Chapters 25 & 27: Simple random effects and mixed effects models

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

25.1 One-way random effects model

If treatment levels come from a larger population, their effects are best modeled as random. A random-effects one-way model is

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

where

$$\alpha_1, \ldots, \alpha_r \overset{iid}{\sim} N(0, \sigma_\alpha^2) \text{ independent of } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2).$$

As usual, $i = 1, \ldots, r$ and $j = 1, \ldots, n_i$.

The test of interest is $H_0 : \alpha_1 = \cdots = \alpha_r = 0$. This happens if and only if $H_0 : \sigma_\alpha = 0$.

$\alpha_1, \ldots, \alpha_r$ are called *random effects* and σ_α and σ are termed *variance components*. This model is an example of a *random effects* model, because it has only random effects in it beyond the intercept μ (which is fixed).

Testing $H_0 : \sigma_\alpha = 0$

The MSE and MSTR are defined as they were before. One can show $E(MSE) = \sigma^2$ and $E(MSTR) = \sigma^2 + n\sigma_\alpha^2$ when $n = n_i$ for all i. SAS provides the expected mean squares on the last page of output.

If $\sigma_\alpha = 0$ we expect $F^* = MSTR/MSE$ to estimate one. In fact, just like the fixed-effects case, $F^* \sim F(r-1, n_T - r)$. This is the test given by `proc glm` when you add a `random A;` statement.

One can also fit the model in `proc mixed`, but this procedure provides a slightly cruder test of $H_0 : \sigma_\alpha = 0$.

Apex Enterprises

$r = 5$ personnel officers were selected at random, and $n_i = 4$ prospective employee candidates assigned at random to each officer. Y_{ij} is the rating of the ith officer on their jth candidate.

Since the personnel officers are chosen randomly from a large population of personnel officers, the random-effects one-way model applies.

```
data apex;
input rating officer @@;
  76  1  65  1  85  1  74  1  59  2  75  2  81  2  67  2
  49  3  63  3  61  3  46  3  74  4  71  4  85  4  89  4
  66  5  84  5  80  5  79  5
;

proc glm; class officer; * Chapter 16, fixed-effects approach;
 model rating=officer;

proc glm; class officer; * Chapter 25, mixed-effects approach;
 model rating=officer;
 random officer;

proc glimmix; class officer; * Tim prefers method=mle;
 model rating= / s;
 random officer;
 covtest zerog; * tests H0: sigma_alpha=0 vs. H0: sigma_alpha>0;
```

25.5 & 27.2 Random block effects and repeated measures

When block levels come from a large population, we can consider a complete randomized block design with random block effects. One very important example of this is the repeated measures design, where each block is an experimental unit in which all treatment levels are randomly applied. In fact, the blocks are retermed "subjects" and we consider a sample of subjects from their population.

$$Y_{ij} = \mu + \underbrace{\rho_i}_{\text{subject}} + \underbrace{\tau_j}_{\text{level}} + \epsilon_{ij},$$

where

$$\rho_1, \dots, \rho_n \overset{iid}{\sim} N(0, \sigma_\rho^2) \text{ independent of } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2).$$

There are $i = 1, \dots, n$ subjects receiving each of $j = 1, \dots, r$ treatments.

Random block effects

Note that the model can be extended to factorial treatment structure, e.g.

$$Y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk},$$

Examples of subjects include people, animals, families, cities, and clinics.

This is an example of a mixed effects model; there is a mix of random (ρ_i's) and fixed (α_j's, β_k's, and $(\alpha\beta)_{jk}$'s) effects in the model.

Random blocks, comments

- We assume subject effects and treatment effects do not interact. Can check via Tukey's 1 df test for additivity. Also look at interaction plots as in fixed-effects C.R.B. designs.
- ANOVA table, sums of squares exactly the same, except now the F-test for blocks tests $H_0 : \sigma_\rho = 0$ instead of $H_0 : \rho_1 = \cdots = \rho_n = 0$.
- Test for treatment is same $H_0 : \tau_1 = \cdots = \tau_r = 0$.
- Every treatment is given to every experimental unit in *randomized order*.
- Two sets of residuals to consider. Both should be normal; e_{ij} should have constant variance.
    1. $e_{ij} = Y_{ij} - \{\hat{\mu} + \hat{\rho}_i + \hat{\tau}_j\}$, and
    2. $\hat{\rho}_i$.
- $\text{corr}(Y_{ij_1}, Y_{ij_2}) = \sigma_\rho^2/(\sigma^2 + \sigma_\rho^2)$ for $j_1 \neq j_2$ tells you how correlated the repeated measures are.
- We will use `proc glimmix` to fit these models.

Road paint wear (p. 1082)

This is homework problems 25.19 and 25.20.

A state highway department studied wear of five paints at eight randomly picked locations. The standard is paint 1. Paints 1, 3, and 5 are white; paints 2 and 4 are yellow. At each location a random ordering of the paints were applied to the road. After an exposure period, a combined measure of wear Y_{ij} was recorded. The higher the score, the better the wearing characteristics.

Recall the model

$$Y_{ij} = \mu + \underbrace{\rho_i}_{\text{location}} + \underbrace{\tau_j}_{\text{paint}} + \epsilon_{ij},$$

where

$$\rho_1, \ldots, \rho_n \overset{iid}{\sim} N(0, \sigma_\rho^2) \text{ independent of } \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2).$$

Here $r = 5$ and $n = 8$.

Road paint wear in SAS

```
data road;
 input wear location paint @@;
datalines;
   11.0       1       1    13.0       1       2    10.0       1       3    18.0       1       4    15.0       1       5
   20.0       2       1    28.0       2       2    15.0       2       3    30.0       2       4    18.0       2       5
    8.0       3       1    10.0       3       2     8.0       3       3    16.0       3       4    12.0       3       5
   30.0       4       1    35.0       4       2    27.0       4       3    41.0       4       4    28.0       4       5
   14.0       5       1    16.0       5       2    13.0       5       3    22.0       5       4    16.0       5       5
   25.0       6       1    27.0       6       2    26.0       6       3    33.0       6       4    25.0       6       5
   43.0       7       1    46.0       7       2    41.0       7       3    55.0       7       4    42.0       7       5
   13.0       8       1    14.0       8       2    12.0       8       3    20.0       8       4    13.0       8       5
;

proc glm plots=all; class location paint; * interaction plot to check additivity;
 model wear=location|paint;

proc glimmix plots=all; class location paint; * 25.20(b,c,d)
 model wear=paint; * only include fixed effects;
 random location;  * only include random effects;
 lsmestimate paint "1 vs 2" 1.00 -1.00  0.00  0.00  0.00,
                   "1 vs 3" 1.00  0.00 -1.00  0.00  0.00,
                   "1 vs 4" 1.00  0.00  0.00 -1.00  0.00,
                   "1 vs 5" 1.00  0.00  0.00  0.00 -1.00 / adjust=bon cl alpha=0.9;
 estimate    "W vs Y" paint 0.34 -0.50  0.33 -0.50  0.33;
```

Wine tasting

$r = 4$ Chardonnary wines of the same vintage were judged by $n = 6$ judges. Each wine was blinded and given to each judge in randomized order. The wines were scored on a 40-point scale Y_{ij}, with higher scores meaning better wine.

The six judges are considered to come from a large population of wine-tasting judges and so a repeated measures model is appropriate.

The analysis of these data are carried out in your textbook on pp. 1132–1137.

Wine tasting in SAS proc glm

```
data wine;
input rating judge wine @@;
datalines;
  20  1  1  24  1  2  28  1  3  28  1  4  15  2  1  18  2  2  23  2  3  24  2  4
  18  3  1  19  3  2  24  3  3  23  3  4  26  4  1  26  4  2  30  4  3  30  4  4
  22  5  1  24  5  2  28  5  3  26  5  4  19  6  1  21  6  2  27  6  3  25  6  4
;

* spaghetti plot figure 27.2 on p. 1133;
proc sgplot noautolegend;
 series x=wine y=rating / group=judge;
 scatter x=wine y=rating / group=judge markerchar=judge;
 run;

* glm works, but is not really designed for repeated measures;
* this duplicates what is in your book;
proc glm plots=all; * gives figure 27.3 on p. 1133;
 class wine judge;
 model rating=wine judge;
 random judge; * need to include 'judge' in model using glm;
 lsmeans wine / pdiff adjust=tukey alpha=0.05 cl;
 run;
```

Analysis in proc glimmix

```
* proc mixed or proc glimmix is a better choice overall;
* note that Tukey intervals are essentially the same;
* conditional residuals are r_ij;
proc glimmix plots=all;
 class wine judge;
 model rating=wine / s chisq; * model includes only 'fixed' effects;
 random judge;                * random includes only 'random' effects;
 lsmeans wine / pdiff adjust=tukey alpha=0.05;
 covtest zerog; * tests H0: sigma_rho=0 vs. H0: sigma_rho>0;
 run;

* obtain estimates of rho_i;
ods listing close; ods output SolutionR=rand; * sends the rho_i to 'rand';
proc glimmix data=wine;
 class wine judge;
 model rating=wine;
 random judge / s; * ask for rho_i;
run;
ods output close;  ods listing;


* check that rho_i estimates are approximately normal;
proc print data=rand; run;
proc univariate data=rand normal; var estimate;
run;
```