Generalizations

Part 3: A Random Effects Model

STAT 705: Regression and Analysis of Variance

More than Random Error

- We have been concerned about random error in all our examples.
- We compute MSE and use it in the denominator of the F statistics to determine whether or not factors are significant.
- However….
- In some studies randomness can occur in more ways than just through random error.
- These sources of randomness are called random effects.

Fixed and Random Effects

- We have assumed that the researcher pre-determines the levels of a factor to be used in an experiment.
- When this occurs, we say the levels of the factor are “fixed”, and their effects are called fixed effects.
- In some cases, *levels of a factor are selected at random from a larger population*, and it is the larger population that the researcher is interested in making inferences about. Their effects are called random effects.

A Single Random Factor

- We will assume that we have a single factor whose levels have been randomly selected from a larger population of such levels.
- Within each level of the random factor, we take measurements independently on experimental units.
- Two questions of interest
  - Is the variance of the random factor 0 (or is it greater than 0)?
  - Assuming the variance of the random factor is not 0, what is the variance and how does it compare to the variance of the random error?

Example

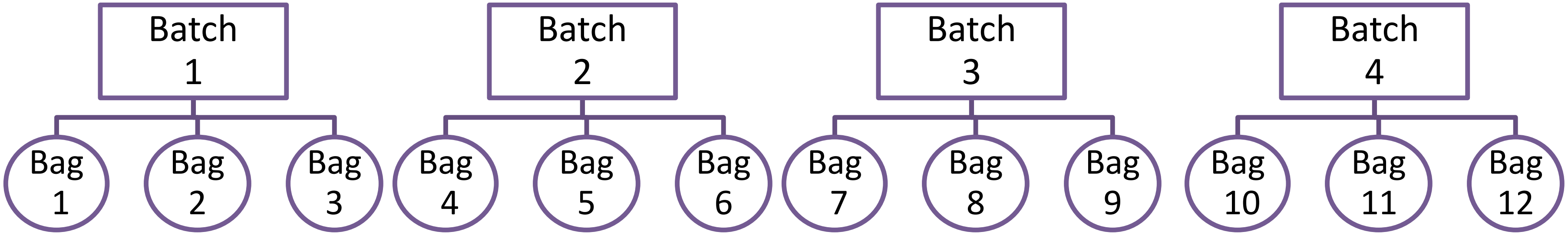
A manufacturer of corn chips receives raw material (corn) in batches. The batches are subdivided and processed to produce bags of corn chips. In terms of consistency of the product, the manufacturer would like to know which contributes more to the variability of the final product: the batches or the bags.

How would they use this information?

- If the batches are the primary contributor to variability, then they should look for a more reliable supplier of the raw material.
- If the bags are the primary contributor to variability, then they should look at their own production facility to identify reasons for the variability.

Example, continued

The manufacturer randomly selects 4 batches for study and within each batch randomly selects 3 bags as the experimental units upon which to measure the quality of the product.



Example, continued

The researcher would like to know if there is significant batch to batch variability and if so, how does it compare to the variance of the bags with in a batch.

If the 4 batches were the only ones the researcher cared about, then this would be like our one-way ANOVA setup. We would have the effects of the batches as fixed effects and the random variability that occurs in making measurements on the bags (our experimental units) would be the random error.

We will see what stays the same and what changes as the result of batches being random.

Random Effects Model ANOVA Table

- When we have an equal number of experimental units for each of the levels of the random factor, the computations are just the same as in one-way ANOVA, and the sources of variability and degrees of freedom are the same.
- The test statistic for testing whether or not the variance of the random factor is 0 is the same as the test for equality of means when the factor is fixed.
  - If the means are all the same, then the variability between the means is 0.
  - If the means are not all the same, then the variability between the means is positive.

Corn Chip Data and ANOVA Table

Below are data for the corn chip example and the ANOVA table from the Proc GLM output.

Batch	Bag	Quality
1	1	90
1	2	92
1	3	98
2	4	84
2	5	88
2	6	91
3	7	89
3	8	92
3	9	93
4	10	75
4	11	80
4	12	70

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	3	609.67	203.22	13.78	0.0016
Error	8	118.00	14.75		
Corrected Total	11	727.67			

The analysis shows that the variance of the population of batches is greater than 0 ($p = 0.0016$), that is, the means are not all the same.

Expected Mean Squares

- Our interest is in obtaining an estimate of the variance of the random factor. To get this, we need to introduce the idea of expected mean squares.
- The expected mean square of a MS in the ANOVA table is the average value we would get if we could repeat the experiment many times under the same condition.
- The expected mean squares for the random effects model are given on the next slide.

Random Effect Expected Mean Squares

- We need some notation:

  MSA = mean square of the random factor A

  MSE = mean square error

  σ^2_A = true variance of the random factor

  σ^2_E = true variance of error

  r = number of observations per level of the random factor

- Expected MSA = $\sigma^2_E + r \cdot \sigma^2_A$
- Expected MSE = σ^2_E
- **σ^2_A = (Expected MSA – Expected MSE)/r**

Corn Chips, Re-visited

- Use the values of MSA and MSE in the ANOVA table in place of the expected values in the equation for σ^2_A on the previous slide, and use r = 3. This gives us an estimate of the variance of the random factor.
- Estimated $\sigma^2_A = (203.22 - 14.75)/3 = 62.82$
- Compare this to MSE of 14.75. Most of the variability of the quality of corn chips can be attributed to Batch rather than to Bag.

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	3	609.67	203.22	13.78	0.0016
Error	8	118.00	14.75		
Corrected Total	11	727.67			

Negative Estimates of σ^2_A

- The true variance of the random factor can never be negative
- However, it is possible for the estimated value to be negative
- When this occurs, we should manually adjust the estimate so that it is a reasonable number
- There are several options, but I recommend that you set the estimate to 0

SAS Proc Mixed

- The Mixed procedure in SAS allows the inclusion of random factors.
- The syntax for Mixed is similar to GLM, with minor modifications
  - The MODEL statement includes only fixed factors
  - The RANDOM statement accommodates random factors
- Proc GLM does not have a RANDOM statement because it does not allow random factors

```
proc mixed data=cornchips;
class Batch;
model Quality = ;
random Batch;
run;
```

Note that the MODEL statement has nothing on the right hand side of the equation, because there are no fixed effects.

Proc Mixed Random Effects Output

- When there are no fixed effects in the MODEL statement
  - The MIXED output will produce estimates of the variance components σ^2_A and σ^2_E
  - These are the same estimates obtained from the ANOVA table

Covariance Parameter Estimates

Cov Parm	Estimate
Batch	62.8241
Residual	14.7500

- When there is more than one factor, it is possible to have combinations of fixed and random factors
- The MIXED procedure provides the flexibility to analyze many experimental designs

What You Should Know

- Recognize the difference between a fixed effect and a random effect
- Be able to use Proc Mixed to obtain estimates of the variance components for the random factors
- Interpret the results of your analysis