How can I determine the correct error term in an ANOVA?

Authored by

stats writer

June 30, 2024

An ANOVA (Analysis of Variance) is a statistical method used to compare means of multiple groups. In order to accurately interpret the results of an ANOVA, it is important to determine the correct error term. The error term represents the variability within each group and is used to estimate the standard error of the mean. To determine the correct error term, one must first identify the type of ANOVA being used (e.g. one-way, two-way), the number of groups being compared, and the type of data being analyzed (e.g. within-subject, between-subject). Once these factors are identified, the appropriate error term can be selected from a table or calculated using a formula. Selecting the correct error term is crucial in accurately assessing the significance of the ANOVA results and making valid conclusions about the differences between the groups.

FAQ: How can I determine the correct error term in an ANOVA?

FAQ: How can I determine the correct error term in an ANOVA?

One method for determining correct denominators in analysis of variance is the Cornfield-Tukey method. This FAQ presents a modified version of the Cornfield-Tukey method for manually deriving the symbolic values for the expected mean squares. It is from these expected mean squares that one can determine appropriate error terms.

Please note that this approach to deriving expected mean squares assumes that the interaction of the fixed and random effects sum to zero over the

fixed effect levels. This approach can be found in a number of classical ANOVA textbooks, such as, Kirk, Winer and Keppel.
This assumption, however, is not universal and is not used in most mixed programs (proc mixed, xtmixed, etc).

If you would like to try a program that automates much of the computation for this algorithm, go to How can I determine the correct term in an anova using Stata?.

Steps in deriving expected mean squares

Step 1 - Write the linear model for the design.
Step 2 - Construct a table with three parts.
Step 3 - The row headings in part 1 contain each of the terms from the linear model including their subscripts but leaving out μ.
Step 4 - The column heading in part 2 contain the subscripts from the linear model, the symbol for the number of levels along with the sampling coefficient. Sampling coefficients are coded 1 for random variables

and 0 for fixed.

Step 5 - If a column heading appears as a row subscript in parentheses

enter a 1 in part 2.

Step 6 - If a column heading appears as a row subscript, not in parentheses,

enter the appropriate sampling coefficient (0 or 1).

Step 7 - If a column heading does not appear as a row subscript enter the

letter for the number of levels

Step 8 - In part 3 list a variance for each term in the linear model that

contains all the row subscripts.

Step 9 - Coefficients for variances are obtained by covering the column headed

by subscripts that appear in the row but not including subscripts in

parentheses. Obviously, terms with zero coefficients drop out.

Example three-way factorial design

In this example, A & C are fixed and B is random. The

subscript for ε is
i(jkl) because the subjects are nested in the A*B*C cells. The subjects themselves
are also random. The term, εi(jkl), is known as error, within cell
or residual.

Step 1 - Yijkl = μ + αj + βk + γl + αβjk + αγjl + βγkl + αβγjkl + εi(jkl)

Part 1 Part 2 Part 3
subscript i j k l
levels n p q m
sampling coef 1 0 1 0
--

αj n 0 q m σ2ε + 0σ2αβγ + 0σ2αγ + nmσ2αβ + nqmσ2α
σ2ε + nmσ2αβ + nqmσ2α

βk n p 1 m σ2ε + 0σ2αβγ + 0σ2βγ + 0σ2αβ + npmσ2β
σ2ε + npmσ2β

γl n p q 0 σ2ε + 0σ2αβγ + npσ2βγ + 0σ2αγ + npqσ2γ
σ2ε + npσ2βγ + npqσ2γ

αβjk n 0 1 m σ2ε + 0σ2αβγ + nmσ2αβ

σ2ε + nmσ2αβ

αγjl n 0 q 0 σ2ε + nσ2αβγ + nqσ2αγ

βγkl n p 1 0 σ2εσ2ε + 0σ2αβγ + npσ2βγ
σ2ε + npσ2βγ

αβγjkl n 0 1 0 σ2ε + nσ2αβγ

εi(jkl) 1 1 1 1 σ2ε

A correctly formed F-ratio will have one more term in the numerator than in the denominator. The additional term in the numerator is the effect of interest. Thus, the F-ration for A main effect would
look something like this:

σ2ε + nmσ2αβ + nqmσ2α MS(A)
F(A) = ----------------------- = ---------
σ2ε + nmσ2αβ MS(A*B)

Here are the terms that go into each of the F-ratios for the above model:

Effect Error Term

numerator denominator
MS(A) MS(A*B)
MS(B) MS(residual)
MS(C) MS(B*C)
MS(A*B) MS(residual)
MS(A*C) MS(A*B*C)
MS(B*C) MS(residual)
MS(A*B*C) MS(residual)

Reference

Kirk, Roger E. (1998) Experimental Design: Procedures for the Behavioral Sciences,
Third Edition. Monterey, California: Brooks/Cole Publishing. ISBN 0-534-25092-0