the within-group information from all groups, usually giving a more stable estimate of the common variance. It does not mix together mean differences as noise. Use $N - k$ degrees of freedom, not merely $n_i - 1$ or $n_i + n_j - 2$.

Confidence intervals: a fully worked calculation

Lecture slides: PDF pages 34–36 (open)

Supplement to the calcium lecture example. Use the preceding calcium data to illustrate individual 95% intervals. These are not simultaneous statements about all possible groups or pairs.

One mean: normal-bone-density group

We have $\bar{x}_1 = 938.33333$, $n_1 = 6$, $MSE = 54655.55556$, and $N - k = 15$. For 95% confidence, $\alpha = 0.05$ and $\alpha/2 = 0.025$; the t cutoff is $t_{0.025;15} = 2.13145$.

1. Divide the variance estimate by the group size: $54655.55556/6 = 9109.25926$.
2. Take the square root to obtain standard error: $\sqrt{9109.25926} = 95.44244$.
3. Multiply by the cutoff to get margin of error: $2.13145(95.44244) = 203.43074$.
4. Subtract and add the margin: $938.33333 \pm 203.43074 = (734.90259, 1141.76408)$.

The result estimates the population mean intake in this group, not the range containing 95% of individual patients' intakes.

A difference: normal bone density minus osteoporosis

Keep the order consistent: the target is $\mu_1 - \mu_3$, and the observed difference is $938.33333 - 715 = 223.33333$.

1. Add reciprocal sample sizes: $1/6 + 1/6 = 1/3$.
2. Multiply by MSE: $54655.55556(1/3) = 18218.51852$.
3. Take a square root: SE $= \sqrt{18218.51852} = 134.97599$.
4. Multiply by 2.13145: margin $= 287.69452$.
5. Form endpoints: $223.33333 \pm 287.69452 = (-64.36118, 511.02785)$.

The interval includes zero and both signs. The data are compatible with a range of differences under the model; they do not precisely determine the direction. Reversing the group order negates and reverses the endpoints.

When many intervals are reported

An ordinary 95% interval is for a *prespecified individual* target. Choosing the most impressive pair after inspecting all pairs changes the inferential problem. One simple simultaneous method is **Bonferroni**: for m planned intervals and total error allowance α, make each interval have error allowance α/m. Use $t_{\alpha/(2m);N-k}$, a larger cutoff. For all three calcium pairs, this is $t_{0.05/6;15} \approx 2.69374$. The resulting intervals are wider. This guarantees at least $1 - \alpha$ simultaneous coverage without requiring independent intervals.

Unequal group sizes: a parallel worked example

Lecture slides: PDF pages 12–21 (open)

Additional biological illustration. Three independently sampled culture conditions give a measured growth response. The group sizes, sample means, and sample variances are:

Quantity	Condition 1	Condition 2	Condition 3
n_i	3	4	5
$\bar{x}_i$	12	15	10
s_i^2	4	5	6

Assume independent observations, normal populations, and common population variance. Test equality of the three population means at $\alpha = 0.05$.

Step 1: Weight the grand mean correctly

There are $N = 3+4+5 = 12$ observations and $k = 3$ groups. The group totals are $3(12) = 36$, $4(15) = 60$, and $5(10) = 50$. Therefore

$$\bar{x} = \frac{36+60+50}{12} = \frac{146}{12} = 12.16667.$$

A simple average $(12+15+10)/3 = 12.33333$ would be wrong because the groups have different sizes.

Step 2: Find each sum of squares

$$\begin{aligned} SST &= 3(12-12.16667)^2 + 4(15-12.16667)^2 + 5(10-12.16667)^2 \\ &= 0.08333 + 32.11111 + 23.47222 = 55.66667, \\ SSE &= (3-1)(4) + (4-1)(5) + (5-1)(6) \\ &= 8 + 15 + 24 = 47. \end{aligned}$$

These expressions use n_i for between-group weights and $n_i - 1$ for within-group weights. Do not interchange them. The given values are *variances*, so do not square them again.

Step 3: Build the table and decide

Source	*df*	*SS*	*MS*	*F*
Treatments	$3-1=2$	55.66667	27.83333	5.32979
Error	$12-3=9$	47	5.22222	
Total	$12-1=11$	102.66667		

Compute $F = 27.83333/5.22222 = 5.32979$. The upper-tail p-value under $F_{2,9}$ is $0.02972 < 0.05$. Reject the all-equal hypothesis: there is evidence of a difference in population mean growth among these conditions. Do not conclude that every pair differs.

A reproducible one-way ANOVA workflow

Lecture slides: PDF pages 13–21, 33 (open)

Before calculating

1. Identify the response and its units, the experimental or sampling unit, and the groups. Check whether the groups are independent or observations are naturally matched.
2. Define the population means. State H_0 (all equal), H_a (not all equal), and the chosen significance level α.
3. Explain the normality, independence, and common-variance assumptions. Plot observations by group if raw data are available. Note serious outliers or unequal spreads rather than quietly ignoring them.

Compute in an order that avoids backtracking

4. Find each n_i, mean $\bar{x}_i$, and sample variance s_i^2; then $N = \sum_i n_i$ and $\bar{x} = \sum_i n_i \bar{x}_i / N$.
5. Find $SST = \sum_i n_i(\bar{x}_i - \bar{x})^2$ and $SSE = \sum_i (n_i - 1)s_i^2$. Add them for total SS.
6. Use treatment $df = k - 1$, error $df = N - k$, and total $df = N - 1$. Divide each source SS by its own df for the mean squares.
7. Compute $F = MST/MSE$. Use the upper tail with numerator $df = k-1$ and denominator $df = N-k$.
8. Compare the p-value with α or the statistic with its critical value. Write a conclusion in context without claiming equality from a nonrejection.

Excel functions and what their arguments mean

For a range of cells containing one group's observations, `COUNT(range)`, `AVERAGE(range)`, and `VAR.S(range)` return its sample size, mean, and sample variance. Use `VAR.S`, not population variance `VAR.P`. If a sample SD is supplied instead, square it to get the variance.

For an observed statistic `Fobs`, numerator degrees of freedom `d1`, and denominator degrees of freedom `d2`:

- `F.DIST.RT(Fobs,d1,d2)` returns the upper-tail p-value.
- `F.INV.RT(alpha,d1,d2)` returns the upper-tail critical value.
- `T.INV.2T(alpha,df)` returns the positive two-sided t cutoff for an individual confidence interval.

Software saves arithmetic but does not determine whether the study supports the assumptions. Supply the correct data layout and design. Retain full precision until the final report; verify the SS and df additions and the meanings of all table labels.

Blocking: remove background variation first

Lecture slides: PDF pages 37 (open)

In a **randomized block design**, experimental units are placed into blocks of similar units before treatments are assigned. A **block** is a collection sharing a background characteristic likely to affect the response. Each block receives every treatment once in the complete design considered here, with treatments randomized to units within that block.

For example, to compare three fertilizer treatments, we might divide each of several fields into three plots and assign all three fertilizers within every field. Field is the block: soil or sunlight differences between fields may otherwise swamp the fertilizer differences. Randomization within each field helps keep other influences from systematically favoring one treatment.

The connection to paired data

For two treatments, a block could be a matched pair, or a participant measured under two appropriately randomized conditions. Comparing treatments within blocks removes background differences between blocks. With more than two treatments, the same idea leads to a block ANOVA.

Do not confuse matching with independent samples. Observations associated with the same block share a background effect. The blocked model accounts for this structure. Simply treating the blocks' repeated or matched measurements as unrelated groups discards that information. Independence is assumed for the remaining errors in the model, not blindly asserted for any repeated-measures data.

Why ignoring blocks can hide a treatment effect

Imagine that every treatment produces higher responses in fertile fields than in poor fields. A one-way analysis puts field-to-field differences into its within-treatment error. The blocked analysis has a separate row for those differences, potentially leaving less unexplained noise in the denominator of the treatment F test.

Blocking is not guaranteed to improve precision: it also uses degrees of freedom. It is useful when the block characteristic explains substantial variation. Blocks should be selected from study design considerations, not invented after seeing which grouping makes a result significant.

A second factor is not automatically a block

In the lecture's phone-cost example, company is the treatment and usage category is a block. The focus is company comparison while accounting for predictable usage differences. If both factors are scientific targets and their effects may depend on each other, we need a factorial analysis that can study *interaction*, defined on page 25, rather than relying on this simple additive block model.

Check your understanding. What does "complete" mean here?

Reasoning. Each block contains an observation for every treatment. The formulas that follow assume one observation per treatment–block cell, with no missing cells.

Write the block model and define its means

Lecture slides: PDF pages 37–39, 41 (open)

Let k be the number of treatments and b the number of blocks. We have one observation for each treatment–block combination, giving $N = kb$ observations. Use $i = 1, \dots, k$ for treatment and $j = 1, \dots, b$ for block; x_{ij} is the observed response in that cell.

Let $\bar{T}_i$ be the mean for treatment i across all blocks, $\bar{B}_j$ the mean for block j across treatments, and $\bar{x}$ the grand mean:

$$\bar{T}_i = \frac{\sum_{j=1}^{b} x_{ij}}{b}, \qquad \bar{B}_j = \frac{\sum_{i=1}^{k} x_{ij}}{k}, \qquad \bar{x} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{b} x_{ij}}{kb}.$$

The letters T and B label treatment and block means; they do not represent new random distributions.

The additive model

Let X_{ij} denote a response before observation, μ an overall baseline, a_i a treatment effect, c_j a block effect, and ε_{ij} the unexplained random error. An **additive** model adds these effects without a term allowing treatment effects to change by block:

$$X_{ij} = \mu + a_i + c_j + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \text{ independently.}$$

Use constraints $\sum_i a_i = 0$ and $\sum_j c_j = 0$ so that the effects and baseline are uniquely defined. The error variance σ^2 is assumed common across all cells. This simple model assumes *no treatment-by-block interaction.*

Predicted values and residuals

The estimates of baseline and effects are $\hat{\mu} = \bar{x}$, $\hat{a}_i = \bar{T}_i - \bar{x}$, and $\hat{c}_j = \bar{B}_j - \bar{x}$. A hat marks an estimate. Thus the fitted value, or model's predicted response for cell (i, j), is

$$\hat{x}_{ij} = \bar{x} + (\bar{T}_i - \bar{x}) + (\bar{B}_j - \bar{x}) = \bar{T}_i + \bar{B}_j - \bar{x}.$$

A **residual**, written e_{ij}, is observed minus fitted:

$$e_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \bar{T}_i - \bar{B}_j + \bar{x}.$$

Why add the grand mean at the end? Subtracting both marginal means removes the baseline twice. Adding it back leaves the unexplained deviation after accounting for treatment and block location.

Build the block ANOVA table

Lecture slides: PDF pages 38–41 (open)

The treatment and block sums of squares measure their mean differences from the grand mean, weighted by the number of observations supporting each mean:

$$SST = \sum_{i=1}^{k} b(\bar{T}_i - \bar{x})^2, \qquad SSB = \sum_{j=1}^{b} k(\bar{B}_j - \bar{x})^2.$$

Here SSB means *blocks* sum of squares. Each treatment mean averages b cells; each block mean averages k cells. Total SS is $\sum_i \sum_j (x_{ij} - \bar{x})^2$, and the error sum of squares is

$$SSE = \sum_i \sum_j e_{ij}^2 = \text{Total SS} - SST - SSB.$$

In a complete balanced design these decompositions agree exactly. Use direct residuals as a check when feasible.

Where the error degrees of freedom come from

There are kb observations. Estimating one baseline, $k-1$ treatment effects, and $b-1$ block effects uses $1 + (k-1) + (b-1)$ independent quantities. Therefore the remaining count is

$$kb - 1 - (k-1) - (b-1) = kb - k - b + 1 = (k-1)(b-1).$$

Divide each source sum of squares by its degrees of freedom. In addition to MST and MSE, define $MSB = SSB/(b-1)$ as the block mean square.

Source	*df*	*SS*	*MS*	*F*
Treatments	$k-1$	SST	$SST/(k-1)$	MST/MSE
Blocks	$b-1$	SSB	$SSB/(b-1)$	MSB/MSE
Error	$(k-1)(b-1)$	SSE	$SSE/[(k-1)(b-1)]$	
Total	$kb-1$	Total SS		

Two distinct tests

For treatments, test $H_0 : a_1 = \cdots = a_k = 0$ against at least one nonzero treatment effect. Use $F_T = MST/MSE$ with numerator $df = k-1$ and denominator $df = (k-1)(b-1)$.

Treating the named blocks as fixed categories of interest, test $H_0 : c_1 = \cdots = c_b = 0$ against at least one nonzero block effect. Use $F_B = MSB/MSE$ with numerator $df = b-1$ and the same error denominator degrees of freedom. Subscripts T and B distinguish these two F statistics. Both tests use upper-tail probabilities. Blocks are often nuisance controls, so the treatment test is usually the main scientific question.

Phone costs: define the layout and compute means

Lecture slides: PDF pages 42 (open)

The lecture compares costs for four phone companies using three usage levels as blocks. This is a numerical illustration of the blocked analysis; actual causal conclusions would require an appropriate study design and model checks.

Usage block	A	B	C	D	Block mean
Low	27	24	31	23	26.25
Medium	68	76	65	67	69.00
High	308	326	312	300	311.50
Treatment mean	134.33333	142	136	130	

There are $k = 4$ treatments (companies) and $b = 3$ blocks (usage levels). Thus $N = 4(3) = 12$ observations. Each company is observed once at each usage level.

Step 1: Average each company down its column

$$\begin{aligned}\bar{T}_A &= (27 + 68 + 308)/3 = 403/3 = 134.33333,\\ \bar{T}_B &= (24 + 76 + 326)/3 = 426/3 = 142,\\ \bar{T}_C &= (31 + 65 + 312)/3 = 408/3 = 136,\\ \bar{T}_D &= (23 + 67 + 300)/3 = 390/3 = 130.\end{aligned}$$

Letters A–D label treatments in place of numeric indices $i = 1, \ldots, 4$.

Step 2: Average each usage block across its row

$$\begin{aligned}\bar{B}_{\text{Low}} &= (27 + 24 + 31 + 23)/4 = 105/4 = 26.25,\\ \bar{B}_{\text{Medium}} &= (68 + 76 + 65 + 67)/4 = 276/4 = 69,\\ \bar{B}_{\text{High}} &= (308 + 326 + 312 + 300)/4 = 1246/4 = 311.5.\end{aligned}$$

Step 3: Average all 12 costs

The overall total is $105 + 276 + 1246 = 1627$, so

$$\bar{x} = 1627/12 = 135.58333.$$

Because all treatments and blocks are balanced, averaging the four treatment means or the three block means also gives this grand mean. Keep the unrounded fraction 1627/12 for subsequent calculations.

Check your understanding. Why is the multiplier in treatment SS equal to 3 rather than 4?

Reasoning. Each company mean represents three observations, one in each block. The multiplier counts observations behind the mean, not the number of means.

Phone costs: compute sums of squares and residuals

Lecture slides: PDF pages 43–44 (open)

Step 4: Treatment sum of squares

Multiply each squared company-mean deviation by $b = 3$:

$$SST = 3\big[(134.33333 - 135.58333)^2 + (142 - 135.58333)^2 + (136 - 135.58333)^2 + (130 - 135.58333)^2\big] = 222.25.$$

Step 5: Block sum of squares

Each usage-level mean represents $k = 4$ companies, so

$$SSB = 4\big[(26.25 - 135.58333)^2 + (69 - 135.58333)^2 + (311.5 - 135.58333)^2\big] = 189335.16667.$$

Step 6: Total sum of squares

A convenient identity is Total SS $= \sum x_{ij}^2 - (\sum x_{ij})^2/N$. It follows by expanding the squares around the grand mean, just as for a sample variance. Here the sum of squared raw costs is

$$\begin{aligned} \text{Low row: } & 27^2 + 24^2 + 31^2 + 23^2 = 729 + 576 + 961 + 529 = 2795; \\ \text{Medium row: } & 68^2 + 76^2 + 65^2 + 67^2 \\ & = 4624 + 5776 + 4225 + 4489 = 19114; \\ \text{High row: } & 308^2 + 326^2 + 312^2 + 300^2 \\ & = 94864 + 106276 + 97344 + 90000 = 388484. \end{aligned}$$

Thus $\sum x_{ij}^2 = 2795 + 19114 + 388484 = 410393$. Therefore

$$\text{Total SS} = 410393 - \frac{1627^2}{12} = 410393 - 220594.08333 = 189798.91667.$$

Step 7: Subtract explained components

$$SSE = 189798.91667 - 222.25 - 189335.16667 = 241.5.$$

The slide's intermediate total and error sum contain small inconsistent rounding/transcription values. These notes use the raw cost data throughout.

Phone costs: verify error from every residual

Lecture slides: PDF pages 41, 42–44 (open)

The preceding calculation found $SSE = 241.5$ by subtraction. A separate check is to calculate each fitted value, subtract it from the observed cost, square the residual, and add the results. This check also makes the word "error" concrete.

Step 1: Calculate a residual from the model

For company A at low usage, the fitted value is $134.33333 + 26.25 - 135.58333 = 25$, so its residual is $27 - 25 = 2$. For company B in that same block, the fitted value is $142 + 26.25 - 135.58333 = 32.66667$, so the residual is $24 - 32.66667 = -8.66667$. A negative residual means the observed cost is below the model's fitted cost, not that the squared error is negative.

Step 2: Repeat observed minus fitted for every cell

Use $e_{ij} = x_{ij} - \bar{T}_i - \bar{B}_j + \bar{x}$ in each cell. This produces:

Block	A	B	C	D
Low	2	−8.66667	4.33333	2.33333
Medium	0.25	0.58333	−4.41667	3.58333
High	−2.25	8.08333	0.08333	−5.91667

Step 3: Square and add within each row

Keep the underlying unrounded values when evaluating the sums. The displayed rounded residuals give the same results to the precision shown:

$$\begin{aligned} \text{Low:} \quad & 2^2 + (-8.66667)^2 + 4.33333^2 + 2.33333^2 \approx 103.33333, \\ \text{Medium:} \quad & 0.25^2 + 0.58333^2 + (-4.41667)^2 + 3.58333^2 \approx 32.75, \\ \text{High:} \quad & (-2.25)^2 + 8.08333^2 + 0.08333^2 + (-5.91667)^2 \approx 105.41667. \end{aligned}$$

Add the row sums:

$$SSE = 103.33333 + 32.75 + 105.41667 = 241.5.$$

This agrees with the subtraction method. A discrepancy larger than rounding would be a reason to check the means, signs, or data entries.

An additional structural check

The residuals sum to zero within each row and each column, up to rounding. For example, the low row gives $2 - 8.66667 + 4.33333 + 2.33333 = 0$. This occurs because the additive fitted model has already accounted for the treatment and block means. Residuals describe what remains after those shifts have been fitted.

Phone costs: finish both tests and interpret

Lecture slides: PDF pages 44–47 (open)

Step 8: Degrees of freedom and mean squares

Treatment $df = 4 - 1 = 3$, block $df = 3 - 1 = 2$, error $df = (4 - 1)(3 - 1) = 6$, and total $df = 12 - 1 = 11$.

$$MST = 222.25/3 = 74.08333, \quad MSB = 189335.16667/2 = 94667.58333,$$

$$MSE = 241.5/6 = 40.25.$$

Step 9: Form the two ratios

$$F_T = 74.08333/40.25 = 1.84058, \qquad F_B = 94667.58333/40.25 = 2351.98965.$$

Source	*df*	*SS*	*MS*	*F*
Treatments	3	222.25	74.08333	1.84058
Blocks	2	189335.16667	94667.58333	2351.98965
Error	6	241.50	40.25	
Total	11	189798.91667		

Step 10: Test treatments at the chosen 5% level

Under the null of no company effect in the additive model, use $F_{3,6}$. Its upper 5% cutoff is 4.75706; the observed 1.84058 is smaller. The p-value is $0.24038 > 0.05$. We do not find sufficient evidence of a difference in mean costs among companies after accounting for usage level.

Step 11: Examine the usage-level effects

For no usage-level effect, use $F_{2,6}$. Its upper 5% cutoff is 5.14325, far below 2351.98965. The upper-tail p-value is approximately 2.07×10^{-9}, or 0.00000000207: very small, not literally zero. Under the model, there is strong evidence of differences by usage level.

What blocking accomplished

If we ignored usage levels, the within-company sum of squares would be $SSB + SSE = 189576.66667$. Separating out usage reduces the residual SS to 241.5. Most observed cost variation is associated with usage, not company. Nevertheless, this example does not prove companies have identical prices or explain their competitive behavior. Statistical evidence should not be replaced with a story unsupported by the design.

Interaction: when the simple block model is not enough

Lecture slides: PDF pages 48–49 (open)

An **interaction** means that the effect of one factor depends on the level of another factor. In plain language: the difference between treatments changes across blocks or groups.

A numerical illustration

Suppose average test scores under no breakfast and full breakfast were:

Age group	No breakfast	Full breakfast	Breakfast difference
Younger	60	80	$80 - 60 = 20$
Older	70	75	$75 - 70 = 5$

These are illustrative population averages, not the lecture's observed data. Breakfast has a 20-point effect for younger participants and a 5-point effect for older participants. The difference of those effects is $20 - 5 = 15$. One common additive breakfast effect cannot describe both. By contrast, differences of 20 in both age groups would be consistent with an additive model.

Connect the idea to the model

The block model $X_{ij} = \mu + a_i + c_j + \varepsilon_{ij}$ includes a treatment shift a_i that is the same in every block. An interaction would require an additional cell-dependent term. With *one observation per cell*, the simple block analysis cannot separately estimate an arbitrary interaction and a pure error term. Interaction may then be hidden in the residual and undermine the ordinary block F-test interpretation.

A **factorial experiment** examines combinations of multiple factors so their main effects and interactions can be studied. Replication, meaning multiple independent experimental units at the same factor combination, can provide information to distinguish interaction from residual noise. This chapter introduces the distinction; it does not derive the full replicated factorial ANOVA.

Design questions to ask before choosing a table

- Is one variable a background characteristic used to make fairer treatment comparisons, or are both variables scientific factors of interest?
- Is it reasonable that treatment differences are similar across blocks? Plot group patterns if possible; a nonparallel pattern can signal interaction.
- Does every block contain every treatment? The simple formulas require complete, balanced treatment–block cells.
- Are observations generated with the necessary randomization and independence structure? Cost categories or repeated measurements do not automatically satisfy these requirements.

The lecture's warning is about choosing a design and model that address the scientific question. Two labels in a data table do not, by themselves, justify a randomized-block analysis.

Final checklist: know what each calculation means

Lecture slides: PDF pages 9–49 (open)

Choose the design first

One-way independent groups	Complete randomized blocks
One factor, k groups, total N observations	k treatments, b blocks, one observation per cell
Total SS $= SST + SSE$	Total SS $= SST + SSB + SSE$
Error $df = N - k$	Error $df = (k-1)(b-1)$
Treatment F $= MST/MSE$	Treatment F $= MST/MSE$; block F $= MSB/MSE$

Checks before handing in an analysis

1. Define the response, its units, groups or blocks, and population parameters. State the hypotheses and significance level before the decision.
2. Explain the model assumptions. Independence is a design issue, not a test result you can infer from an ANOVA table.
3. Distinguish sample SD from sample variance. Square an SD if the formula needs a variance; do not square a variance again.
4. Weight the grand mean by sample sizes. In one-way ANOVA, use n_i in treatment SS and $n_i - 1$ in error SS.
5. Check that the SS and degrees-of-freedom columns add to their totals. Divide each SS by its own degrees of freedom before taking the F ratio.
6. Use the correct numerator and denominator degrees of freedom and an *upper-tail* p-value. Do not double ANOVA's p-value.
7. Keep sufficient precision. Verify suspicious negative error sums or changing answers using unrounded arithmetic.
8. Write "at least one population mean differs" after rejection. Write "insufficient evidence" after nonrejection, not "proved equal."
9. For intervals, state the target and group order, use the pooled MSE only under its common-variance model, and distinguish individual from simultaneous coverage.

Short checks with reasoning

1. Why does a method called analysis of variance test means? It asks whether mean separation is unusually large relative to individual variability; the variances form the test, while the hypotheses concern means.

2. Can F be less than 1? Yes. Sampling can make between-group mean square smaller than within-group mean square. This is not evidence for rejecting the all-equal hypothesis.

3. Does a small block p-value prove a treatment works? No. Treatment and block hypotheses have different numerators and different interpretations.