

### 3.1: The Model

The effects model for one way ANOVA is a linear additive statistical model which relates the response to the treatment and can be expressed as

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad (3.1.1)$$

where  $\mu$  is the grand mean,  $\tau_i$  ( $i = 1, 2, \dots, T$ ) are the deviations from the grand mean due to the treatment levels and  $\epsilon_{ij}$  are the error terms. The quantities  $\tau_i$  ( $i = 1, 2, \dots, T$ ) which add to zero, are also referred to as the treatment level effects and the errors show the amount "left over" after considering the grand mean and the effect of being in a particular treatment level.

Here's the analogy in terms of the greenhouse experiment. Think of someone who is not aware that different fertilizers have been used walking into the greenhouse to simply inquire about plant heights in general. The overall sample mean, an estimate of the grand mean, will be a suitable response to this inquiry. On the other hand, the overall mean would not be satisfactory to the experimenter of the study, who obviously suspects that there will be height differences among different fertilizer types. Instead, what is more acceptable to the experimenter are the plant height estimates after including the effect of the treatment  $\tau_i$ .

#### Note

The actual plant height can never be known because there is an unknown measurement error associated with any observation. This unknown error is associated with the  $i$ th treatment level, and the  $j$ th observation is denoted  $\epsilon_{ij}$  ( $i = 1, 2, \dots, T$ ,  $j = 1, 2, \dots, n_i$ ) is a random component (noise) that reflects the unexplained variability among plants within treatment levels.

**Under the null hypothesis** where the treatment effect is zero, **the reduced model** can be written  $Y_{ij} = \mu + \epsilon_{ij}$ .

**Under the alternative hypothesis**, where the treatment effects are not zero, **the full model** for at least one treatment level can be written  $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$ .

If  $SSE(R)$  denotes the error sums of squares associated with the reduced model and  $SSE(F)$  denotes the error sums of squares associated with the full model, we can utilize the General Linear Test approach to test the null hypothesis by using the test statistic:

$$F = \frac{\left( \frac{SSE(R) - SSE(F)}{df_R - df_F} \right)}{\left( \frac{SSE(F)}{df_F} \right)} \quad (3.1.2)$$

which under the null hypothesis has an  $F$  distribution with the numerator and denominator degrees of freedom equal to  $df_R - df_F$  and  $df_F$  respectively, where  $df_R$  is the degrees of freedom associated with  $SSE(R)$  and  $df_F$  is the degree of freedom associated with  $SSE(F)$ . It is easy to see that  $df_R = N - 1$  and  $df_F = N - T$  where  $N = \sum_{i=1}^N n_i$ . Also,

$$SSE(R) = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = SS_{Total} \quad \text{See Section 2.2} \quad (3.1.3)$$

Therefore,

$$F = \frac{\left( \frac{SS_{Total} - SSE}{T - 1} \right)}{\left( \frac{SSE}{df_{Error}} \right)} \quad (3.1.4)$$

$$= \frac{\left( \frac{SS_{Treatment}}{df_{Treatment}} \right)}{\left( \frac{SSE}{df_{Error}} \right)} \quad (3.1.5)$$

$$= \frac{MS_{Trt}}{MSE} \quad (3.1.6)$$

Note that this is the same test statistic derived in Section 2.2 for testing the treatment significance. If the null hypothesis is true, then the treatment effect is not significant. If we reject the null hypothesis, then we conclude that the treatment effect is significant, which leads to the conclusion that at least one treatment level is better than the others!

---

This page titled [3.1: The Model](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#).