Article Talk

Read Edit View history | Search Wikipedia

One-way analysis of variance

From Wikipedia, the free encyclopedia

In statistics, one-way analysis of variance (abbreviated one-way ANOVA) is a technique that can be used to compare means of two or more samples (using the F distribution). This technique can be used only for numerical response data, the "Y", usually one variable, and numerical or (usually) categorical input data, the "X", always one variable, hence "one-way".[1]

The ANOVA tests the null hypothesis, which states that samples in all groups are drawn from populations with the same mean values. To do this, two estimates are made of the population variance. These estimates rely on various assumptions (see below). The ANOVA produces an F-statistic, the ratio of the variance calculated among the means to the variance within the samples. If the group means are drawn from populations with the same mean values, the variance between the group means should be lower than the variance of the samples, following the central limit theorem. A higher ratio therefore implies that the samples were drawn from populations with different mean values.[1]

Typically, however, the one-way ANOVA is used to test for differences among at least three groups, since the two-group case can be covered by a t-test (Gosset, 1908). When there are only two means to compare, the t-test and the F-test are equivalent; the relation between ANOVA and t is given by $F = t^2$. An extension of one-way ANOVA is two-way analysis of variance that examines the influence of two different categorical independent variables on one dependent variable.

Contents [hide]

Assumptions [edit]

The results of a one-way ANOVA can be considered reliable as long as the following assumptions are met:

- Response variable residuals are normally distributed (or approximately normally distributed).
- Variances of populations are equal.
- Responses for a given group are independent and identically distributed normal random variables (not a simple random sample (SRS)).

If data are ordinal, a non-parametric alternative to this test should be used such as Kruskal–Wallis one-way analysis of variance. If the variances are not known to be equal, a generalization of 2-sample Welch's t-test can be used.[2]

Departures from population normality [edit]

ANOVA is a relatively robust procedure with respect to violations of the normality assumption.[3]

The one-way ANOVA can be generalized to the factorial and multivariate layouts, as well as to the analysis of covariance.[clarification needed]

It is often stated in popular literature that none of these F-tests are robust when there are severe violations of the assumption that each population follows the normal distribution, particularly for small alpha levels and unbalanced layouts.[4] Furthermore, it is also claimed that if the underlying assumption of homoscedasticity is violated, the Type I error properties degenerate much more severely.[5]

However, this is a misconception, based on work done in the 1950s and earlier. The first comprehensive investigation of the issue by Monte Carlo simulation was Donaldson (1966).[6] He showed that under the usual departures (positive skew, unequal variances) "the F-test is conservative", and so it is less likely than it should be to find that a variable is significant. However, as either the sample size or the number of cells increases, "the power curves seem to converge to that based on the normal distribution". Tiku (1971) found that "the non-normal theory power of F is found to differ from the normal theory power by a correction term which decreases sharply with increasing sample size."[7] The problem of non-normality, especially in large samples, is far less serious than popular articles would suggest.

The current view is that "Monte-Carlo studies were used extensively with normal distribution-based tests to determine how sensitive they are to violations of the assumption of normal distribution of the analyzed variables in the population. The general conclusion from these studies is that the consequences of such violations are less severe than previously thought. Although these conclusions should not entirely discourage anyone from being concerned about the normality assumption, they have increased the overall popularity of the distribution-dependent statistical tests in all areas of research."[8]

For nonparametric alternatives in the factorial layout, see Sawilowsky.[9] For more discussion see ANOVA on ranks.

The case of fixed effects, fully randomized experiment, unbalanced data [edit]

The model [edit]

The normal linear model describes treatment groups with probability distributions which are identically bell-shaped (normal) curves with different means. Thus fitting the models requires only the means of each treatment group and a variance calculation (an identical variance within the treatment groups is used). Calculations of the means and the variance are performed as part of the hypothesis test.

The commonly used normal linear models for a completely randomized experiment are:[10]

$$y_{i,j} = \mu_j + \varepsilon_{i,j} \quad \text{(the means model)}$$

or

$$y_{i,j} = \mu + \tau_j + \varepsilon_{i,j} \quad \text{(the effects model)}$$

where

$i = 1, \ldots, I$ is an index over experimental units
$j = 1, \ldots, J$ is an index over treatment groups
I_j is the number of experimental units in the jth treatment group
$I = \sum_j I_j$ is the total number of experimental units
$y_{i,j}$ are observations
μ_j is the mean of the observations for the jth treatment group
μ is the grand mean of the observations
τ_j is the jth treatment effect, a deviation from the grand mean
$\sum \tau_j = 0$
$\mu_j = \mu + \tau_j$
$\varepsilon \sim N(0, \sigma^2), \varepsilon_{i,j}$ are normally distributed zero-mean random errors.

The index i over the experimental units can be interpreted several ways. In some experiments, the same experimental unit is subject to a range of treatments; i may point to a particular unit. In others, each treatment group has a distinct set of experimental units; i may simply be an index into the j-th list.

The data and statistical summaries of the data [edit]

One form of organizing experimental observations y_{ij} is with groups in columns:

ANOVA data organization, Unbalanced, Single factor

Lists of Group Observations

	I_1	I_2	I_3	...	I_j
1	y_{11}	y_{12}	y_{13}		y_{1j}
2	y_{21}	y_{22}	y_{23}		y_{2j}
3	y_{31}	y_{32}	y_{33}		y_{3j}
⋮	⋮	⋮	⋮		⋮
i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}

	Group Summary Statistics					**Grand Summary Statistics**	
# Observed	I_1	I_2	...	I_j	...	I_J	# Observed	$I = \sum_j I_j$
Sum				$\sum_i y_{ij}$			Sum	$\sum_j \sum_i y_{ij}$
Sum Sq				$\sum_i (y_{ij})^2$			Sum Sq	$\sum_j \sum_i (y_{ij})^2$
Mean	m_1	...		m_j	...	m_J	Mean	m
Variance	s_1^2	...		s_j^2	...	s_J^2	Variance	s^2

Comparing model to summaries: $\mu = m$ and $\mu_j = m_j$. The grand mean and grand variance are computed from the grand sums, not from group means and variances.

The hypothesis test [edit]

Given the summary statistics, the calculations of the hypothesis test are shown in tabular form. While two columns of SS are shown for their explanatory value, only one column is required to display results.

ANOVA table for fixed model, single factor, fully randomized experiment

Source of variation	Sums of squares		Degrees of freedom	Mean square	
	Explanatory SS[11]	Computational SS[12]	DF	MS	F
Treatments	$\sum_{Treatments} I_j(m_j - m)^2$	$\sum_j \frac{(\sum_i y_{ij})^2}{I_j} - \frac{(\sum_j \sum_i y_{ij})^2}{I}$	$J-1$	$\frac{SS_{Treatment}}{DF_{Treatment}}$	$\frac{MS_{Treatment}}{MS_{Error}}$
Error	$\sum_{Treatments} (I_j - 1)s_j^2$	$\sum_j \sum_i y_{ij}^2 - \sum_j \frac{(\sum_i y_{ij})^2}{I_j}$	$I-J$	$\frac{SS_{Error}}{DF_{Error}}$	
Total	$\sum_{Observations} (y_{ij} - m)^2$	$\sum_j \sum_i y_{ij}^2 - \frac{(\sum_j \sum_i y_{ij})^2}{I}$	$I-1$		

MS_{Error} is the estimate of variance corresponding to σ^2 of the model.

Analysis summary [edit]

The core ANOVA analysis consists of a series of calculations. The data is collected in tabular form. Then

- Each treatment group is summarized by the number of experimental units, two sums, a mean and a variance. The treatment group summaries are combined to provide totals for the number of units and the sums. The grand mean and grand variance are computed from the grand sums. The treatment and grand means are used in the model.
- The three DFs and SSs are calculated from the summaries. Then the MSs are calculated and a ratio determines F.
- A computer typically determines a p-value from F which determines whether treatments produce significantly different results. If the result is significant, then the model provisionally has validity.

If the experiment is balanced, all of the I_j terms are equal so the SS equations simplify.

In a more complex experiment, where the experimental units (or environmental effects) are not homogeneous, row statistics are also used in the analysis. The model includes terms dependent on i. Determining the extra terms reduces the number of degrees of freedom available.

Example [edit]

Consider an experiment to study the effect of three different levels of a factor on a response (e.g. three levels of a fertilizer on plant growth). If we had 6 observations for each level, we could write the outcome of the experiment in a table like this, where a_1, a_2, and a_3 are the three levels of the factor being studied.

a_1	a_2	a_3
6	8	13
8	12	9
4	9	11
5	11	8
3	6	7
4	8	12

The null hypothesis, denoted H_0, for the overall F-test for this experiment would be that all three levels of the factor produce the same response, on average. To calculate the F-ratio:

Step 1: Calculate the mean within each group:

$$\overline{Y}_1 = \frac{1}{6}\sum Y_{1i} = \frac{6+8+4+5+3+4}{6} = 5$$
$$\overline{Y}_2 = \frac{1}{6}\sum Y_{2i} = \frac{8+12+9+11+6+8}{6} = 9$$
$$\overline{Y}_3 = \frac{1}{6}\sum Y_{3i} = \frac{13+9+11+8+7+12}{6} = 10$$

Step 2: Calculate the overall mean:

$$\overline{Y} = \frac{\sum_i \overline{Y}_i}{a} = \frac{\overline{Y}_1 + \overline{Y}_2 + \overline{Y}_3}{3} = \frac{5+9+10}{3} = 8$$

where a is the number of groups.

Step 3: Calculate the "between-group" sum of squared differences:

$$S_B = n(\overline{Y}_1 - \overline{Y})^2 + n(\overline{Y}_2 - \overline{Y})^2 + n(\overline{Y}_3 - \overline{Y})^2$$
$$= 6(5-8)^2 + 6(9-8)^2 + 6(10-8)^2 = 84$$

where n is the number of data values per group.

The between-group degrees of freedom is one less than the number of groups

$$f_b = 3 - 1 = 2$$

so the between-group mean square value is

$$MS_B = 84/2 = 42$$

Step 4: Calculate the "within-group" sum of squares. Begin by centering the data in each group

a_1	a_2	a_3
6−5=1	8−9=−1	13−10=3
8−5=3	12−9=3	9−10=−1
4−5=−1	9−9=0	11−10=1
5−5=0	11−9=2	8−10=−2
3−5=−2	6−9=−3	7−10=−3
4−5=−1	8−9=−1	12−10=2

The within-group sum of squares is the sum of squares of all 18 values in this table

$$S_W = (1)^2 + (3)^2 + (-1)^2 + (0)^2 + (-2)^2 + (-1)^2 +$$
$$(-1)^2 + (3)^2 + (0)^2 + (2)^2 + (-3)^2 + (-1)^2 +$$
$$(3)^2 + (-1)^2 + (1)^2 + (-2)^2 + (-3)^2 + (2)^2$$
$$= 1 + 9 + 1 + 0 + 4 + 1 + 1 + 9 + 0 + 4 + 9 + 1 + 9 + 1 + 1 + 4 + 9 + 4$$
$$= 68$$

The within-group degrees of freedom is

$$f_W = a(n-1) = 3(6-1) = 15$$

Thus the within-group mean square value is

$$MS_W = S_W/f_W = 68/15 \approx 4.5$$

Step 5: The F-ratio is

$$F = \frac{MS_B}{MS_W} \approx 42/4.5 \approx 9.3$$

The critical value is the number that the test statistic must exceed to reject the test. In this case, $F_{crit}(2, 15) = 3.68$ at $\alpha = 0.05$. Since $F = 9.3 > 3.68$, the results are significant at the 5% significance level. One would reject the null hypothesis, concluding that there is strong evidence that the expected values in the three groups differ. The p-value for this test is 0.002.

After performing the F-test, it is common to carry out some "post-hoc" analysis of the group means. In this case, the first two group means differ by 4 units, the first and third group means differ by 5 units, and the second and third group means differ by only 1 unit. The standard error of each of these differences is $\sqrt{4.5/6 + 4.5/6} = 1.2$. Thus the first group is strongly different from the other groups, as the mean difference is more times the standard error, so we can be highly confident that the population mean of the first group differs from the population means of the other groups. However, there is no evidence that the second and third groups have different population means from each other, as their mean difference of one unit is comparable to the standard error.

Note $F(x, y)$ denotes an F-distribution cumulative distribution function with x degrees of freedom in the numerator and y degrees of freedom in the denominator.

See also [edit]

- Analysis of variance
- F test (includes a one-way ANOVA example)
- Mixed model
- Multivariate analysis of variance (MANOVA)
- Repeated measures ANOVA
- Two-way ANOVA
- Welch's t-test

Notes [edit]

1. ^ a b Howell, David (2002). Statistical Methods for Psychology[?]. Duxbury. pp. 324–325[?]. ISBN 0-534-37770-X.

2. ^ Welch, B. L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach". Biometrika. 38 (3/4): 330–336. doi:10.2307/2332579[?]. JSTOR 2332579[?].

3. ^ Kirk, RE (1995). Experimental Design: Procedures For The Behavioral Sciences (3 ed.). Pacific Grove, CA, USA: Brooks/Cole.

4. ^ Blair, R. C. (1981). "A reaction to 'Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance.'". Review of Educational Research. 51 (4): 499–507. doi:10.3102/00346543051004499[?].

5. ^ Randolf, E. A.; Barkowiwski, R. S. (1989). "Type I error rate when real study values are used as population parameters in a Monte Carlo study". Paper Presented at the 11th Annual Meeting of the Mid-Western Educational Research Association, Chicago.

6. ^ Donaldson, Theodore S. (1966). "Power of the F-Test for Nonnormal Distributions and Unequal Error Variances"[?]. Paper Prepared for United States Air Force Project RAND.

7. ^ Tiku, M. L. (1971). "Power Function of the F-Test Under Non-Normal Situations". Journal of the American Statistical Association. 66 (336): 913–916. doi:10.1080/01621459.1971.10482371[?].

8. ^ "Archived copy"[?]. Archived from the original[?] on 2018-12-04. Retrieved 2016-09-22.

9. ^ Sawilowsky, S. (1990). "Nonparametric tests of interaction in experimental design". Review of Educational Research. 60 (1): 91–126. doi:10.3102/00346543060001091[?].

10. ^ Montgomery, Douglas C. (2001). Design and Analysis of Experiments (5th ed.). New York: Wiley. p. Section 3–2. ISBN 9780471316497.

11. ^ Moore, David S.; McCabe, George P. (2003). Introduction to the Practice of Statistics (4 ed.). W H Freeman & Co. p. 764. ISBN 0716796570.

12. ^ Winkler, Robert L.; Hays, William L. (1975). Statistics: Probability, Inference, and Decision (s (2nd ed.). New York: Holt, Rinehart and Winston. p. 761[?].

Further reading [edit]

- George Casella (18 April 2008). Statistical design[?]. Springer. ISBN 978-0-387-75965-4.

Categories: Analysis of variance | Statistical tests

This page was last edited on 17 December 2020, at 15:49 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

Privacy policy About Wikipedia Disclaimers Contact Wikipedia Mobile view Developers Statistics Cookie statement