Displays for Statistics 5303

Lecture 31

November 15, 2002

Christopher Bingham, Instructor

612-625-7023 (St. Paul)
612-625-1024 (Minneapolis)

Class Web Page
http://www.stat.umn.edu/~kb/classes/5303

Mixed effects (continued)

A model with two or more factors may have both fixed and random factors.

Any interaction involving a random factor will be a random effect, too, even if some or all of the other factors in the interaction are fixed.

Thus, when A is a fixed factor and B is a random factor, in the two-factor model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

all β_j, j = 1,...,b and $\alpha\beta_{ij}$. i = 1,..., a, j = 1,...,b are random variables.

The mean or expectation of any random effect is 0, so $E(\alpha\beta_{ij}) = 0$, even when A is fixed.

There are three variance components here σ_β^2, $\sigma_{\alpha\beta}^2$ and $\sigma^2 = \sigma_\varepsilon^2$.

Restrictions on mixed effects

For a purely fixed effect, the usual restrictions are that the effects sum to 0 over each subscript.

That means, when A, B and C are fixed

- $\sum_i \alpha_i = 0$ (main effect).
- $\sum_i \alpha\beta_{ij} = 0$ (sum over levels of A)
  $\sum_j \alpha\beta_{ij} = 0$ (sum over levels of B)
- $\sum_i \alpha\beta\gamma_{ijk} = 0$ (sum over levels of A)
  $\sum_j \alpha\beta\gamma_{ijk} = 0$ (sum over levels of B)
  $\sum_k \alpha\beta\gamma_{ijk} = 0$ (sum over levels of C)

For mixed interactions, the situation is somewhat more complicated.

There are two different approaches which depend on what you think a random interaction is.

Let's return to the experiment in which 10 operators made cartons on each of 10 machines. Previously the machines were assumed selected randomly from a large population of machines.

Let A refer to machines with 10 levels and B to operators with 10 levels.

Suppose now that the 10 machines are of different specific types (brands, say), so that A is a fixed factor. If the operators are still selected from a population, then B is a random factor.

Notation: At least for the present I will use an upper case I, J, or K instead of i, j and k as the subscript on a random factor.

Thus, for example, when you see $\alpha\beta_{iJ}$ you know it is an interaction effect between a fixed factor A and random factor B.

You have at least two choices for how you view the machine×operation interaction .

Restricted model

For each type i machine, operator J in the population has a random mean

$$\mu_{iJ} = \mu + \alpha_i + \beta_J + \alpha\beta_{iJ}$$

- β_J is an effect of the characteristics of the operator, including skill, as they affect the production on any machine
- $\alpha\beta_{iJ}$ is an effect of the characteristics of the operator's production specific to working on a on type i machine.

For operator J, the actual machine effect (difference of μ_{iJ} from $\mu + \beta_J$) is $\alpha_i + \alpha\beta_{iJ}$. This applies whenever operator J makes boxes on a machine of this type

In this context, since $\alpha_i + \alpha\beta_{iJ}$ is a machine effect it makes sense to apply the same restriction to $\alpha_i + \alpha\beta_{iJ}$. as to machine main effect α_i:

$$\sum_i(\alpha_i + \alpha\beta_{iJ}) = 0$$

But since $\sum_i\alpha_i = 0$, this means $\sum_i\alpha\beta_{iJ} = 0$.

This is characteristic of the **restricted model**:

> Sums of a mixed effect over a subscript associated with a fixed effect are 0.

You should use the restricted model when you think of an interaction such as $\alpha\beta_{iJ}$ as "attached" to level J of factor B in the sense that every time level J is randomly selected, you would get the same interaction effects $\alpha\beta_{1J}, \alpha\beta_{2J},\ldots,\alpha\beta_{aJ}$.

Unrestricted model

Another way you might view the interaction $\alpha\beta_{iJ}$ is that it is not "attached" to operator J, that is it does not come from specific characteristics of operator J.

Instead $\alpha\beta_{iJ}$ reflects the random circumstances of the particular session of making boxes on machine i. These might be such things as temperature and humidity, how tired the worker is, whether he/she has a hangover, etc.

The same operator J at another time would have a different $\alpha\beta_{iJ}$ on the same type machine.

In this case, a restriction such as $\sum_i\alpha\beta_{iJ} = 0$ doesn't make much sense.

This is the **unrestricted model**.

These ideas can clearly be extended to experiments with higher order interactions.

In the **restricted** model, you assume the sums over the subscripts corresponding to the fixed effects are 0.

In the **unrestricted** model there is no assumption that any sums over subscripts on a random effect are 0.

In the restricted model, the variance, say $\sigma_{\alpha\beta}^2 = V(\alpha\beta_{iJ})$ is not the proper way to summarize the contribution to an EMS. $\sigma_{\alpha\beta}^2$ overstates the contribution because it ignores the loss of variability coming from the restrictions.

Because $\sum_i \alpha\beta_{iJ} = 0$, the effects for the same J and different i are negatively correlated.

In tables of expected mean squares, you need to interpret each symbol σ_X^2 as meaning r_xVar(X) where $r_x < 1$ depends on the number of levels in the fixed factors in the effect X.

This reinterpretation of the symbol σ_X^2 doesn't change how you decide on what mean squares appear in F-tests.

The factor "shrinking" a variance component is $r_x = r_2/r_1$ where

- r_1 = product of all the levels of the fixed factors in a term (ac for ABC interaction with A and C fixed)
- r_2 = product of these levels – 1 ((a-1)(c-1) for ABC interaction with A and C fixed)

In a three factor experiment with A and C fixed and B random, for the ABC variance component

- $r_1 = ac$
- $r_2 = (a-1)(c-1)$

$r_{ac} = (a-1)(c-1)/(ac) = (1 - 1/a)(1 - 1/c)$

By default, `ems()`, `mixed()` and `varcomp()` assume the restricted model. If you believe the unrestricted model is more appropriate you should use `restrict:F` as an argument to these macros.

We return to a modified form of the experiment in which b = 2 analysts in each of a = 10 labs made two determ-inations on each of two samples. Originally I viewed the analysts as a nested random effect.

Now suppose that one analyst in each lab is experienced and the other is newly hired so you have a fixed crossed factor **experience** with two levels.

For this reason I use a renamed copy of factor `analyst`.

```
Cmd> exper <- analyst
```

Restricted model EMS and var components

```
Cmd> ems("y = lab + exper + lab.exper.sample",\
      vector("lab","sample")) # restricted
EMS(CONSTANT) = V(ERROR1) + 2V(lab.exper.sample) + 8V(lab)
    + 48Q(CONSTANT)
EMS(lab) = V(ERROR1) + 2V(lab.exper.sample) + 8V(lab)
EMS(exper) = V(ERROR1) + 2V(lab.exper.sample) + 4V(lab.exper)
    + 24Q(exper)
EMS(lab.exper) = V(ERROR1) + 2V(lab.exper.sample)
    + 4V(lab.exper)
EMS(lab.exper.sample) = V(ERROR1) + 2V(lab.exper.sample)
EMS(ERROR1) = V(ERROR1)

Cmd> varcomp("y = lab + exper + lab.exper + lab.exper.sample",\
      vector("lab","sample"))
                  Estimate         SE         DF
lab                0.00941   0.0070378     3.5755
lab.exper        0.0088221   0.0078058     2.5547
lab.exper.sample 0.0030646   0.0029115     2.2158
ERROR1           0.0071958   0.0020773         24
```

Unrestricted model EMS and components

```
Cmd> ems("y = lab + exper + lab.exper + lab.exper.sample",\
      vector("lab","sample"),restrict:F)
EMS(CONSTANT) = V(ERROR1) + 2V(lab.exper.sample) + 4V(lab.exper)
    + 8V(lab) + 48Q(CONSTANT)
EMS(lab) = V(ERROR1) + 2V(lab.exper.sample) + 4V(lab.exper
    + 8V(lab)
EMS(exper) = V(ERROR1) + 2V(lab.exper.sample) + 4V(lab.exper)
    + 24Q(exper)
EMS(lab.exper) = V(ERROR1) + 2V(lab.exper.sample)
    + 4V(lab.exper)
EMS(lab.exper.sample) = V(ERROR1) + 2V(lab.exper.sample)
EMS(ERROR1) = V(ERROR1)
```

The underlined terms are not present in the restricted model EMS.

```
Cmd> varcomp("y = lab + exper + lab.exper + lab.exper.sample",\
      vector("lab","sample"),restrict:F)
                  Estimate         SE         DF
lab               0.004999   0.0079899     0.7829
lab.exper        0.0088221   0.0078058     2.5547
lab.exper.sample 0.0030646   0.0029115     2.2158
ERROR1           0.0071958   0.0020773         24
```


So deciding on a model and doing an analysis require a number of things

- Determine the sources of variation (factors).
- Decide which are crossed and which are nested. A factor is crossed if a particular subscript value has the same meaning for all levels of other factors
- Decide which factors are fixed and which are random.
- Decide which interactions are in the model.
- Decide whether the model should be restricted or unrestricted.

Hasse Diagrams

You have seen how important it is to

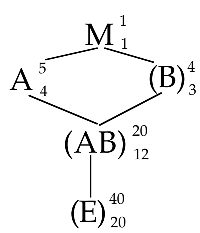
- find expected mean squares
- select denominators for tests

For balanced data there is an important tool -- the **Hasse Diagram**

This represents ANOVA model in semi-graphical form giving information on:

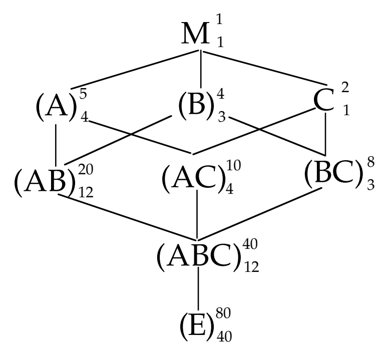
- Which factors are fixed and random
- The number of effects for each term
- The degrees of freedom for each term

Two way factorial with A fixed (a=5) and B random (b=4). There are 5 *nodes*, one for each term. (...) means random term. Subscript = DF Superscript = number of effects.



M_1^1 is *above* everything. A_4^5 and $(B)_3^4$ are above $(AB)_{12}^{20}$ which is above $(E)_{20}^{40}$

3-way factorial with A and B random (a=5, b=4), C fixed (c=2).



Fully nested with A and B random (a = 5, b = 4) and C fixed (c = 2). C is in (...) because it is nested in a random factor. This makes the actual levels of C present random, even though they are fixed once the levels of B are selected.