6.2 ANOVA in a nutshell

The one-way ANOVA model

In one-way univariate analysis of variance (ANOVA), interest lies in comparing the means among several groups. More formally, ANOVA tests the null hypothesis of no differences in the population means among groups.

Let y_{ij} be the jth observation for variable Y in the ith group, with $i = 1, \ldots, a$ groups and $j = 1, \ldots n_i$ observations per group. The ANOVA linear model is: $$ y_{ij} = \mu + \alpha_i + \varepsilon_{ij} $$ where μ is the overall population mean parameter, α_i is the population group effect parameter for a particular group i and ε_{ij} is the error parameter associated with y_{ij}, the jth observation in group i. The population means for each group i can be defined as: $$ \mu_i = \mu + \alpha_i $$

and our null hypothesis in ANOVA is that all of the population group means are equal to one another, written formally as: $$ \text{H}_ 0 : \mu_1 = \mu_2 = \ldots = \mu_a $$ or, equivalently, that all of the population group effect parameters are equal to zero: $$ \text{H}_ 0: \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0. $$

Although interest lies in the comparison of *means*, it is possible that the groups differ from one another in other ways as well. For example, the groups may have different *variances*; that is, the *spread* or *dispersion* of the observations occurring within each group may differ from one another, with some groups being more spread out/dispersed than others. We may denote the population variances associated with the errors belonging to any particular group i as σ^2_i.

The F ratio

The test statistic used in ANOVA is a ratio of two mean squares. Specifically the classical univariate F ratio for the one-way ANOVA case may be defined as: $$ F = \frac{ \sum_{i=1}^a n_i(\bar{y}_ {i\cdot} - \bar{y}_ {\cdot\cdot})^2 / (a - 1) } { \sum_{i=1}^a (n_i - 1) s_i^2 / (N - a) } $$ where
$ \hspace{1cm} N = \sum_{i=1}^a n_i \hspace{0.05 cm}$, the sum of all observations;
$ \hspace{1cm} \bar{y}_ {i\cdot} = \sum_{j=1}^{n_i} y_{ij}/n_i \hspace{0.05 cm}$, the sample mean of group i;
$ \hspace{1cm} \bar{y}_ {i\cdot\cdot} = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij} / N \hspace{0.05 cm}$, the sample mean of all observations; and
$ \hspace{1cm} s_i^2 = \sum_{j=1}^{n_i} (\bar{y}_ {ij} - \bar{y}_ {i\cdot})^2 / (n_i - 1) \hspace{0.05 cm}$, the sample standard deviation of group i.

For the one-way case, we can think of F as a ratio of two measures of variation: the among-group mean square in the numerator measures the variation among the groups; the within-group

mean square in the denominator measures the variation within the groups.

Assumptions of classical ANOVA

In addition to the linear model itself (articulated above), classical ANOVA asserts the following (three-part) assumption to maintain the validity of the F test:

- The errors, ε_{ij}, are *independent* and identically distributed random variables, drawn from a *normal* distribution with a mean of 0 and a *common variance* of σ^2_ε.

We may write the 'common variance' (or 'homogeneity of variances') aspect of this assumption as a statement that all of the within-group variances are equal to one another: $$ \sigma^2_1 = \sigma^2_2 = \ldots = \sigma^2_a $$

Calculating a p-value

If the null hypothesis is true, and all of the assumptions above are fulfilled, then F is a random variable distributed as F_0, a ratio of two chi-square random variables, having degrees of freedom $(a - 1)$ and $(N - a)$ in the numerator and denominator, respectively; i.e., $$ F \sim F_0 = \frac{X_\text{num}}{X_\text{denom}} $$ where $$ X_\text{num} \sim \chi^2_{(a-1)} \hspace{1cm} \text{and} \hspace{1cm} X_\text{denom} \sim \chi^2_{(N - a)} $$

By knowing this distribution, the *p*-value for the test can then be calculated directly for any observed value F_{obs} calculated from data, as follows: $$ P = \text{Pr}(F_0 \ge F_{\text{obs}}) $$

Test by permutation

We can rather easily dispense with the assumption of normality altogether, however, by using a *permutation test*. For example, if we ran a PERMANOVA on univariate data (based on Euclidean distances), then we would obtain a classical ANOVA partitioning and associated F-ratio test-statistic, but with the *p*-value calculated empirically using *permutations*. In that case, we do not assume normality (or any other distribution), but only *exchangeability* of the observations among the groups under a true null hypothesis.

Specifically, we calculate the observed value of F for the original data, F_obs, then we randomly shuffle (permute) and re-allocate all of the N observations across all of the groups, maintaining the original sample size n_i for every group i. After the re-allocation, we get a value of F under permutation, F^{π}. We repeat this random re-allocation and re-calculation of F under permutation many times to get an entire permutation distribution of values of F^{π} under the null hypothesis of no differences among the groups, i.e., all observations y_{ij} are exchangeable. The *p*-value under permutation is then calculated empirically by tallying the number of $F^{\pi} \ge F_{\text{obs}}$ and looking at this as a proportion of the total number of permutations done, n_perm: $$ P = \frac{(\text{no. of } F^{\pi} \ge F_{\text{obs}}) + 1}{ (n_\text{perm} + 1) } $$

Note that '$+1$' in the numerator and denominator acknowledge the observed value F_{obs} as a member of this distribution (being one possible realised allocation). If we systematically do *all* possible re-allocations (in which case the '$+1$' in the numerator and denominator of the above equation would not be needed), then the resulting empirical permutation *p*-value is exact.† If we do a random sub-set of all possible permutations, then we get an estimate of the *p*-value that is nevertheless accurate (unbiased), and it gets more and more precise, the larger the value of n_perm we are able to achieve.

Violations of assumptions

Independence of the errors is typically not difficult to achieve in practice, simply by taking care with the study design itself and the way observations are sampled (e.g., using random representative sampling). A thoughtful discussion of the consequences of non-independence (positively or negatively, either within or among groups) is provided by Underwood (1997) .

Normality of the errors is typically much more difficult to fulfill; however violation of this assumption does not typically have a strong impact on the validity of the test. Even if errors are not normally distributed, the *central limit theorem* ensures that the distribution of *means* will be approximately normal, and the ANOVA F test remains quite robust. Also, the use of a permutation test to calculate the *p*-value avoids having to make this particular assumption.

The assumption of *homogeneity of variances* across all groups is also rather easy to violate in practice. For example, count data (such as the abundances of a species) typically show intrinsic mean-variance relationships, so any differences in means among groups will almost surely be accompanied by differences in variances as well (McArdle & Anderson (2004)). The assumption of *homogeneity of variances* across all groups, if violated, will not affect the validity of the test appreciably provided the design is *balanced*; i.e., if there are equal sample sizes across the groups. However, if the design is *unbalanced*, then heterogeneity of variances *will* potentially affect either the Type I error rate or the Type II error rate of the test, depending on the nature of the heterogeneity.¶

Effects of heterogeneity (univariate)

In classical univariate ANOVA, in cases where there is heterogeneity of variances and the design is unbalanced, then:

- if there is *greater dispersion* in one or more groups that have a *small sample size*, then the tendency will be to *inflate the Type I error* of the ANOVA test;
- if there is *greater dispersion* in one or more groups that have a *large sample size*, then the tendency will be to *inflate the Type II error* of the ANOVA test.

For more details of these effects, see Welch (1938) , Horsnell (1953) , Box (1954) and Glass *et al.* (1972) .

Permutation tests do not provide a solution to this issue. They, too, are sensitive to differences in dispersion; groups with different dispersions cannot strictly be considered to be 'exchangeable' under a true null hypothesis of no differences in means (e.g., see Boik (1987) and Hayes (1996)).

What is needed is *a method for testing differences in means when variances differ*.

¶*Recall that:*

- *the probability of a Type I error is the probability of rejecting H_0 when it is true; and*
- *the probability of a Type II error is the probability of failing to reject H_0 when it is false.*

† *By an exact test, we mean that the Type I error of the test is exactly equal to the a priori chosen significance level of the test. For an exact test, if you choose a significance level of (say) 0.05, and you reject the null hypothesis any time you get a p-value less than or equal to 0.05, then the probability that you will reject a true null hypothesis is indeed precisely 0.05; that is, you will be wrong 5% of the time.*

Revision #140
Created 31 March 2025 22:44:10 by Marti
Updated 30 November 2025 01:33:11 by Marti