

4.1: How is ANOVA Calculated?

In the past lessons, we carried out the ANOVA computations conceptually in terms of deviations from means. For the calculation of total variance, we used the deviations of the individual observations from the overall mean, while the treatment SS was calculated using the deviations of treatment level means from the overall mean, and the residual or error SS was calculated using the deviations of individual observations from treatment level means. In practice, however, to achieve higher computational efficiency, SS for ANOVA is computed utilizing the following mathematical identity:

$$SS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N} \quad (4.1.1)$$

This identity is commonly called the **working formula** or **machine formula**. The second term on the right-hand side is often referred to as the **correction factor (CF)**.

For computing the SS for the total variance of the responses, the formula above can be used as it is, but modifications need be made for others. For example, to compute the treatment SS, the above equation has to be modified as:

$$SS_{treatment} = \sum_{i=1}^T \frac{(\sum_{j=1}^{n_i} Y_{ij})^2}{n_i} - \frac{(\sum Y_i)^2}{N} \quad (4.1.2)$$

We will examine three new ANOVA models (Models 1, 2, 3), as well as the effects model (Model 4) from the previous lesson, defined as follows:

Model 1 - The Overall Mean Model

$$Y_{ij} = \mu + \epsilon_{ij} \quad (4.1.3)$$

which simply fits an overall or "grand" mean'. This model reflects the situation where H_0 is true, implying that $\mu_1 = \mu_2 = \dots = \mu_T$.

Model 2 - The Cell Means Model

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad (4.1.4)$$

where μ_i , $i = 1, 2, \dots, T$ are the factor level means. Note that in this model, there is no overall mean being fitted.

Model 3 - Dummy Variable Regression

$$Y_{ij} = \mu + \mu_i + \epsilon_{ij}, \text{ fitted as } Y_{ij} = \beta_0 + \beta_{Level\ 1} + \beta_{Level\ 2} + \dots + \beta_{Level\ r-1} + \epsilon_{ij} \quad (4.1.5)$$

where $\beta_{Level\ 1}, \beta_{Level\ 2}, \dots, \beta_{Level\ T-1}$ are regression coefficients for $T - 1$ indicator-coded regression "dummy" variables that are correspond to the $T - 1$ categorical factor levels. The T^{th} factor level mean is given by the regression intercept β_0 .

Model 4 - The Effects Model

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad (4.1.6)$$

where τ_i are the the deviations of each factor level mean from the overall mean so that $\sum_{i=1}^T \tau_i = 0$.

Each of these four models can be written as a **general linear model (GLM)**: $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$ simply by changing the design matrix \mathbf{X} . Thus to perform the data analysis, in terms of the computer coding instructions, the appropriate numerical values for the \mathbf{X} matrix elements will need to be inputted.

This page titled 4.1: How is ANOVA Calculated? is shared under a [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](https://stats.libretexts.org/@go/page/33489).