TWO WAY ANOVA

Next we consider the case when we have two factors, categorizations, e.g. lab and manufacturer. If there are I levels in the first factor and J levels in the second factor then we can think of this situation as one where there are $I \times J$ levels of the combined factors.

Notation

Notation-wise, we simply add another subscript to the response, that is, y now has a triple subscript, where $y_{i,j,k}$ represents the measurement on the kth subject that belongs to both the ith group (lab) of the first factor and the j group (manufacturer) of the second factor, $i = 1, \ldots I$, $j = 1, \ldots, J$, and $k = 1, \ldots, n_{ij}$.

For simplicity, we will only work with the special case of $n_{i,j} = K$, i.e. all subgroups have the same number of responses. Then we write the model as follows:

$$y_{i,j,k} = \alpha + \eta_{i,j} + E_{i,j,k}$$

So, when $i = 1$ and $j = 1$, $y_{1,1,k} =$ ---------- $+ E_{1,1.k}$
and when $i = 2$ and $j = 1$, $y_{2,1,k} =$ ---------- $+ E_{2,1.k}$
Again, we need a constraint because our model is over-parameterized. We add the constraint that $\sum_i \sum_j \eta_{i,j} = 0$.

A Simpler Sub-model

In our example of the study of the measurement process, we find that with 7 labs and 4 manufacturers, we have 28 levels. If the effect of the lab is the same, regardless of which manufacturer the tablets are coming from, and if the effect of the manufacturer is the same regardless of which lab is measuring the tablets then we could express the model as

$$y_{i,j,k} = \alpha + \beta_i + \gamma_j + E_{i,j,k}$$

Note that now we have only $I + J$ levels, rather than $I \times J$. This model is called an *additive* model. It puts structure on the levels. That is the difference between measurements at LAb 1 and Lab 2 of tablets from Manufacturer A is $\beta_2 - \beta_1$, and this difference is the same for the measurements at Labs 1 and 2 for tablets from Manufacturer B, i.e. there is no interaction between lab and manufacturer.

Degrees of Freedom

To see that the additive model is a sub-model of the full model, we can we express the full model as follows:

$$y_{i,j,k} = \alpha + \beta_i + \gamma_j + \nu_{i,j} + E_{i,j,k}$$

Now again, we need to put constraints on the parameterization. If we think about it from the geometric perspective, we see that the $\mathbf{1}$ vector lies in both the space spanned by the lab indicators (the $\mathbf{e}_i$) and the space spanned by the manufacturer indicators (the $\mathbf{u}_j$). So, the $\mathbf{1}$ vector, and $I - 1$ of the $\mathbf{e}_i$ vectors and $J - 1$ of the $\mathbf{u}_j$ vectors are all that is needed for the additive part of the model. As for the rest, suppose we have vectors $\mathbf{v}_{i,j}$ that indicate whether a response belongs in group i, j or not.

Note that $\sum_j \mathbf{v}_{i,j} =$ ________, and
and that $\sum_i \mathbf{v}_{i,j} =$ ________.
So we need only __________ of these $I \times J$ vectors.
All together that gives us $1 + (I - 1) +$ (_________) + (___________) = (___________) of the $1 + I + J + IJ$ vectors.

If we are to put all of the parameters in then we must add constraints. Traditionally these are

$$\begin{aligned} \sum_i \beta_i &= 0 \\ \sum_j \gamma_j &= 0 \end{aligned}$$

$$
\begin{aligned}
\sum_i \text{\_\_\_\_\_\_\_\_\_\_\_} &= 0, for \\
\sum_j \text{\_\_\_\_\_\_\_\_\_\_\_} &= 0, for
\end{aligned}
$$

How many constraints do we have? _____________

Sums of Squares

The Anova table of the sums of squared deviations helps us assess whether the simple additive model is adequate to describe the variation in the means, and whether there is a lab effect or a manufacturer effect (i.e. whether all of the $\beta_i = 0$ or all of the $\gamma_j = 0$).

The decomposition of the sums of squares is a bit more complex here. First we need to introduce some more notation,

$$
\begin{aligned}
\bar{y}_{..} &= \frac{1}{IJK} \sum_i \sum_j \sum_k y_{ijk} \\
\bar{y}_{i.} &= \frac{1}{JK} \sum_j \sum_k y_{ijk}, \text{ for } i = 1 \ldots I \\
\bar{y}_{.j} &= \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_} \\
\bar{y}_{ij} &= \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_}
\end{aligned}
$$

Now let's look at the sums of squares:

$$
\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{..})^2
$$

To begin, let's add and subtract the IJ means $\bar{y}_{ij}$.

$$\sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{..})^2 = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij})^2 + \sum_i \sum_j \sum_k (\bar{y}_{ij} - \bar{y}_{..})^2$$

Show that the cross product term is 0.

We call the first sum on the right-hand side of the equation the error sum of squares, or SS_E. We want to further decompose the second term.

$$\sum_i \sum_j \sum_k (\bar{y}_{ij} - \bar{y}_{..})^2$$

What do we add and subtract – $\bar{y}_{i.}$ or $\bar{y}_{.j}$? Both:

$$\begin{aligned}
\sum_i \sum_j \sum_k & \quad (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} + \bar{y}_{i.} - \bar{y}_{..} + \bar{y}_{.j} - \bar{y}_{..})^2 \\
&= \sum_i \sum_j K(\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \\
&+ \sum_i JK(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_j IK(\bar{y}_{.j} - \bar{y}_{..})^2
\end{aligned}$$

The three terms on the right-hand side of the equality are called, the interaction sum of squares, or SS_{LM}, the sum of squares due to Lab, or SS_L, and the sum of squares due to Manufacturer, or

SS_M.

Show that the cross products are all 0.

ANOVA Table

Arrange the sum of squares into an ANOVA table.

Source	DF	Sum of Squares	Mean Square	F-statistic
Labs				
Manufacturer		3		
Interaction		8		
Error		60		
Total		85		

The first F statistic is used to test whether there is a difference between labs, i.e. whether there is a lab effect. The second F statistic is used to test whether there is a difference between manufactureres. The third is to test the additive model, i.e. is there an interaction between lab and manufacturer.