

Univariate 1-Way ANOVA as a Linear Model with Fixed Regressors

Group 1	Group 2	Group 3
x_{11}	x_{12}	x_{13}
x_{21}	x_{22}	x_{23}
x_{31}	x_{32}	x_{33}

Linear Model

$$x_{i,j} = \mu + \alpha_j + \varepsilon_{i,j} \quad (1)$$

Reparameterized Linear Model

$$x_{i,j} = \mu_j + \varepsilon_{i,j} \quad (2)$$

Matrix Form

$$\mathbf{x} = \mathbf{D}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (3)$$

D is a “design matrix” with 1’s and 0’s, and

$$\boldsymbol{\mu}' = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_J \quad \mu] \quad (4)$$

Example.

$$\begin{bmatrix} x_{1,1} \\ x_{2,1} \\ x_{3,1} \\ x_{1,2} \\ x_{2,2} \\ x_{3,2} \\ x_{1,3} \\ x_{2,3} \\ x_{3,3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,1} \\ \varepsilon_{2,1} \\ \varepsilon_{3,1} \\ \varepsilon_{1,2} \\ \varepsilon_{2,2} \\ \varepsilon_{3,2} \\ \varepsilon_{1,3} \\ \varepsilon_{2,3} \\ \varepsilon_{3,3} \end{bmatrix} \quad (5)$$

Notice that the \mathbf{D} matrix is not of full rank. (Why not? C.P.)

\mathbf{D} is of rank 3. We can partition \mathbf{D} into two submatrices, one of which is full rank, and the other of which contains columns that are “superfluous” in the sense that they are linearly dependent.

$$\mathbf{D} = [\mathbf{D}_1 \quad \mathbf{D}_2] \quad (6)$$

We can similarly partition

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad (7)$$

Note that we could redefine $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu}^* = \begin{bmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad (8)$$

In which case

$$\mathbf{x} = \mathbf{D}_1 \boldsymbol{\mu}^* + \boldsymbol{\varepsilon} \quad (9)$$

The null hypothesis in ANOVA can be expressed as a “general linear hypothesis” of the form

$$\mathbf{H}'\boldsymbol{\mu}^* = \mathbf{0} \quad (10)$$

Let's try it.

Suppose

$$\mathbf{H}' = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad (11)$$

Then

$$\mathbf{H}'\boldsymbol{\mu}^* = \begin{bmatrix} \alpha_1 + \mu - \alpha_3 - \mu \\ \alpha_2 + \mu - \alpha_3 - \mu \end{bmatrix} = \begin{bmatrix} \alpha_1 - \alpha_3 \\ \alpha_2 - \alpha_3 \end{bmatrix} \quad (12)$$

Notice that if both elements of the expression on the right in Equation (12) are equal, then all 3 α_j must be equal, and the null hypothesis of equal means must be true.

Notice also that there are infinitely many ways you can write a “hypothesis matrix” \mathbf{H} that expresses

the same null hypothesis. All of them will be of rank 2, or, in the case of J groups, $J - 1$.

Let \mathbf{H}' be of order $p \times q$, and let \mathbf{x} have N elements. (Note that this N is the *total* N in the case of ANOVA.)

Under H_0 , the following statistic has an F distribution:

$$F_{p, N-q} = \frac{\mathbf{x}' \mathbf{D}_1 (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{H} \left[\mathbf{H}' (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{H} \right]^{-1} \mathbf{H}' (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \mathbf{x}}{p} \quad (13)$$

$$\frac{\mathbf{x}' \left[\mathbf{I} - \mathbf{D}_1 (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \right] \mathbf{x}}{N - q}$$

This looks incomprehensible at first glance! Let's examine the formula with reference to our simple 3 group ANOVA example, and let's assume that sample sizes are equal in all groups.

Look closely at the denominator. Note that it is the complementary projection operator for the column

space of \mathbf{D}_1 . So the entire denominator can be written as $\mathbf{x}'\mathbf{Q}_{\mathbf{D}_1}\mathbf{Q}_{\mathbf{D}_1}\mathbf{x} = \mathbf{x}^{*'}\mathbf{x}^*$. So it is a sum of squares of the projection of \mathbf{x} into the space orthogonal *simultaneously* to all the columns of \mathbf{D}_1 . If you study \mathbf{D}_1 , you will see that in order for \mathbf{x} to simultaneously be orthogonal to all the columns of \mathbf{D}_1 , each group of 3 scores must have a sum of zero for its group. So this is just a fancy way of computing *mean square within*.

Now consider the matrix expression in the numerator. Recognizing that

$$\left(\mathbf{D}_1' \mathbf{D}_1\right)^{-1} \mathbf{D}_1' \mathbf{x} = \bar{\mathbf{x}} \quad (14)$$

On the left and the right, we have expressions of the form $\bar{\mathbf{x}}'$ and $\bar{\mathbf{x}}$. Now, with equal n per group, it is easy to see that

$$\mathbf{D}_1' \mathbf{D}_1 = n\mathbf{I} \quad (15)$$

and so we can write the entire numerator matrix expression as

$$\bar{\mathbf{x}}' \mathbf{H} \left(\mathbf{H}' n^{-1} \mathbf{I} \mathbf{H} \right)^{-1} \mathbf{H}' \bar{\mathbf{x}} = n \bar{\mathbf{x}}' \mathbf{P}_{\mathbf{H}} \bar{\mathbf{x}} \quad (16)$$

At this point, deciphering the meaning of Equation (16) requires us to take a step back and look at $\mathbf{P}_{\mathbf{H}}$. It is a column space projector for \mathbf{H} , which has two columns in 3 dimensional space. But note that these two columns are *both orthogonal to the unit vector $\mathbf{1}$* , and are linearly independent.

Consequently, they span the space orthogonal to $\mathbf{1}$, and so it must be that $\mathbf{P}_{\mathbf{H}} = \mathbf{Q}_{\mathbf{1}} = \mathbf{I} - \mathbf{P}_{\mathbf{1}} = \mathbf{I} - \frac{\mathbf{1}\mathbf{1}'}{\mathbf{1}'\mathbf{1}}$.

Consequently, the matrix quantity in Equation (16)

is simply n times the sum of squared deviations of the sample means around their overall mean. Since p in equation (13) is one less than the number of groups, we recognize that the entire expression is simply

$$F = \frac{nS_{\bar{x}}^2}{\hat{\sigma}^2} \quad (17)$$

Two-Way ANOVA

Notice that in the model $\mathbf{x} = \mathbf{D}_1\boldsymbol{\mu}^* + \boldsymbol{\varepsilon}$, all the scores in the experiment are in a single vector \mathbf{x} , and in the hypothesis statement of Equation (10), all the means are in a single vector. Moreover, the F statistic depends on the data, N , and just two matrix quantities, i.e., the hypothesis matrix \mathbf{H} and the design matrix \mathbf{D}_1 . The design matrix is a simple function of the linear model in scalar form (i.e., Equation (2)). However, it is less obvious how to

construct the hypothesis matrices for row, column, and interaction, effects. We need a heuristic!

Constructing the Hypothesis Matrix

For simplicity, suppose we have a 2×2 ANOVA, with 2 observations per cell. The cell means are

$$\mathbf{U} = \begin{bmatrix} \mu_{1,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{2,2} \end{bmatrix} \quad (18)$$

Notice that it is easy to write the null hypothesis in the form $\mathbf{AUB} = \mathbf{0}$. We can write the null hypothesis for rows as

$$\begin{aligned} H_0 : & \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{1,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{2,2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ & = \mu_{1,1} + \mu_{1,2} - (\mu_{2,1} + \mu_{2,2}) = 0 \end{aligned} \quad (19)$$

This can be written

$$H_0 : \mathbf{OU1} = \mathbf{0} \quad (20)$$

We call \mathbf{O} an *omnibus contrast matrix*, and it is always of the same basic form, i.e.,

$$\mathbf{O} = [\mathbf{I} \mid -\mathbf{1}] \quad (21)$$

(Examples. C.P.)

In a similar vein, we can write the column effect null hypothesis as

$$\begin{aligned} H_0 &: [1 \quad 1] \begin{bmatrix} \mu_{1,1} & \mu_{1,2} \\ \mu_{2,1} & \mu_{2,2} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \mu_{1,1} + \mu_{2,1} - (\mu_{1,2} + \mu_{2,2}) = 0 \end{aligned} \quad (22)$$

This is of the form

$$H_0 : \mathbf{1}'\mathbf{U}\mathbf{O}' = \mathbf{0} \quad (23)$$

The interaction null hypothesis is that there are no differences of differences, and consequently is of the form

$$H_0 : \mathbf{O}\mathbf{U}\mathbf{O}' = \mathbf{0} \quad (24)$$

(C.P. Write the null hypotheses for a 3×2 Anova row effect.)

(C.P. Write the null hypothesis for a simple main effect for columns at level 1 of the row effect.)

As straightforward as this system is, unfortunately the null hypothesis requires all the means in a single vector μ , not in a matrix, and there is only *one* hypothesis matrix, not two as in the examples above. So how do we proceed?

First, we need to define the *Kronecker Product*.

Kronecker (Direct) Product

The *Kronecker product* of two matrices \mathbf{A} and \mathbf{B} is denoted $\mathbf{A} \otimes \mathbf{B}$, and can be written in partitioned form as

$${}_p \mathbf{A}_q \otimes {}_r \mathbf{B}_s = \begin{bmatrix} a_{1,1} \mathbf{B} & a_{1,2} \mathbf{B} & \cdots & a_{1,q} \mathbf{B} \\ a_{2,1} \mathbf{B} & a_{2,2} \mathbf{B} & \cdots & a_{2,q} \mathbf{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p,1} \mathbf{B} & a_{p,2} \mathbf{B} & \cdots & a_{p,q} \mathbf{B} \end{bmatrix} \quad (25)$$

The Kronecker product above is of order $pq \times rs$.

Kronecker products have some interesting properties. Let $\text{vec}_c(\mathbf{X})$ be a column vector consisting of the columns of \mathbf{X} stacked on top of each other. Let $\text{vec}_r(\mathbf{X})$ be a column vector consisting of the rows of \mathbf{X} transposed into columns and stacked on top of each other. Then

$$\text{vec}_c(\mathbf{BSA}') = (\mathbf{A} \otimes \mathbf{B}) \text{vec}_c(\mathbf{S}) \quad (26)$$

and, since $\text{vec}_r(\mathbf{S}') = \text{vec}_c(\mathbf{S})$, we also have

$$\text{vec}_r(\mathbf{ASB}') = (\mathbf{A} \otimes \mathbf{B}) \text{vec}_r(\mathbf{S}) \quad (27)$$

$$\text{vec}_r(\mathbf{ASB}) = (\mathbf{A} \otimes \mathbf{B}') \text{vec}_r(\mathbf{S}) \quad (28)$$

Consider the null hypothesis of Equation (20). This can be written as $H_0 : \mathbf{OU1} = \mathbf{0}$, or as

$$H_0 : (\mathbf{O} \otimes \mathbf{1}') \text{vec}_r(\mathbf{U}) = \mathbf{0} \quad (29)$$

Example. Consider the null hypothesis of no row effect in the 2×2 ANOVA.

$$\begin{aligned}
 (\mathbf{0} \otimes \mathbf{1}')\boldsymbol{\mu} &= ([1 \quad -1] \otimes [1 \quad 1]) \begin{bmatrix} \mu_{1,1} \\ \mu_{1,2} \\ \mu_{2,1} \\ \mu_{2,2} \end{bmatrix} \\
 &= [1 \quad 1 \quad -1 \quad -1] \begin{bmatrix} \mu_{1,1} \\ \mu_{1,2} \\ \mu_{2,1} \\ \mu_{2,2} \end{bmatrix}
 \end{aligned} \tag{30}$$

Equation (30) gives $(\mu_{1,1} + \mu_{1,2}) - (\mu_{2,1} + \mu_{2,2})$,
which is the quantity that is zero when there is no
row main effect.

General Specification of the Hypothesis Matrix in Factorial Anova

1. Call the effects A, B, C etc.
2. A conformable *Omnibus Matrix* for effect A is denoted \mathbf{O}_A .
3. A conformable *Summing Vector* for effect A is denoted $\mathbf{1}'_A$.
4. A *selection vector* $\mathbf{s}'_{A,j}$ is a row vector conformable with factor A with a 1 in position j , and zeroes elsewhere.

1. To construct an **H** for a main effect, use an **O** matrix for that effect and a summing vector for all other effects.
2. To construct an **H** for an interaction effect, use an **O** matrix for all effects in the interaction, and a summing vector for all effects not in the interaction.
3. To construct the **H** for a simple main effect, use selection vector(s) to select the levels for the

simple main, then use an \mathbf{O} matrix for the main effect tested.

Examples. 2-way ($A \times B$) ANOVA

A Main Effect $\mathbf{O}_A \otimes \mathbf{1}'_B$

B Main Effect $\mathbf{1}'_A \otimes \mathbf{O}_B$

AB Interaction Effect $\mathbf{O}_A \otimes \mathbf{O}_B$

Simple Main Effect of A at Level 1 of B $\mathbf{O}_A \otimes \mathbf{s}'_{B,1}$

Examples. 3-way (A × B × C) ANOVA

ABC Interaction Effect $\mathbf{O}_A \otimes \mathbf{O}_B \otimes \mathbf{O}_C$

A Main Effect $\mathbf{O}_A \otimes \mathbf{1}'_B \otimes \mathbf{1}'_C$

Now that we understand how to express univariate ANOVA in matrix notation, we move on to MANOVA.