Spring 2014
Stat 8004: Statistical Methods II
Lecture 11

1 Random Effect Model

1.1 Examples

Example 1: Personnel management in a large enterprise. Does the interviewer have an effect on the rating of job candidates?

Data: 5 interviewers selected at random, each interviews 4 candidates selected at random. (Figure 1)

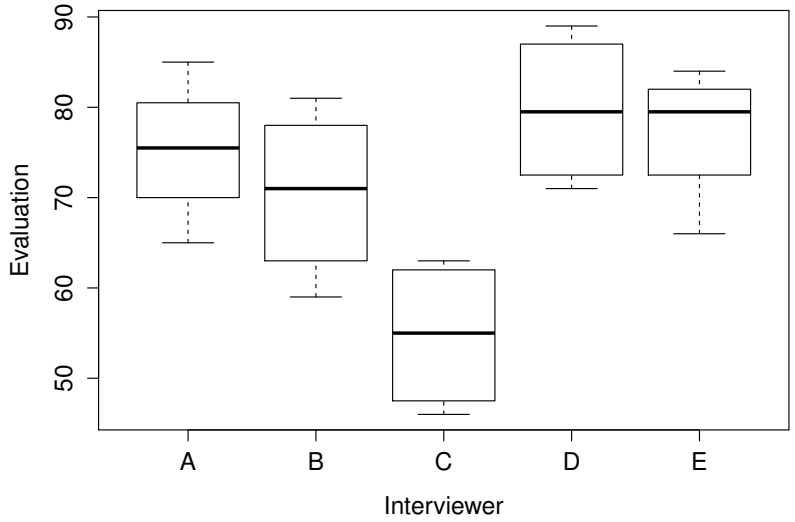

Figure 1: Boxplot of Rating vs. Interviewer

What's different about this data set?

Compare to previous example.

Example 2: Eye flicker frequency vs. eye colors. Does eye flicker frequenc differ across the brown, blue and green eye colors? (Figure 2)

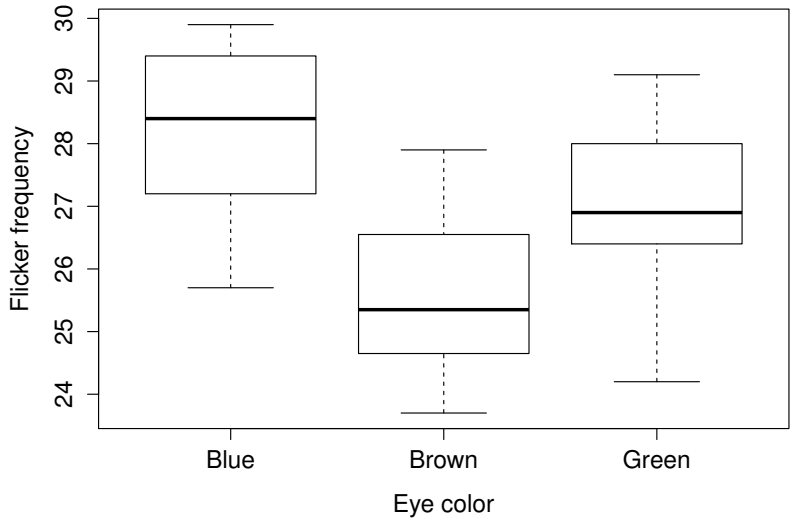

Figure 2: Boxplot of Rating vs. Interviewer

In Example 1, the interviews are *random draws* from a larger population. We are interested in the larger population and not just these 5 specific interviews. Meanwhile, in Example 2, we are just interested in these three eye colors. We are not trying to generalize our conclusion.

More precisely, *fixed effects* can be thought of as "treatment" levels that we have selected for inclusion in a study, which are the only levels of the variable in question in which we have an interest. In an experiment, we might have a treatment group and a control group. The purpose of the study is to compare these two groups – we are not trying to generalize to other treatments that we might have included, but didn't. In a non-experimental setting, a variable may have a small number of levels, and we have included them all – gender, for example, might be treated as a fixed effect because we have included all possible levels of this "treatment" variable in our study (i.e. males and females). Or, we may have a variable that has, many possible levels, but we are only interested in generalizing study results to the ones that we happened to include. For example, suppose that we did a survey using cluster sampling – and selected 10 cities at random. Normally, "city" could be thought of as a factor in the design, with 10 levels. If we did not want to generalize our results to all of the cities that might have been selected (for example, the 10 cities that we did include could be regarded as a sample of 10 drawn from a larger population of cities) – but only wanted to generalize to the 10 we included, then "city" could be treated as a "fixed" effect.

In the case of selecting 10 cities from a larger population of cities, just discussed, we would more naturally treat the variable "city" as a *random effect*. That is, we would regard the effects of "city" as a random sample of the effects of all the cities in the full population of cities. Conceptually, a variable's effects might be treated as random effects if we can think of

the levels of the variable that we included in the study as a sample drawn from some larger (conceptual) population of levels that could (in principle) have been selected.

One key difference between fixed and random effects is in the kind of information we want from the analysis of the effects. In the case of fixed effects, we are usually interested in making explicit comparisons of one level against another. For example, we very much would want to compare the mean of the "control group" to the mean of the "treatment group" in an experiment. If explicit comparison of the levels of a variable against one another is the goal of the research, then the levels of the variable are usually treated as "fixed". If, on the other hand, our primary interest is in the effects of other variables or treatments across the levels of a factor (e.g. the effect of gender on voting, across samples from 15 nations), then the "blocking" or "control" variable might be treated as a "random" effect. In this example, the dependent variable is voting, the independent variable is gender (which would be treated as a fixed factor so we can compare men and women), and the national context or sampling design variable (which of 15 nations) might be treated as a random factor.

In the case of a fixed factor, then, we are usually interested in comparing the scores on the dependent variable among the levels of the factor; our interest will be in differences between means. In the case of a random factor, we are not really interested in the specific differences in means from one level of the factor to another – but, we are interested in the extent to which the random factor accounts for variance in the dependent variable, because we want to control for this. So, rather than being interested in the individual means across the levels of the fixed factor, we are interested in the variance of means across the levels of a random factor.

Example 3. Does the sodium in beer differ between brands? 6 randomly chosen brands, 8 bottles tested per brand. (Figure 3)

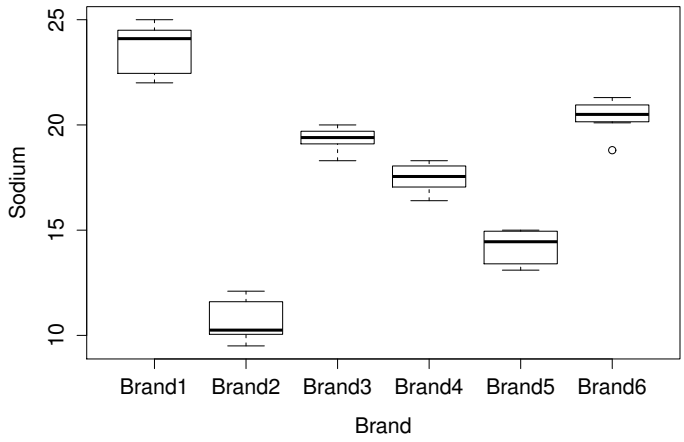

Figure 3: Sodium Content vs. Beer Brand

Shall we treat beer brand as random effect or fixed effect?

1.2 Random Effects Model

Assuming that cell-sizes are the same, *i.e.*, equal observations for each "subject" (brand of beer).

$$Y_{ij} = \mu + a_i + \epsilon_{ij}, \tag{1}$$

where $\epsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $a_i \overset{i.i.d.}{\sim} N(0, \sigma_\alpha^2)$, $1 \leq i \leq r$, $1 \leq j \leq n$.

Parameters:

- μ is the population mean;
- σ^2 is the measurement variance;
- σ_a^2 is the population variance of effect (*i.e.*, variation in sodium content of beer).

1.3 Inference and ANOVA Table

Decomposition of Variance of Covariance

$$\mathrm{Var}(Y_{ij}) = \sigma_a^2 + \sigma^2.$$

But only one parameter in mean function:

$$\mathbb{E}\,(Y_{ij}) = \mu.$$

The observations are no longer independent:

$$\mathrm{Cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma_a^2 + \sigma^2, & i = i',\ j = j'; \\ \sigma_a^2, & i = i',\ j \neq j'; \\ 0, & i \neq i',\ j \neq j'. \end{cases}$$

Random effects models are also called "variance components" models.

When cell sizes are the same (balanced),

$$\hat{\mu} = \bar{Y}_{\cdot\cdot} = \frac{1}{nr} \sum_{i,j} Y_{ij}.$$

This also changes estimates of σ^2 – see ANOVA table 2. You might guess that $df = nr - 1$ and $\hat{\sigma}^2 = \frac{1}{nr-1} \sum_{ij} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$. This is *not* the case.

Now let's compare the one-way ANOVA table. Suppose there are r levels for one factor, and there are n observations in each level.

Source	SS	df	$\mathbb{E}(MS)$
Treatment	$SSTR = \sum_{i=1}^{r} n(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r-1$	$\sigma^2 + n\frac{\sum_{i=1}^{r} a_i^2}{r-1}$
Error	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

Table 1: Fixed effect model

Source	SS	df	$\mathbb{E}(MS)$
Treatment	$SSTR = \sum_{i=1}^{r} n(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r-1$	$\sigma^2 + n\sigma_a^2$
Error	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

Table 2: Random effect model

- Only change here is the expectation of MSTR which reflects randomness of a_i's.
- ANOVA table is still useful to setup tests: the same F statistics for fixed effect models will work here.
- Test for random effect: $\mathrm{H}_0: \ \sigma_a^2 = 0$ based on

$$F = \frac{MSTR}{MSE} \sim F_{r-1,(n-1)r} \text{ under } \mathrm{H}_0.$$

How to estimate σ_a^2?

- From the ANOVA table:

$$\sigma_a^2 = \frac{\mathbb{E}(SSTR/(r-1)) - \mathbb{E}(SSE/((n-1)r))}{n}$$

- Natural estimate:

$$\hat{\sigma}_a^2 = \frac{SSTR/(r-1) - SSE/((n-1)r)}{n}$$

Sometimes, the estimate can be negative. If it is, set it to 0.

Influence about μ:

It is easy to check that

$$\mathbb{E}(\bar{Y}_{\cdot\cdot}) = \mu;$$
$$\mathrm{Var}(\bar{Y}_{\cdot\cdot}) = \frac{n\sigma_a^2 + \sigma^2}{rn}.$$

To come up with a t-statistic that we can use for test, we need to find an estimate of $\mathrm{Var}(\bar{Y}_{\cdot\cdot})$. ANOVA table says

$$\mathbb{E}(MSTR) = n\sigma_a^2 + \sigma^2.$$

Therefore,

$$\frac{\bar{Y}_{\cdot\cdot} - \mu_{\cdot}}{\sqrt{\frac{SSTR}{(r-1)rn}}} \sim t_{r-1}.$$

In fact, we use $SSTR/(r-1)$ to estimate $n\sigma_a^2 + \sigma^2$. Why $r-1$ degrees of freedom?

- Imagine we could record an infinite number of observations for each group, so that $\bar{Y}_{i\cdot} \to \mu_i$, or that $\sigma_a = 0$.
- To learn anything about μ, we still only have r oberavations $(\mu_1, \ldots, \mu_r)$.
- Sample more with an individual cannot narrow the CI for μ.

1.4 Data Analysis

Now let's get back to Example 1. To fit an random effect model, we can use the *nlme* package in R. The function to fit a random effect/mixed effect model is called *lme*.

```
> library("nlme")
> fm1 <- lme(Evaluation ~ 1, data = apex,
          random = ~ 1|Officer)
> summary(fm1)
Linear mixed-effects model fit by REML
 Data: apex
       AIC      BIC    logLik
  151.2466 154.0799 -72.62328

Random effects:
 Formula: ~1 | Officer
        (Intercept) Residual
StdDev:     8.99375 8.555699

Fixed effects: Evaluation ~ 1
            Value Std.Error DF  t-value p-value
(Intercept)  71.5  4.453932 15 16.05323       0

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-1.3839147 -0.8919203  0.2485326  0.6501819  1.2590574

Number of Observations: 20
Number of Groups: 5
> anova(fm1)
            numDF denDF  F-value p-value
(Intercept)     1    15 257.7063  <.0001
```

In practice, the model containing both fixed effects and the random effects have a much wider application. Due to time limits, we cannot further discuss the topic here. If you are interested, you can read some books on the topic *mixed effect model*, or take a course in *longitudinal analysis*.

2 Mixed Effect Model

2.1 Example

Example 4: Suppose we have a clinical trial in which a drug is administtered at four different dose levels. And the test was conducted in 20 different clinics in New York City which are randomly chosen. In each clinic each patient was randomly assigned to one of the dose levels. If y_{ijk} is the datum for patient k on dose level j in clinic i, then a suitable model for the data is:

$$\mathbb{E}\left(Y_{ijk}\right) = \mu + a_i + \beta_j + c_{ij},$$

where a_i is a random effect representing clinic i, β_j is a fixed effect for dose j of a drug, and c_{ij} is a random effect for interaction. This, with its mixure of fixed and random effects, is a *linear mixed effect model* (LMM). A special case of an LMM is when there are no fixed effects other than the intercept μ, which is a *random effect model*.

In linear effect models, fixed effects are used for modeling the mean of $\mathbf{y}$ while random effects govern the variance-covariance structure of $\mathbf{y}$. In fact, a prime reason for having random effects is to simplify the otherwise difficult task of specifying the $N(N+1)/2$ distinct elements of $\mathrm{Var}(\mathbf{y}_{N\times 1})$. Without using random effects, we would have to deal with elements of $\mathrm{Var}(\mathbf{y})$ being a variety of forms; but with random factors we can conveniently deal with variances and covariances attributable to factors acknowledged to be affecting the data.

2.2 Model

In a more general way, we can write our model as

$$\mathbb{E}\left(\mathbf{y} \mid \boldsymbol{\mu}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}. \tag{2}$$

Here, $\mathbf{X}\boldsymbol{\beta}$ are just as before, representing the fixed effect part. $\mathbf{Zu}$ represents the random effect part, where $\mathbf{Z}$, like $\mathbf{X}$, is a known matrix and $\mathbf{u}$ is the vector of random effects that occur in the data vector $\mathbf{y}$. We write the model using $\mathbb{E}\left(\mathbf{y} \mid \mathbf{u}\right)$ rather than $\mathbb{E}\left(\mathbf{y}\right)$ since $\mathbf{u}$ is just random realization of the random variable $\mathbf{U}$.

Question: How to write the example model into the matrix form as (2)?

In order to handle the first and second moments of $\mathbf{y}$, those of $\mathbf{u}$ are needed. They get specified by

$$\mathbf{u} \sim (\mathbf{0}, \mathbf{D}), \quad \text{meaning that } \mathbb{E}\left(\mathbf{u}\right) = \mathbf{0} \text{ and } \mathrm{Var}(\mathbf{u}) = \mathbf{D}.$$

There is no loss of generality in taking $\mathbb{E}(\mathbf{u}) = \mathbf{0}$. Why?

And also, we specify $\text{Var}(\mathbf{y} \mid \mathbf{u}) = \mathbf{R}$.

It is easy to show that

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}} + \mathbf{R}),$$

indicating the fixed effects enter only the mean whereas the random effects model matrix and variance enter only the variance of $\mathbf{y}$.

Now let's get back to Example 4. What would be the mean and variance for $\mathbf{y}$?

2.3 Another Example: Longitudinal Data

Example 5: Longitudinal data are successive observations on each of a collection of observational units (often people). An example is blood pressure measurements taken weekly on a group of patients. If $\mathbf{y}_i$ is the vector of n_i measurements on patient i, then we can use the model

$$\mathbb{E}(\mathbf{y}_i \mid \mathbf{u}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i,$$

with vectors $\boldsymbol{\beta}$ and $\mathbf{u}_i$ consisting of fixed and random effects, respectively. $\boldsymbol{\beta}$ is the same for all the patients and $\mathbf{u}_i$ is specific to patients i.

Suppose there are m such patients. Then for $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)^{t}rans$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)^{\mathrm{T}}$, and $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$ and $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_m)^{\mathrm{T}}$, we have

$$\mathbb{E}(\mathbf{y} \mid \mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}.$$

2.4 Estimating Fixed Effects

2.4.1 When $\mathbf{V} = \text{Var}(\mathbf{y})$ is known

Suppose $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, where $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}} + \mathbf{R}$. It is just like a general linear regression. The log-likelihood is

$$l = -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) - \frac{1}{2}\log|\mathbf{V}| - \frac{N}{2}\log(2\pi), \tag{3}$$

where $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$. Taking the derivative with respect to $\boldsymbol{\beta}$ and set to zero, we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y}. \tag{4}$$

It is easy to show that

$$\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

To test the null hypothesis $\mathrm{H}_0: \ \mathbf{K}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{m}$, we can derive a chi-square statistic using

$$X^2 = (\mathbf{K}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \mathbf{m})^{\mathrm{T}}\{\mathbf{K}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K}\}^{-1}(\mathbf{K}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \mathbf{m}).$$

Under H_0, X^2 has a central χ^2 distribution with Rank($\mathbf{K}$) degrees of freedom.

2.4.2 When $\mathbf{V} = \mathrm{Var}(\mathbf{y})$ is unknown

With $\mathbf{V}$ unknown but not being a function of $\boldsymbol{\beta}$, the log-likelihood function l in (3) has to be minimized with respect to both $\boldsymbol{\beta}$ and $\mathbf{V}$. Setting the derivatives to zero will lead to the same result as $\hat{\boldsymbol{\beta}}$ in (4), but just replacing $\mathbf{V}$ by $\hat{\mathbf{V}}$.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{V}}^{-1}\mathbf{y}.$$

Now we need to derive MLE for $\mathbf{V}$. Suppose $\mathbf{V}$ are parameterized by $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)^{\mathrm{T}}$.

$$\frac{\partial l}{\partial \phi_k} = -\frac{1}{2}\left\{\mathrm{tr}(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi_k}) - (\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi_k}\mathbf{V}^{-1}(y - \boldsymbol{\mu})\right\}.$$

On writing $\partial \mathbf{V}/\partial \phi_k \mid_{\mathbf{V}=\hat{\mathbf{V}}}$ as $\hat{\mathbf{V}}_{\phi_k}$, this gives

$$\mathrm{tr}(\hat{\mathbf{V}}^{-1}\hat{\mathbf{V}}_{\phi_k}) = \mathbf{y}^{\mathrm{T}}\hat{\mathbf{P}}\hat{\mathbf{V}}_{\phi_k}\hat{\mathbf{P}}\mathbf{y}$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}$$

and $\hat{\mathbf{P}}$ is $\mathbf{P}$ with $\mathbf{V}$ replaced by $\hat{\mathbf{V}}$. The detailed estimation will depend on the structure and the parametrization of $\mathbf{V}$.