Generalizations
Part 5: A Mixed Effects Model

STAT 705: Regression and Analysis of Variance

Mixed Effects

- In two-way ANOVA involving factors A and B, we assume that the levels of the factors are fixed by the researcher in advance of the study. We say that the factors are fixed effects.
- In this lesson we will consider a case in which the levels of one the factors are fixed and the levels of the other factor are selected randomly from a population of such levels.
- This leads us to what are know as mixed-effects.

Example

- Suppose a medical researcher wishes to compare the recovery times of two medical procedures and selects 3 hospitals at random from a population of hospitals as locations for doing the study. Each hospital does both procedures 3 times.
- It is possible that there is a medical procedure by hospital interaction because not all the hospitals have the same patients or doctors. Thus the hospital effect is not additive as it would be if it were a block.

Example, continued

- The hospitals are random effects because they have been selected at random from a larger population, but the medical procedures are fixed effects because they were predetermined for the study. (The procedures do not come from a random sample of all possible procedures.)
- Because both fixed effects and random effects are involved in the study, we say the that we have mixed-effects in our study.

The Mixed Effects Model

- Suppose we have two factors A and B, where A is the fixed effects factor and B is the random effects factor.
- The model for a response looks similar to the two-way ANOVA model with interaction:

  $$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}$$

- The difference is that the β_j's and γ_{ij}'s are assumed to be selected randomly from normally distributed populations, with mean 0 and variances σ_β^2 and σ_γ^2, respectively.
- We still assume that the ε_{ij}'s are normal, independent and identically distributed, with mean 0 and constant variance, but we now denoted the common variance as σ_E^2

KANSAS STATE UNIVERSITY

F-Tests for Mixed Effects

- Assume A is a fixed effect and B is a random effect and that there are an equal number of observations for all A and B combinations.
- Here are the ratios of mean squares for testing A and B main effects and the A*B interaction:

  F_A = MSA / MS(A*B)

  F_B = MSB / MS(A*B)

  F_{AB} = MS(A*B) / MSE
- In the mixed effects analysis the denominator of the F-statistics for A and B is the A*B interaction mean square, but in the fixed effects analysis this denominator is the MSE.

Degrees of Freedom

- We are still assuming A is a fixed effect and B is a random effect and that there are an equal number of observations for all A and B combinations.
- Suppose factor A has ‘a’ levels and factor B has ‘b’ levels
- Then dfA = a – 1, dfB = b – 1, and dfAB = (a – 1)(b – 1)
- For each of the F statistics, the numerator and denominator degrees of freedom are as follows

  F_A = MSA / MS(A*B); df numerator = dfA; df denominator = dfAB

  F_B = MSB / MS(A*B); df numerator = dfB; df denominator = dfAB

  F_{AB} = MS(A*B) / MSE; df numerator = dfAB; df denominator = dfE

Hospital Data

These are the hospital data that we use to illustrate the analysis. They are also in a separate file on the course website.

patient	Hospital	Procedure	Recovery
1	1	a	10
2	1	a	12
3	1	a	18
4	1	b	15
5	1	b	20
6	1	b	23
7	2	a	9
8	2	a	11
9	2	a	12
10	2	b	13
11	2	b	15
12	2	b	20

patient	Hospital	Procedure	Recovery
13	3	a	10
14	3	a	12
15	3	a	16
16	3	b	24
17	3	b	26
18	3	b	28
19	4	a	10
20	4	a	11
21	4	a	16
22	4	b	24
23	4	b	28
24	4	b	30

GLM for Hospital Data

- We used GLM two-way ANOVA to obtain the following table of sums of squares, degrees of freedom, and mean squares (ignoring F statistics and p-values)

Source	DF	SS	MS
hosp	3	163.125	54.375
pro	1	590.042	590.042
hosp*pro	3	110.792	36.931
Error	16	164.000	10.250

- Here are the mixed-effects F statistics:

For medical procedure: F = 590.042 / 36.931 = 15.98, df = (1, 3)

For hospital: F = 54.375 / 36.931 = 1.47, df = (3, 3)

For interaction: F = 36.931 / 10.250 = 3.60, df = (3, 16)

Hospital Data: Compare Procedures

- Suppose the researcher asks:

  “Would there be a difference between the average recovery times if these procedures were performed across the population of hospitals?”
- In other words, the researcher is interested in the main effect of procedure.
- The F statistic for procedure is 15.98 with 1 df for numerator and 3 df for denominator. The 5% critical value is 10.13, so there is a significant difference between the marginal means of the two procedures.
- From computer output we find p = 0.0281.

Hospital Data: Fixed vs. Mixed

- We now compare the mixed effects analysis to the fixed effects analysis.
- In the fixed effect analysis (as is done in GLM), the F-statistic has MSE in the denominator of the F-statistic.
- From the information in the ANOVA table with this example, we see that the F-statistic for the main effect of procedure is

$$F = MS(\text{procedure})/MSE = 590.042/10.250 = 57.56$$

- There is 1 df for the numerator and 16 df for the denominator. The critical value is 4.49, so ‘procedure’ is a significant factor.
- The p-value for this F-statistic turns out to be $p<.0001$ compared to $p=0.0281$ for the mixed effect analysis.

Questions We Want to Answer

- In mixed effects analysis
  - We are most often interested in the main effect of the fixed factor
  - We want to know how the fixed factor behaves when averaged across the levels of the random factor
- In fixed effects analysis
  - We are interested in main effects only if there is no interaction.
  - Thus, testing for interaction is often the most important thing in a fixed effects analysis

Hospital Data: Fixed vs. Mixed

- In the hospital example, the researcher might like to make a recommendation to the population of hospitals. How a procedure does "on average" across the population of hospitals would be important. This is what we would get out of a mixed effects analysis.
- In fixed effects analysis, we would be interested only in the 3 hospitals in the study. Of particular importance would be whether or not the effectiveness of the procedures depends on the hospital in which they are performed (i.e., is there an interaction?), and if so, which procedure to recommend for each hospital.

Blocks as Random Factors

- In studies involving blocks, many statisticians consider blocks as random effects.
- The random effects analysis assumes that the blocks are selected randomly from a hypothetical population of blocks. Thinking of it this way, the researcher can make inferences about how the treatments compare across a population of such blocks.
- Unfortunately, the population of blocks is often not well-defined so it is not always clear exactly what population of blocks the results might apply to.

Special Case of RCB

- If the design is an RCB, the fixed effects and mixed effects analyses give the same F-statistics and p-values for comparing treatment means.
- This is because the mean square error for the RCB is the treatment*block mean square whether the analysis is for fixed effects or mixed effects.
- If the blocks are incomplete or if there are multiple experimental units for each treatment within blocks, then the fixed effects and random effects analyses will not give the same tests for treatment effects.

SAS PROC MIXED

- The MIXED procedure in SAS performs mixed effects analysis.
- The code is similar to GLM except that there is both a MODEL and a RANDOM statement
  - The MODEL statement contains the response and the fixed effects
  - The RANDOM statement contains only the random effects
- The complete code for the hospital data on the course website. You should be familiar with this code.
- Note: The interaction between a fixed effect and a random effect is another random effect. So if A is fixed and B is random, then A would go into the MODEL statement and both B and A*B would go into the RANDOM statement.

What You Should Know

- Understand the relationship between mixed effects analysis and fixed effects analysis
- Be able to use Proc Mixed to perform a mixed effects analysis
- Interpret the results of your analysis