Print

Applying the Two-Way ANOVA Model

Let's review a few modeling terms. ANOVA and regression are used to estimate parameters in statistical models. A model is simply the mathematical relationship between predictor variables and the response variable, and the same model can be expressed in a variety of ways. The independent variables in ANOVA are called predictors or factors. A factor is simply an explanatory categorical variable with two or more levels.

In a two-way ANOVA, there are two variables, or factors, which affect the response variable. Each factor has multiple levels. Treatments, or treatment groups, are formed by combining the two factors. If the first factor has three levels, and the second factor has four levels, then there are three times four, or twelve treatment groups.

The term effect refers to the expected change in the response variable due to the change in value of a predictor variable. That is, if we change the value of a single variable, holding everything else constant, the response changes accordingly given the model equation. The variables themselves are often referred to as effects in the model.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} ...$$

A main effect is the effect of a single predictor variable, such as x1, x2, or x3.

Interaction effects occur when the relationship between the response and a predictor changes according to another predictor in the model. These effects are coded and built into the model simply by crossing effects (for example, x1 by x2 or x1 crossed with x2 and x3).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{1i} X_{2i} ...$$

When you use two-way ANOVA, you examine the effects of your two factors concurrently to determine whether they interact with respect to their effects on the response variable. This possible interaction means that the effect of one factor depends on the value of the other factor.

Here's the two-way ANOVA model.

$$Y_{ijk} = \mu + a_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

With each additional predictor variable, a new parameter is introduced. Notice that beta has been introduced into the model as the second factor and now our subscript has three unique levels. Let's apply the model to our Ames Housing data in order to consider the effect of heating system quality and season sold on home sale prices.

Y_{ijk} is the response variable, SalePrice.

μ, the overall population mean of the response, is the average sale price of all homes sold.

α_i is the effect of the ith heating system category. This is the difference between the overall population mean of the i^{th} heating system category and the overall mean, μ.

β_j is the effect of the j^{th} season sold. This is the difference between the population mean of the j^{th} season sold and the overall mean, μ.

$(\alpha\beta)_{ij}$ is the effect of the interaction between the i^{th} heating system category and the j^{th} season. Interaction terms are also called product terms or crossed effects. Notice that the interaction terms involves both main effects, Heating_QC and Season_Sold.

ε_{ijk} is the error term or residual in your model.

As with one-way ANOVA, there are three assumptions. The observations are independent. The data is normally distributed. The population variances are equal for each treatment combination.

The null hypothesis for two-way ANOVA without interaction is that none of the effects in the model are statistically different. That is, no differences exist among the group means. When testing main effects only, we're looking for differences in sale price means among the four seasons, or among the four heating quality levels.

$$H_0 : a_i = 0$$

and

$$\beta_j = 0$$

In a model with interactions, the null hypothesis is that no differences exist among the 16 different combinations of quality and season.

$$H_0 : a_i = 0$$

and

$$\beta_j = 0$$

and

$$(\alpha\beta)_{ij} = 0$$

Close