

Analysis of Variance

Comparing group means by partitioning outcome variability

Ratiomera Statistics

Table of contents

Course-Page Summary	2
One-Way ANOVA	2
Design and Model Checks	2
Extensions	3
How the Ideas Connect	3
Expanded Reference	4
Purpose and foundations	4
Core ideas	4
Formula guide	5
Reading the explanatory figure	8
Interpretation checklist	8
How this topic connects	9

Course-Page Summary

Analysis of variance, or ANOVA, compares population means across levels of one or more categorical factors. Its omnibus F test compares variation represented by the factor with the remaining variation inside the fitted groups.

One-Way ANOVA

For p factor levels,

$$H_0 : \mu_1 = \dots = \mu_p,$$

while the alternative says that at least two population means differ. It does not say that every pair differs.

The observed variation is partitioned exactly:

$$SS_{total} = SS_A + SS_e.$$

Table 1: The factor and error rows compare mean separation with within-group variation.

Source	Degrees of freedom	Mean square	Test statistic
Factor A	$p - 1$	$MS_A = SS_A / (p - 1)$	$F = MS_A / MS_e$
Error	$N - p$	$MS_e = SS_e / (N - p)$	
Total	$N - 1$		

A significant omnibus result is followed by a question-matched comparison plan. A contrast uses weights that sum to zero. Planned contrasts are specified before inspecting outcomes. Post hoc families such as Tukey all-pairs or Dunnett reference comparisons protect the familywise Type I error rate for their stated family. Bonferroni does not require independent tests, though it can be conservative.

Design and Model Checks

- A quantitative outcome and categorical factor structure must match the research question.
- Independent cases, an adequate error-distribution model, and reasonably equal error variances support the usual fixed one-way inference.
- Random assignment supports causal comparison of assigned conditions; an observed group difference alone does not.
- Residual plots can reveal unequal spread or shape concerns, but they cannot verify assignment or independence.
- A significant result does not identify every differing pair or establish practical importance.

Extensions

Table 2: Each extension changes the design question and therefore the model structure.

Extension	New question
Factorial ANOVA	How do several categorical factors and their interactions relate to the outcome?
Random factor	How much model variance is associated with levels sampled from a wider population of levels?
Repeated measures	How do conditions differ when the same people are measured repeatedly?
Analysis of covariance	How do categorical and quantitative predictors work together in one model?

For a balanced one-way random factor,

$$\hat{\sigma}_A^2 = \frac{MS_A - MS_e}{n}, \quad ICC = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_e^2}.$$

In repeated-measures ANOVA, the same person's observations are linked. Sphericity asks whether all population variances of pairwise difference scores are equal. A Greenhouse-Geisser correction multiplies the numerator and denominator degrees of freedom by $\hat{\epsilon}$; it leaves the observed F unchanged and recalculates its reference.

How the Ideas Connect

ANOVA is not isolated from regression. Dummy variables represent factor levels, fitted group means are regression fitted values, and SS_e is residual variation. Correlation, simple regression, partial correlation, multiple regression, and ANOVA therefore form one developing family of linear questions. The design, variable roles, dependence, and population target determine which version is appropriate.

Expanded Reference

Purpose and foundations

Analysis of variance, abbreviated ANOVA, compares mean outcomes across groups by studying variability. The name can feel surprising because the research question concerns means. The method works by separating total outcome variability into a part associated with group differences and a part that remains among cases within the same groups. If the group-related component is large relative to residual variation, the data provide evidence that not all population group means are equal.

In a one-way between-groups ANOVA, there is one categorical factor and one numerical outcome. A **factor** is a categorical predictor; its categories are called **levels**. Each case belongs to one level, and different cases appear in different groups. The null hypothesis states that every population group mean is equal. The alternative states that not all population means are equal, which means that at least two differ. Rejecting the null does not identify which groups differ or how large those differences are.

Begin with the design. Identify the observational or experimental unit, the factor levels, the outcome and its scale, and whether observations are independent or repeated. Plot the outcome by group and report group sample sizes, means, standard deviations, and intervals. A test statistic cannot repair a mismatch between the design and the model.

ANOVA quantity	What it records	Degrees of freedom in a one-way design
Total sum of squares	Every case's squared deviation from the grand mean	$N - 1$
Factor sum of squares	Group-mean deviations from the grand mean, weighted by group size	$k - 1$
Error sum of squares	Individual deviations from their own group means	$N - k$
Mean square	Sum of squares divided by its degrees of freedom	Depends on component

Core ideas

The **grand mean** is the mean across all observations. Total sum of squares measures how far every outcome lies from that grand mean. Factor sum of squares asks how far each group mean lies from the grand mean and weights that squared distance by the group's size. Error sum of squares measures how far each observation lies from its own group mean. In the ordinary one-way model with an intercept, factor and error sums of squares add exactly to total sum of squares.

Sums of squares grow with sample size and do not account for how many independent pieces of information were used. Dividing each component by its degrees of freedom gives a mean square. The F statistic divides factor mean square by error mean square. Under the null hypothesis and model assumptions, both estimate the same error variance in different ways, so a ratio near one is plausible. A large ratio indicates that group-mean separation is large relative to typical within-group variation.

Follow-up question	Appropriate tool	Interpretation focus
Did a planned scientific comparison differ?	Planned contrast	The specified weighted mean comparison and its uncertainty

Follow-up question	Appropriate tool	Interpretation focus
Which pairs differ after an omnibus result?	Multiplicity-adjusted pairwise comparisons	Pair differences with simultaneous error control
How should the omnibus analysis be documented?	Complete ANOVA table	Sums of squares, degrees of freedom, mean squares, F , and p-value
Does one factor's pattern depend on another?	Factorial ANOVA with interaction	Differences of differences rather than isolated main effects

Multiple unadjusted tests increase the probability of at least one Type I error across a family of comparisons. An omnibus ANOVA controls one overall question but does not replace thoughtfully chosen follow-ups. Planned contrasts should come from the research question. Post hoc pairwise procedures use an adjustment designed for the family being interpreted. Report the estimated mean differences or contrasts, their uncertainty, and adjusted p-values.

A factorial ANOVA contains more than one factor. Main effects summarize average differences for one factor across levels of the other. An interaction asks whether the effect of one factor changes across the other factor's levels. When an interaction is meaningful, interpret conditional group means and contrasts rather than relying on main effects alone.

Repeated-measures data require a model that recognizes that several observations belong to the same person or unit. Those observations are correlated and cannot be treated as independent groups. Sphericity is a repeated-measures condition concerning the variances of pairwise differences among levels. When it is not plausible, a degree-of-freedom correction or a suitable repeated-data model is required. A random-effects perspective separates variation between clusters or people from variation within them; the intraclass correlation summarizes how strongly observations from the same cluster resemble one another.

The ordinary between-groups model assumes independent observations, an appropriate mean structure, and residual variances that are suitable for the intended F inference. Residual diagnostics and group displays matter. With unequal variances and group sizes, the standard pooled-error test may be unsuitable. The response should follow the design and material's stated procedure rather than an automatic transformation or deletion of cases.

Formula guide

The one-way model writes each outcome as a grand mean, a group effect, and an individual error:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

Here i identifies a case within group j , μ is the grand reference, α_j is the group component, and ε_{ij} is residual variation. The total sum of squares is:

$$SS_{\text{total}} = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2$$

Its exact partition is:

$$SS_{\text{total}} = SS_{\text{factor}} + SS_{\text{error}}$$

The two components are calculated as:

$$SS_{\text{factor}} = \sum_{j=1}^k n_j (\bar{y}_j - \bar{y})^2, \quad SS_{\text{error}} = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

Mean squares divide by their degrees of freedom, and the omnibus test compares them:

$$F = \frac{MS_{\text{factor}}}{MS_{\text{error}}} = \frac{SS_{\text{factor}}/(k-1)}{SS_{\text{error}}/(N-k)}$$

With exactly two independent groups, this one-way fixed-effects test and the two-sided pooled-variance independent-samples t test are equivalent only when they use the same equal-variance model:

$$F(1, N-2) = t(N-2)^2.$$

The omnibus result coordinates the group comparison, but it does not identify the differing means. A focused **contrast** combines group means with weights that sum to zero:

$$D = \sum_{i=1}^k c_i \bar{y}_i, \quad \sum_{i=1}^k c_i = 0.$$

Positive and negative weights place levels on opposite sides of the intended comparison. For a balanced design with n cases in every level, the contrast calculation used in the supplied material is

$$SS_D = \frac{nD^2}{\sum_i c_i^2}, \quad F_D = \frac{SS_D}{MS_{\text{error}}},$$

with 1 numerator degree of freedom and the omnibus error degrees of freedom in the denominator. A comparison is planned only when its weights were selected before the outcomes were inspected.

The number of distinct pairs among k levels is

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

For m mutually independent tests, each with testwise Type I error probability α_{test} , the exact familywise error rate is

$$\alpha_{\text{family}} = 1 - (1 - \alpha_{\text{test}})^m.$$

Solving that relationship for a target familywise level gives the Sidak threshold, while Bonferroni gives a bound that does not require independent tests:

$$\alpha_{\text{test,Sidak}} = 1 - (1 - \alpha_{\text{family}})^{1/m}, \quad \alpha_{\text{test,Bonferroni}} = \frac{\alpha_{\text{family}}}{m}.$$

The Sidak equality is exact only for mutually independent tests. Pairwise comparisons that share groups are generally dependent. Bonferroni controls the familywise error through an upper bound without that independence requirement, although it can be conservative.

A two-factor fixed-effects ANOVA writes a cell outcome as a grand mean, two main-effect components, their interaction, and an individual error:

$$y_{ijm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijm}.$$

A **cell mean** belongs to one exact combination of factor levels. A **marginal mean** averages across the cells belonging to one level of one factor. Main effects compare marginal means. The interaction asks whether the effect of one factor changes across levels of the other, which is a difference-of-differences question. Nonparallel mean profiles show an interaction pattern; the lines do not have to cross.

For a balanced one-way random-factor model with n observations per sampled level, the supplied material estimates the between-level and within-level variance components by

$$\hat{\sigma}_A^2 = \frac{MS_A - MS_{\text{error}}}{n}, \quad \hat{\sigma}_{\text{error}}^2 = MS_{\text{error}},$$

and summarizes within-level resemblance with

$$ICC = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_A^2 + \hat{\sigma}_{\text{error}}^2}.$$

These equations belong to that balanced one-way random-factor setting. They are not a universal ICC formula for every clustered or repeated design.

For one repeated factor, the person term preserves the link among measurements from the same person:

$$y_{im} = \mu + \alpha_i + \pi_m + \varepsilon_{im},$$

where α_i is the fixed occasion or condition component and π_m is the random person component. The corresponding variation partition is

$$SS_{\text{total}} = SS_{\text{condition}} + SS_{\text{person}} + SS_{\text{error}}.$$

For two repeated levels j and k , the variance of the within-person difference is

$$Var(Y_j - Y_k) = Var(Y_j) + Var(Y_k) - 2Cov(Y_j, Y_k).$$

Sphericity requires the population variances of all such pairwise difference scores to be equal. When the stated Greenhouse-Geisser procedure is used, its estimate $\hat{\varepsilon} \leq 1$ reduces both reference degrees of freedom:

$$df_{\text{condition}}^* = \hat{\varepsilon} df_{\text{condition}}, \quad df_{\text{error}}^* = \hat{\varepsilon} df_{\text{error}}.$$

The observed F statistic does not change. Its reference degrees of freedom and resulting p-value or critical value do.

Design question	Quantity or comparison	Essential limit
Are all fixed population group means equal?	Omnibus one-way F	Rejection does not locate the difference
Which prespecified weighted means differ?	Planned contrast D and F_D	Planning must occur before inspecting outcomes
Does one factor's effect depend on another?	Factorial interaction	Main effects alone can hide the cell pattern
How much variation belongs to sampled levels?	Random-factor variance component and ICC	Formula depends on the stated random-effects design
Do linked occasions differ?	Repeated-measures condition effect	Within-person dependence and sphericity procedure matter

Reading the explanatory figure

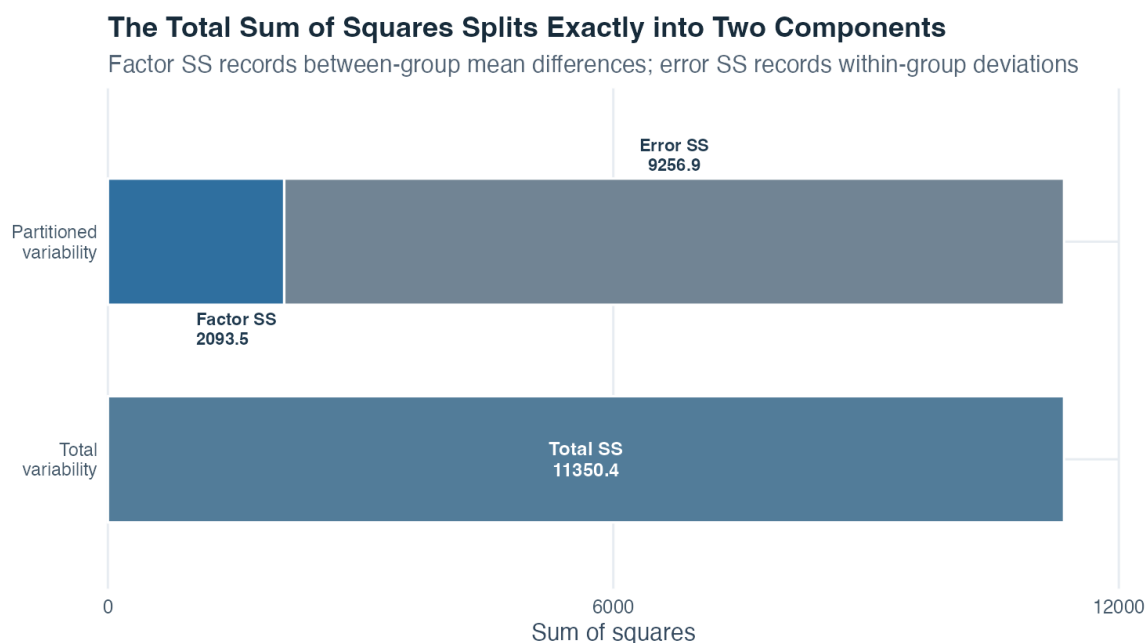


Figure 1: Two bars show total sum of squares beside an equally tall stacked bar divided into factor and error sums of squares with numerical labels.

The left bar contains the total sum of squares, 11,350.4. It represents squared deviations of every observed score from the grand mean. The right bar has the same total height but is stacked. The lower blue segment is factor sum of squares, 2,093.5, and represents group-mean separation. The upper gray segment is error sum of squares, 9,256.9, and represents differences among cases around their respective group means.

The two right-hand labels add to the total on the left. This visual equality is the core ANOVA identity. The error component is larger in this dataset, but the F test does not compare the raw segment heights directly. Each sum of squares is first divided by its degrees of freedom. The resulting mean-square ratio is then evaluated against an F distribution under the null model.

The bars do not show which group mean is highest, which pairs differ, whether the residual assumptions are plausible, or whether the design supports causation. Those questions require the group plot, descriptive table, diagnostics, and planned contrasts. The displayed values come from simulated teaching data and therefore illustrate the calculation rather than provide evidence about a real population.

Interpretation checklist

Identify the factor, levels, outcome, unit of analysis, and between-groups or repeated structure. Report group counts, means, standard deviations, and a clear group display. State the omnibus null and alternative. Check independence from the design and inspect residual variation and unusual observations. Report the complete ANOVA table with sums of squares, degrees of freedom, mean squares, F , and p-value, followed by the estimates and uncertainty for planned or adjusted comparisons.

After an omnibus result, answer the substantive question with planned or multiplicity-adjusted comparisons. For a factorial design, interpret interactions before averaging across them. For repeated measurements, preserve within-person dependence and address the documented sphericity procedure. Avoid saying that a nonsignificant result proves equal means or that a significant result proves an important or causal difference.

How this topic connects

ANOVA completes the learning sequence by returning to variance, the quantity introduced in descriptive statistics. Probability and inference explain the F reference distribution and decision process. Covariance and correlation show how variables share linear variation. Simple regression partitions outcome variation into fitted and residual parts. Partial correlation and multiple regression explain conditional associations after other information is considered.

Group membership in ANOVA can be coded as indicator predictors in a regression model. The factor sum of squares is model-related variation; error sum of squares is residual variation; the omnibus F test compares the full group model with an intercept-only model. Factorial interactions match regression interactions, and planned contrasts are targeted linear comparisons of fitted means. This shared framework is best understood as the **general linear model connection**: the later topics are different views of how structured predictor information is used to explain outcome variability while uncertainty and residual variation remain visible.

Document ID: topic-08-analysis-of-variance-summary-en

Notes 1 of 3

Analysis of Variance | Topic 8

Notes 2 of 3

Analysis of Variance | Topic 8

Notes 3 of 3

Analysis of Variance | Topic 8