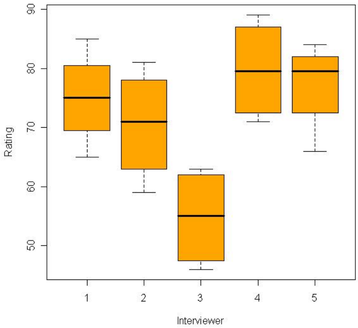
Lecture 7: Random Effect Models

Nancy R. Zhang

Statistics 203, Stanford University

February 2, 2010

Example



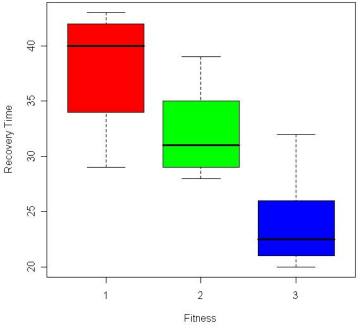
Setting: Personnel management in a large enterprise.

Question: 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.

What is different about this data set?

Compare to previous cases

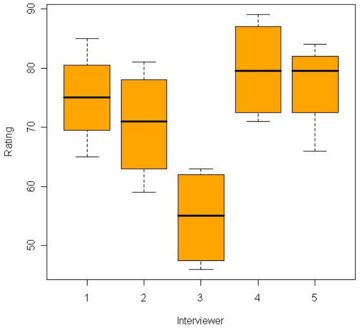


How does prior fitness affect recovery from surgery? Observations: 24 subjects' recovery time.

Three fitness levels: below average (8), average (10), above average (6).

Here, fitness level is of intrinsic interest. They are not random.

Example



Setting: Personnel management in a large enterprise.

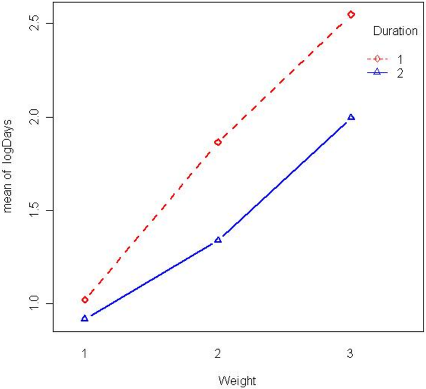
Question: Does the interviewer have an effect on the rating of job candidates?

Data: 5 interviewers selected at random, each interviews 4 candidates selected at random.

The interviewers are *random draws* from a larger population.

We are interested in the larger population and not these 5 specific interviewers.

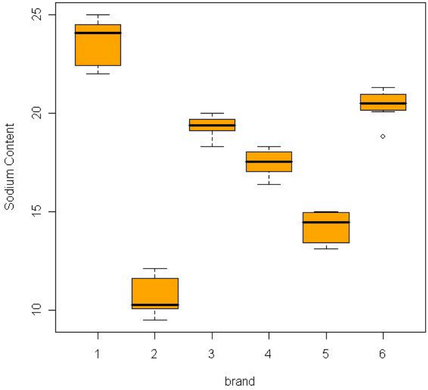
Another Example



Recovery time depends on weight gain between treatments and duration of treatment.

Two levels of duration, three levels of weight gain.

Another Example



How does the sodium in beer differ between brands?

6 randomly chosen brands, 8 bottles tested per brand.

Random Effects Model

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

$$Y_{ij} \sim \mu. + \alpha_i + \varepsilon_{ij}, \quad 1 \leq i \leq r, 1 \leq j \leq n$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$\alpha_i \sim N(0, \sigma_\alpha^2)$$

Parameters:

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

Decomposition of Variance and Covariance

$$Var(Y_{ij}) = \sigma_\alpha^2 + \sigma^2$$

But only one parameter in mean function:

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

- The observations are no longer independent:

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

- Random effects models are also called “variance components” models.

One way ANOVA: r groups, n observations in each group.

- Fixed effect model:

Source	SS	df	$E(MS)$
Treatments	$SSTR = \sum_{i=1}^{r} n\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$	$r-1$	$\sigma^2 + n\frac{\sum_{i=1}^{r}\alpha_i^2}{r-1}$
Error	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

- Random effect model:

Source	SS	df	$E(MS)$
Treatments	$SSTR = \sum_{i=1}^{r} n\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$	$r-1$	$\sigma^2 + n\sigma_\alpha^2$
Error	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

Inference for population mean: $\mu.$

- Easy to check that

$$E(\overline{Y}_{..}) = \mu.$$
$$\text{Var}(\overline{Y}_{..}) = \frac{n\sigma_\alpha^2 + \sigma^2}{rn}.$$

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

$$E(MSTR) = n\sigma_\alpha^2 + \sigma^2.$$

- Therefore,

$$\frac{\overline{Y}_{..} - \mu.}{\sqrt{\frac{SSTR}{(r-1)rn}}} \sim t_{r-1}$$

Inference for population mean: $\mu.$

$$\frac{\overline{Y}_{..} - \mu.}{\sqrt{\frac{SSTR}{(r-1)rn}}} \sim t_{r-1}$$

- Why $r - 1$ degrees of freedom? Imagine we could record an infinite number of observations for each group, so that $\overline{Y}_{i.} \to \mu_i$, or that $\sigma^2 = 0$.
- To learn anything about $\mu.$ we still only have r observations $(\mu_1, \ldots, \mu_r)$.
- Sampling more within an individual cannot reduce the degree of freedom for $\mu.$.

One-way ANOVA (random)

Source	SS	df	$E(MS)$
Treatments	$SSTR = \sum_{i=1}^{r} n\left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$	$r-1$	$\sigma^2 + n\sigma_\alpha^2$
Error	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n}(Y_{ij} - \overline{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

- Only change here is the expectation of *MSTR* which reflects randomness of α_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: $H_0 : \sigma_\alpha^2 = 0$ based on

$$F = \frac{MSTR}{MSE} \sim F_{r-1,(n-1)r} \qquad \text{under } H_0.$$

Estimating $\sigma_\alpha^2/(\sigma_\alpha^2+\sigma^2)$

As before, the MSTR and MSE are independent.

$$\frac{MSTR/(n\sigma_\alpha^2+\sigma^2)}{MSE/\sigma^2} \sim F_{r-1,r(n-1)},$$

Thus,

$$P\left\{F_{r-1,r(n-1)}(\alpha/2) \leq \frac{MSTR/(n\sigma_\alpha^2+\sigma^2)}{MSE/\sigma^2} \leq F_{r-1,r(n-1)}(1-\alpha/2)\right\} = 1-\alpha.$$

Rearranging, we have:

$$L = \frac{1}{n}\left[\frac{MSTR}{MSE}\frac{1}{F(1-\alpha/2)} - 1\right]$$

$$U = \frac{1}{n}\left[\frac{MSTR}{MSE}\frac{1}{F(\alpha/2)} - 1\right]$$

are lower and upper confidence bounds for σ_α^2/σ^2. For $\sigma_\alpha^2/(\sigma_\alpha^2+\sigma^2)$, they are:

$$L^* = \frac{L}{1+L} \quad L^* = \frac{U}{1+U}.$$

Estimating σ_α^2

- From the ANOVA table

$$\sigma_\alpha^2 = \frac{E(SSTR/(r-1)) - E(SSE/((n-1)r))}{n}.$$

- Natural estimate:

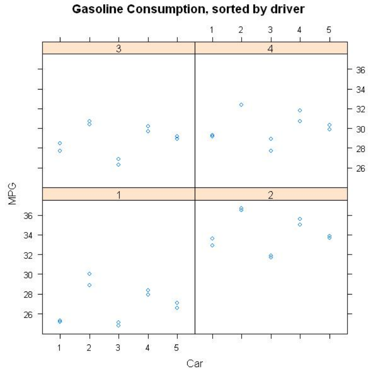
$$S_\alpha^2 = \frac{SSTR/(r-1) - SSE/((n-1)r)}{n}$$

- Problem: this estimate can be negative. If it is, set to 0.

Two-way ANOVA (random)

An automobile manufacturer wishes to study the differences between drivers and between cars on gasoline consumption.

Example: Gasoline Consumption



Are there effects of:

- different cars
- different drivers

on gasoline consumption?

Data:

- 5 cars
- 4 drivers
- Each driver drove each car twice.

Two-way ANOVA (random)

Observations, for $1 \leq i \leq a, 1 \leq j \leq b, 1 \leq k \leq n$:

$$Y_{ijk} \sim \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2),$$

$$\alpha_i \sim N(0, \sigma^2_\alpha),$$

$$\beta_j \sim N(0, \sigma^2_\beta),$$

$$(\alpha\beta)_{ij} \sim N(0, \sigma^2_{\alpha\beta}).$$

Sums of squares

Identical to fixed effects models

$$SSA = nb \sum_{i=1}^{a} \left(\overline{Y}_{i..} - \overline{Y}_{...} \right)^2$$

$$SSB = na \sum_{j=1}^{b} \left(\overline{Y}_{.j.} - \overline{Y}_{...} \right)^2$$

$$SSAB = n \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...} \right)^2$$

ANOVA tables: Two-way (random)

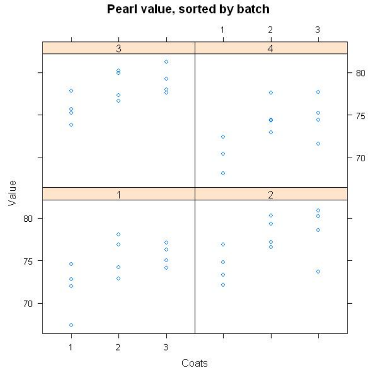
SS	df	$E(MS)$
SSA	$a-1$	$\sigma^2 + nb\sigma_\alpha^2 + n\sigma_{\alpha\beta}^2$
SSB	$b-1$	$\sigma^2 + na\sigma_\beta^2 + n\sigma_{\alpha\beta}^2$
$SSAB$	$(b-1)(a-1)$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
SSE	$(n-1)ab$	σ^2

- To test $H_0 : \sigma_\alpha^2 = 0$ use SSA and $SSAB$.
- To test $H_0 : \sigma_{\alpha\beta}^2 = 0$ use $SSAB$ and SSE.

Two-way ANOVA (mixed)

Market research: How does number of coats of a special lacquer applied to a plastic bead affect its market value?

Example: Imitation Pearls



How does the number of lacquer coating on the pearl affect its market value?

Data:

- 4 batches of 12 beads each.
- 3 levels of coating (fixed in advance) applied to each batch.

Two-way ANOVA (mixed)

$$Y_{ijk} \sim \mu_{\cdot\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, 1 \leq i \leq a, 1 \leq j \leq b, 1 \leq k \leq n$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2),$$

$$\alpha_i \sim N(0, \sigma^2_\alpha),$$

$$\beta_j \quad \text{are constants.}$$

$$(\alpha\beta)_{ij} \sim N(0, (b-1)\sigma^2_{\alpha\beta}/b).$$

Constraints:

- $\sum_{j=1}^{b} \beta_j = 0$
- $\sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, 1 \leq i \leq a.$

Details can safely be ignored here.

Two-way ANOVA

MS	df	Fixed	Random	Mixed
MSA	$a-1$	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1}$	$\sigma^2 + nb\sigma_\alpha^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + na\sigma_\alpha^2$
MSB	$b-1$	$\sigma^2 + na\frac{\sum \beta_j^2}{b-1}$	$\sigma^2 + na\sigma_\beta^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + nb\frac{\sum_{j=1}^b \beta_i^2}{b-1} + n\sigma_{\alpha\beta}^2$
$MSAB$	$(b-1)(a-1)$	$\sigma^2 + n\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\sigma^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
MSE	$(n-1)ab$	σ^2	σ^2	σ^2

Mixed Effect Models:

- To test $H_0 : \sigma_\alpha^2 = 0$ use MSA and MSE.
- To test $H_0 : \beta_1 = \cdots = \beta_b = 0$ use MSB and $MSAB$.
- To test $H_0 : \sigma_{\alpha\beta}^2$ use $MSAB$ and MSE.

The `R` library for fitting random and mixed effect models is called `nlme`. You can use the function `lme` from this library. We are only going to skim the surface here.

For `lme`, you need to give what `R` calls a model formulae:

- `miles.lme = lme(MPG~ 1, data=miles,random=~ 1|Driver/Car)`
- `pearls.lme = lme(Value~ Coats, data=pearls,random=~ 1|Batch)`