Models & Matrices in One-Way Analysis of Variance

The analysis of variance is a linear model because it can be written in the form $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, Where $\mathbf{Y}$ is the random response vector, $\mathbf{X}$ is the design matrix, β is the coefficient vector, and $\boldsymbol{\varepsilon}$ is the error vector. (The forms of the design matrix and coefficient vector depend on the model being used.) As in regression, the least square estimate of the coefficients is given by $\hat{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathrm{Y}$, and the vector of fitted values (which estimates the mean response at different treatments) is $\hat{\mathrm{Y}} = \mathbf{X}\hat{\beta}$.

A Tale of Two Models

Two equivalent models are commonly employed in one-way ANOVA: The **Cell Means Model**, and the **Factor Effects Model**. We'll compare the two models, and we'll investigate their matrix representations.

Cell Means Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$

The following is meant to explain the notation used in the cell means model: As this is one-way analysis of variance, I will refer to the levels of the factor variable as treatments. The cell means model is the model we've primarily used for one-way ANOVA in this course.

- n_i , $1 \le i \le k$, is the size of the sample drawn from the i^{th} treatment.
- $n = \sum n_i$ is the total sample size drawn from all treatments.
- μ_i : The mean value of the quantitative dependent (response) variable under the i^{th} treatment.
- $\boldsymbol{Y}_{ij}$: The value of the response variable for the j^{th} observation from the i^{th} treatment in the sample.
- $\boldsymbol{\varepsilon}_{ij}$: The difference between $\boldsymbol{Y}_{ij}$ and μ_i , i.e., $\boldsymbol{\varepsilon}_{ij} = \boldsymbol{Y}_{ij} - \mu_i$. Note: The assumption made in ANOVA is that the errors are independent and identically distributed as $N\left(0,\sigma^2\right)$.

For the j^{th} observation on the i^{th} treatment, the cell means model is given by $\boxed{\boldsymbol{Y}_{ij} = \mu_i + \boldsymbol{\varepsilon}_{ij}}$. Since the treatment mean μ_i is a constant, we have $E\left(\mathbf{Y}_{ij}\right) = \mu_i$ and $\sigma^2_{\mathbf{Y}_{ij}} = \sigma^2$. Finally, this leads to the conclusion that, under the assumptions made about the distribution of the errors, the $\boldsymbol{Y}_{ij}$ are independent $N\left(\mu_i, \sigma^2\right)$.

Matrix Representation: Cell Means Model

To demonstrate the method, we'll use the following "toy" data set with two observations on each of three levels (treatments) of the factor.

Factor Level	A	A	B	B	C	C
y	11	19	32	42	22	30

In the cell means model, we'll create dummy variables, one for each of the three treatments (levels) to make up the three columns of the design matrix. Notice that we don't leave a dummy undefined as we would if we were using dummy variables in a regression model. The resulting matrix is $\mathbf{X} = \begin{array}{c} \begin{matrix} A & B & C \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$,

where the least squares solution $\hat{\beta}$ is the solution to $\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T Y$. Solving, as in regression,

- $\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Note: Diagonals are treatment sample size.

- $\mathbf{X}^T Y = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 11 \\ 19 \\ 32 \\ 42 \\ 22 \\ 30 \end{bmatrix} = \begin{bmatrix} 30 \\ 74 \\ 52 \end{bmatrix}$, where the entries of $\mathbf{X}^T Y$ are treatment totals $\sum_{j=1}^{n_s} Y_{ij}$.

- $\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$. (Diagonal matrices have simple inverses.)

- $\hat{\beta} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T Y = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 30 \\ 74 \\ 52 \end{bmatrix} = \begin{bmatrix} 15 \\ 37 \\ 26 \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \end{bmatrix}$. The estimated treatment means are just the sample means for the three treatments, which is intuitively satisfying.

Factor Effects Model: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

The following is meant to explain the notation used in the factor effects model:

- μ: Under the most common interpretation, this is the unweighted average of all treatment means, i.e., $\mu = \frac{\sum \mu_i}{k}$. (In some situations, μ may be defined as a weighted average of the treatment means.)

- α_i : The effect of the i^{th} treatment upon the mean. If μ is the unweighted mean of all treatment means, then $\sum \alpha_i = 0$. (The α_i represent "offsets" from the overall mean.)

For the j^{th} observation on the i^{th} treatment, the factor effects model is given by $\boxed{\mathbf{Y}_{ij} = \mu + \alpha_i + \boldsymbol{\varepsilon}_{ij}}$. Since the treatment mean $\mu + \alpha_i$ is a constant, we have $E\left(\mathbf{Y}_{ij}\right) = \mu + \alpha_i$ and $\sigma^2_{\mathbf{Y}_{ij}} = \sigma^2$. Finally, this leads to the conclusion that, under the assumptions made about the distribution of the errors, the $\mathbf{Y}_{ij}$ are independent $N\left(\mu + \alpha_i, \sigma^2\right)$.

Matrix Representation: Factor Effects Model

To demonstrate the method, we'll return to the "toy" data set with two observations on each of three levels (treatments) of the factor.

Factor Level	A	A	B	B	C	C
y	11	19	32	42	22	30

The Factor Effects Model is sometimes thought of as being more "regression-like". The first column of the design matrix is a column of ones (similar to the first column in a regression analysis). Then, for k treatments we'll create $k-1$ variables as we would if we were using dummy variables in a regression model, but with a twist. These dummy variables have the special form shown below,

$X_{ij} = \begin{cases} 1 & \text{,if the } i^{\text{th}} \text{ factor is involved} \\ 0 & \text{,if another factor is involved} \\ -1 & \text{,if the } k^{\text{th}} \text{ factor is involved} \end{cases}$. For the "toy" dataset above, the resulting design matrix is

$$\begin{array}{cc} & \begin{array}{ccc} 1 & X_1 & X_2 \end{array} \\ \mathbf{X} = & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \end{array}$$

, where the least squares solution $\hat{\beta}$ solves the equation $\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\beta} = \mathbf{X}^{\mathrm{T}}\mathrm{Y}$ as before.

Solving, as in regression,

- $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 4 \end{bmatrix}$.

- $$\mathbf{X}^{\mathrm{T}}\mathrm{Y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 11 \\ 19 \\ 32 \\ 42 \\ 22 \\ 30 \end{bmatrix} = \begin{bmatrix} 156 \\ -22 \\ 22 \end{bmatrix}$$, where the first entry of $\mathbf{X}^{\mathrm{T}}\mathrm{Y}$ is $\sum_{i=1}^{3}\sum_{j=1}^{2} Y_{ij}$.

- $$\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 1/6 & 0 & 0 \\ 0 & 1/3 & -1/6 \\ 0 & -1/6 & 1/3 \end{bmatrix}$$. ($\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is a Block Diagonal matrix.)

- $$\hat{\beta} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathrm{Y} = \begin{bmatrix} 1/6 & 0 & 0 \\ 0 & 1/3 & -1/6 \\ 0 & -1/6 & 1/3 \end{bmatrix} \begin{bmatrix} 156 \\ -22 \\ 22 \end{bmatrix} = \begin{bmatrix} 26 \\ -11 \\ 11 \end{bmatrix} = \begin{bmatrix} \bar{\bar{Y}} \\ \bar{Y}_1 - \bar{\bar{Y}} \\ \bar{Y}_2 - \bar{\bar{Y}} \end{bmatrix}$$, where $\bar{\bar{Y}}$ estimates the overall mean μ, and $\bar{Y}_1 - \bar{\bar{Y}}$ estimates the effect of treatment 1, $\alpha_1 = \mu_1 - \mu$. Similarly, $\bar{Y}_1 - \bar{\bar{Y}}$ estimates the effect of treatment 2, $\alpha_2 = \mu_2 - \mu$.

Note that the two models produce similar results, they simple emphasize different aspects of the problem. For instance, the produce the same estimated treatment means.

$$\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \end{bmatrix} = \begin{bmatrix} 15 \\ 37 \\ 26 \end{bmatrix} = \begin{bmatrix} 26-11 \\ 26+11 \\ 26-0 \end{bmatrix} = \begin{bmatrix} \bar{\bar{Y}} + \hat{\alpha}_1 \\ \bar{\bar{Y}} + \hat{\alpha}_2 \\ \bar{\bar{Y}} + \hat{\alpha}_3 \end{bmatrix}$$