STATS660.1 Overview of the linear model (work in progress)

Anonymous

November 2, 2010

Copyright This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/1.0/ or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Contents

1 The linear model

1.1 Simple linear regression

Consider the model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

used to describe the relationship between pairs of numbers (x_i, y_i), $i = 1, \dots, n$.

We commonly assume that the errors are Gaussian $\varepsilon_i \sim N(0, \sigma^2)$ and independent.

We can write this in matrix form

$$y = X\beta + \varepsilon$$

with

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

1.2 Ordinary least squares (OLS)

Ordinary least squares (OLS). We can estimate β_1, β_2 by minimizing

$$\sum_{i=1}^{n} (y_i - (\beta_1 + \beta_2 x_i))^2$$

by β_1, β_2. The resaults, obtained by setting derivatives to zero are

$$\hat{\beta}_2 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

1.3 Normal distribution

The normal (Gaussian) p.d.f is given by

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\big(-\frac{(t-\mu)^2}{2\sigma^2}\big)$$

and a random variable T with this distribution is denoted by $T \sim n(\mu, \sigma^2)$. If $Y_1, \dots, Y_n \sim n(\mu, \sigma^2)$ are independent then the joint density is the product of the individual p.d.f's

$$\prod_{i=1} f(y_i) = \frac{1}{(2\pi)^(n/2)\sigma^n} \exp\big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\big)$$

For fixed data, $y_1, \dots, y_n$, this can be viewed as a function of the parameters denoted $L(\mu, \sigma^2)$ where L is termed the likelihood function.

1.4 Maximum likelihood estimation (MLEs) for SLR

The likelihood for the SLR model β based on realizing that here

$$y_i \sim n(\beta_1 + \beta_2 x_i, \sigma^2)$$

and these are independent so (ignoring [2])

$$L(\beta) = L(\beta_1, \beta_2) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\big(-\frac{1}{2\sigma^2}\sum(y_i - (\beta_1 + \beta_2 x_i))^2\big)$$

The MLE of $(\beta_1, \beta_2) = \beta$ is the value wich maximizes $L(\beta)$ over β, or equivalently minimizes $-\ln(L(\beta))$ over β. This is numerically equivalent to OLS and the solution is the same.

1.5 Multiple linear regression ((general) linear model)

When we have several explanatory variables x_j and a variable y to be predicted the linear model becomes

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \varepsilon_i \quad i = 1, \ldots, n$$

or

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon.$$

The OLS solution

$$\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|^2 = \min_{\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} (\beta_j x_{i,j}))^2$$

is

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

1.6 Multivariate normal distribution

Suppose $Z_1, \ldots, Z_n$ are independent Gaussian with mean zero and variance one ($Z_1, \ldots, Z_n \sim n(0,1)$ iid) so their joint density is

$$f(\xi) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\xi_i^2/2\right) = \frac{1}{(2\pi)^{n/2}} \exp(-(1/2)\xi^T\xi)$$

Let A be an invertible $n \times n$ matrix and $\mu \in^n$ and define $\mathbf{Y} = A\mathbf{Z} + \mu$. Recall from calculus that if g is a $1-1$ function $g :^n \to^n$

$$\int f(\xi) d\xi = \int f(g(\mathbf{y}))|J| d\mathbf{y}$$

where J is the Jacobian of the transformation

$$J = \left|\frac{d\xi}{d\mathbf{y}}\right| = \left|\frac{\partial g(\mathbf{y})}{\partial \mathbf{y}}\right|$$

and the integrals are over corresponding regions. It follows that the joint pdf of $\mathbf{Y}$ is h with $h(y) = f(g(y))|J|$.
Some linear algebra gives

$$h(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu))$$

where $\Sigma = AA^T$. This describes the mutivariate normal distribution $Y \sim n(\mu, \Sigma)$.

1.7 Distribution of the MLE

Distribution of the MLE

$$\hat{\beta} \sim n(\beta, (X'X)^{-1}\sigma^2)$$

1	2	…	I
$y_{1,1}$	$y_{2,1}$	…	$y_{I,1}$
$y_{1,2}$	$y_{2,2}$	…	$y_{I,2}$
…	…	⋱	…
y_{1,n_1}	y_{2,n_2}	…	y_{I,n_I}

1.8 Testing hypothesis

$$\frac{||y - X\beta||^2}{\sigma^2} \sim \chi^2_{n-p}$$

and is independent of $\hat{\beta}$ (if X is og full rank) so we can use t-tests to test $H_0 : \beta_i = \beta_{0_i}$ etc.

2 Analysis of variance in the one-way layout

2.1 The one-way layout

$$\boxed{y_{i,j} = \mu + \alpha_i + \varepsilon_{i,j}}$$

Consider measurements of response of a grouping variable (factor)

$$y_{i,j} = \mu + \alpha_i + \varepsilon_{i,j} \quad j = 1, \ldots, n \quad i = 1, \ldots, I$$

where $y_{i,j}$ is the j'th measurement within level i of the grouping factor and $\varepsilon_{i,j} \sim n(0, \sigma^2)$.

2.2 Partitioning the SS

Partitioning the SS without any model we can look at the total sum of squares

$$\text{SSTOT} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \overline{y})^2$$

where $\overline{y} = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{n_i} y_{ij}$ and $n = \sum_{i=1}^{I} n_i$.
Consider first the obvious estimates $y_{ii} = \frac{1}{n_i} \sum_{i=1}^{n_i} y_{ij}$ of the expected values of each y_{ij}, which have expected values $\mu + \alpha_i$.
Similarly, $\overline{y}_{\cdot\cdot}$ is an obious estimate of μ (at least if $n_i = J$ for all i).
Consider the equality

$$y_{ij} - \overline{y}_{\cdot\cdot} = (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}) + (y_{ij} - \overline{y}_{i\cdot})$$

and note that

$$\underbrace{\sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_{\cdot\cdot})^2}_{\text{SSTOT}} = \sum_{i=1}^{I} n_i (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 + \underbrace{\sum_{i=1}^{I} \sum_{j=1}^{n_i} (\overline{y}_{ij} - \overline{y}_{\cdot\cdot})^2}_{\text{SSE}}$$

which holds because the products sum to zero. Here, SSE is the unexplained or residual variation explained by the model.
Under the normality assumption, SSA and SSE are independent and

$$\frac{\text{SSA}}{\sigma^2} \sim \chi^2 \qquad \frac{\text{SSE}}{\sigma^2} \sim \chi^2$$

Source	df	SS	MS	F
Model	$I-1$	$SSA = \sum_{i=1}^{I} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$\frac{SSA}{I-1}$	$\frac{MSA}{MSE}$
Error	$n-I$	$SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2$	$\frac{SSE}{n-I}$	
Total	$n-1$	$SSTOT = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\cdot\cdot})^2$		

Note that for fixed i:

$$\frac{\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{n_i-1}$$

and these sums are independent and the total has

$$\text{df} = \sum_{i=1}^{I}(n_i - 1) = n - I$$

Aslo $\bar{y}_{i.}, \ldots, \bar{y}_{I.}$ are independent so SSR is like a SS for I independent r.v.s.
Under

$$H_0 = \alpha_1 = \ldots = \alpha_I$$

the F statistic $F = \text{MSA}/\text{MSE}$ has an F-distribution with I-1 and n-I df.

2.3 The analysis of variance in the one-way-layout

Source	df	SS	MS	F
Model	$I-1$	$SSA = \sum_{i=1}^{I} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$\frac{SSA}{I-1}$	$\frac{MSA}{MSE}$
Error	$n-I$	$SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2$	$\frac{SSE}{n-I}$	
Total	$n-1$	$SSTOT = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\cdot\cdot})^2$		

under $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_I$ the F-statistic $F = MSA/MSE$ has a F-distribution with $I-1$ and $n-I$ df.

3 Factors in the linear model

3.1 A matrix version of the one-way layout

The basic model for a one-way layout is

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

where $j = 1, \ldots, n$ and $i = 1, \ldots, I$

or

$$\begin{array}{lll} y_{1j} & = \mu_1 + \varepsilon_{1j} & j = 1, \ldots, n_1 \\ y_{2j} & = \mu_2 + \varepsilon_{2j} & j = 1, \ldots, n_2 \\ \vdots & \quad\vdots & \quad\vdots \\ y_{Ij} & = \mu_I + \varepsilon_{Ij} & j = 1, \ldots, n_I \end{array}$$

$\mu_i = \mu + \alpha_i$ in the earlier notation.

In a matrix form this can be written as

$$
\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ y_{I2} \\ \vdots \\ y_{In_I} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_I \end{bmatrix} + \varepsilon
$$

or $\mathbf{y} = \mathbf{X}\beta + \varepsilon$

Note that in the model $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, the $\mathbf{X}$ matrix would have a first column of all the 1s' and $\beta = \mu, \alpha_1, \alpha_2, \dots, \alpha_I$.

3.2 OLS estimates

Note that

$$
\underbrace{\underbrace{X'}_{I\times n} \cdot \underbrace{X}_{n\times I}}_{I\times I} = \begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & n_I \end{bmatrix}
$$

and

$$
X'y = \begin{bmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{n_I} y_{Ij} \end{bmatrix}
$$

so

$$
\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 1/n_1 & & \\ & \ddots & \\ & & 1/n_I \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{j=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{j=1}^{n_I} y_{Ij} \end{bmatrix}
$$

$$
\Rightarrow \hat{\beta} = \begin{bmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \vdots \\ \bar{y}_{I\cdot} \end{bmatrix}
$$

is the OLS estimate, and the MLE in this model

$$
y \sim N(X\beta, \sigma^2 I)
$$

3.3 Indicator variables

When a linear model has a factor as in

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}.$$

This model can always be rewritten in matrix form by using inicator varibles, as in the above example. Formally we can define x-variables which are 0 and 1 as needed.

3.4 Estimability

The original model formulation:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

The corresponding X-matrix would be

$$\begin{bmatrix} 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & 1 & \cdots & \cdots & 0 \\ 1 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ 1 & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & \cdots & 1 \\ 1 & 0 & 0 & \cdots & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & \cdots & 1 \end{bmatrix}_{n \times (I+1)}$$

and here $rank(X) = I$ so $X'X$ does not have an inverse and some restrictions are needed to estimate $\mu, \alpha_1, \ldots, \alpha_I$. Common restrictions include

$$\begin{aligned} \alpha_1 &= 0 \\ \alpha_I &= 0 \\ \sum \alpha_i &= 0 \\ \sum n_i \alpha_i &= 0 \end{aligned}$$

4 Analysis of variance: Implementation examples

4.1 An ANOVA model

Example:

	x	f	y
1	1	1	79.00
2	2	1	129.00
3	3	1	164.00
4	4	1	194.00
5	1	2	113.00
6	2	2	139.00
7	3	2	204.00
8	4	2	250.00

First model $y_i = \mu + \alpha_i + \varepsilon_{ij}$ where α_i is the effort of the factor, f, at level i. We can test the hypothesis of no factor effect

$$H_0: \quad \alpha_1 = \ldots = \alpha_I \tag{1}$$

using the one-way anova :

source	df	SS	MS	F
Between groups	I - 1	$\text{SSA} = \sum_{i=1}^{I} n_i(\bar{y}_{i.} - \bar{y}_{..})^2$	$\text{MSA} = \frac{\text{SSA}}{I-1}$	$\frac{\text{MSA}}{\text{MSE}}$
Within groups	n - I	$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2$	$\text{MSE} = \frac{\text{SSE}}{n-I}$	
Total	N-1	$\text{SSTOT} = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2$		

and we reject H_0 if $F > F_{I-1,n-I,1-\alpha}$.

f					$\sum y_{ij}$	$\bar{y}_{i.}$	$\sum(y_{ij} - \bar{y}_{..})^2$	$n_i(\bar{y}_{i.} - \bar{y}_{..})^2$
1	79	129	164	194	566	141.5	7325	1225
2	113	139	204	250	706	176.5	11597	1225
					1272		18922	2450

and the resulting ANOVA table is:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
f	1	2450.00	2450.00	0.78	0.4120
Residuals	6	18922.00	3153.67		

$F^* = F_{1,6,0.95} = 5.99$ so $F < F^*$ and we cannot reject H_0. Note that we have $\hat{\mu} = \bar{y}_{..} = 159$, $\hat{\alpha}_1 = \bar{y}_{1.} - \bar{y}_{..} = 17.5$ and $\hat{\alpha}_2 = \bar{y}_{2.} - \bar{y}_{..} = 17.5$

- Matrix version 1 :

$$X = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \Rightarrow (X'X)^{-1}X'y = \begin{bmatrix} 141.5 \\ 176.5 \end{bmatrix} \tag{2}$$

- Matrix version 2 :

$$X = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} \Rightarrow (X'X)^{-1}X'y = \begin{bmatrix} 176.5 \\ -35 \end{bmatrix} \tag{3}$$

Note that $\text{SSE} = ||\mathbf{y} - \mathbf{X}\hat{\beta}||^2$ in all formulations.

4.2 A regression model

We can test $H_0 : \beta_2$ in the model $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ using a t-test or an F-test from a appropiate anova table.

Example: same data

We fit using OLS to obtain

$$\hat{\beta}_2 = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_i - \overline{x})^2}$$
$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

and set

$$SSE = ||y - X\beta||^2$$
$$= \sum (y_i - (\beta_1 + \beta_2 x_i))^2$$

where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

and

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (X'X)^{-1}X'y$$

(Method 2)

Method 3

```
>summary(lm(y~x))
```

also gives the t-test. The anova table is obtained with

```
>anova(lm(y~x))
```

Table 1: Anova table

Source	df	SS	MS	F
Regression	$p-1=1$	SSR=SSTOT-SSE		
Error	$n-p=n-2$	SSE		
Total	$n-1$	SSTOT		

and we can test $H_0 : \beta_2 = 0$ by rejecting if $F > F^* = F_{1,n-2,1-\alpha}$

$$\hat{\beta}_1 = 52$$
$$\hat{\beta}_2 = 42.8$$
$$SSE = 3053.6$$
$$MSE = 508.93$$

4.3 The general F test

If model R is a restricted version of a full model then we can fit each one to obtain SSE(R) and SSE(F) and use an F-test to test weather R fits significantly worse than F.

$$F := \frac{\frac{SSE(R)-SSE(F)}{df(R)-df(F)}}{SSE(F)/df(F)}$$

Reject if $F > F*$.

4.4 The likelihood ratio test (LRT)

Under very general conditions $\{\log L(\hat{\theta}_F) - \log L(\hat{\theta}_R)\} \sim \chi^2_\gamma$ when $\hat{\theta}_R$ and $\hat{\theta}_F$ are from a fit of a reduced and full model, respectively and γ is the difference in the number of free parameters. Here L is the likelihood, which is maximized to obtain parameter estimates.

4.5 Type 1 vs Type 2 errors SS

The drop1() command evalates the effect of dropping individual terms from a model - evaluates the marginal effects (like summary for regression).
The anova() command evaluates the effect of inserting the terms in sequence.

5 The one way layout with random effects

5.1 The model

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$
$$\alpha_i \sim n(0, \sigma_A^2), \quad \varepsilon_{ij} \sim n(0, \sigma^2) \quad \text{independent}$$
$$1 \le j \le n_i, \;\; 1 \le i \le I$$

Note the changes from the fixed-effect version here

$$\begin{aligned} \mathrm{E}[y_{ij}] &= \mu \\ \mathrm{Var}[y_{ij}] &= \sigma_A^2 + \sigma^2 \\ \mathrm{Cov}[y_{ij}, y_{i'j'}] &= 0 \text{ if } i \neq i' \\ \mathrm{Cov}[y_{ij'}, y_{ij'}] &= \mathrm{Cov}(\alpha_i, \alpha_i) = \sigma_A^2 \end{aligned}$$

and

$$\rho = \mathrm{corr}(y_{ij}, y_{ij'}) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2}$$

Ex. Typically, the factor A and index correspond to a randomly selected subset of possible factor levels.

5.2 The ANOVA in the one-way-layout

Note that the partitioning of $y_{ij} - \overline{y}_{\cdot\cdot} = (\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}) + (y_{ij} - \overline{y}_{i\cdot})$ is still of interest. Assume $n_i = J$. If we look at $\overline{y}_{\cdot\cdot}$, we see that $\overline{y}_{\cdot\cdot} = \mu + \alpha_i + \overline{\varepsilon}_{i\cdot}$ and $\overline{y}_{\cdot\cdot} = \mu + \overline{\alpha}_i + \overline{\varepsilon}_{\cdot\cdot}$ so

$$\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot} = (\alpha_i - \overline{\alpha}_i) + (\overline{\varepsilon}_{i\cdot} - \overline{\varepsilon}_{\cdot\cdot})$$

and $\bar{y}_{i\cdot}$ are independent with

$$\begin{aligned} \mathrm{E}[\bar{y}_{i\cdot}] &= \mu \\ \mathrm{Var}[\bar{y}_{i\cdot}] &= \sigma_A^2 + \frac{\sigma^2}{J}, \quad i = 1,..,I \end{aligned}$$

and thus we have $\bar{y}_{i\cdot} \sim n(\mu, \sigma_A^2 + \sigma^2/J) \Rightarrow$

$$\frac{\sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{\sigma_A^2 + \sigma^2/J} = \frac{\sum_i J(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{J\sigma_A^2 + \sigma^2} = \frac{\mathrm{SSR}}{J\sigma_A^2 + \sigma^2} \sim \chi^2_{I-1}$$

Aslo note that

$$y_{ij} - \bar{y}_{i\cdot} = (\mu + \alpha_i + \varepsilon_{ij}) - (\mu + \alpha_i + \bar{\varepsilon}_{i\cdot}) = \varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}$$

(is like a deviation based on ε_{ij}) but $\varepsilon_{ij} \sim n(0, \sigma^2)$ iid so for fixed i

$$\frac{\sum_{j=1}^{J} (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2}{\sigma^2} \sim \chi^2$$

and therefore

$$\frac{\mathrm{SSE}}{\sigma^2} = \frac{\sum\sum(\varepsilon_{ij} - \varepsilon_{i\cdot})^2}{\sigma^2} \sim \chi^2_{I(J-1)}$$

We can also show that SSE and SSR are independent, so

$$F := \frac{\frac{\sum J(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{J\sigma_A^2 + \sigma^2} / (I-1)}{\frac{\sum\sum(y_{ij} - \bar{y}_{i\cdot})^2}{\sigma^2} / I(J-1)} \sim F_{I-1, I(J-1)}$$

and

$$F = \frac{\mathrm{SSR}/(I-1)}{\mathrm{SSE}/I(J-1)} \quad \text{if} \quad H_0 : \sigma_A^2 = 0 \text{ is true.}$$

$\star$ We can use the same ANOVA table.

5.3 Matrix formulation

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ y_{I2} \\ \vdots \\ y_{In_I} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{bmatrix} + \varepsilon$$

or

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{I1} \\ y_{I2} \\ \vdots \\ y_{In_I} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} [\mu] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{bmatrix} + \varepsilon$$

or

$$\mathbf{y} = X\beta + Z\mathbf{u} + \varepsilon$$

Where β contains the fixed effect parameters and $\mathbf{u}$ contains the random effects.

5.4 The ANOVA approach

Starting with the anova table

Source	df	SS	MS	F
Between group	$I-1$	$SSA = J\sum(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$\frac{SSA}{I-1} = MSA$	$\frac{MSA}{MSE}$
Within group	$I(J-1)$	$SSE = \sum\sum(y_{ij} - \bar{y}_{i\cdot})^2$	$\frac{SSE}{I(J-I)} = MSE$	
Total	$IJ-1$	$SSTOT = \sum\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2$		

From the above we know that

$$\begin{aligned} \mathrm{E}[MSE] &= \sigma^2 \\ \mathrm{E}[MSA] &= J\sigma_A^2 + \sigma^2 \end{aligned}$$

from with we obtain unbiased estimators

$$\begin{aligned} \hat{\sigma}^2 &= \text{MSE} \\ \hat{\sigma}_A^2 &= \frac{\text{MSA} - \text{MSE}}{J} \end{aligned}$$

5.5 Variance components

σ_A^2, σ^2 are the variance components
ρ intra-class correlation
The anova estimates are "natural", but

a) Do not extend easily

b) σ_A^2 can be negative

6 Estimation and prediction in the lmm

6.1 REML background

Recall that in the linear model, $\mathbf{y} \sim n(X\beta, \sigma^2\mathbf{I})$, we can write

$$||\mathbf{y} - X\beta||^2 = ||X\hat{\beta} - X\beta||^2 + ||\mathbf{y} - X\hat{\beta}||^2 (\text{ show this})$$

and hence the pdf/likelihood is

$$\begin{array}{c} \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{||\mathbf{y} - X\beta||^2}{2\sigma^2}\right) = \\ \frac{1}{(2\pi)^{p/2}\sigma^p} \exp\left(-\frac{||X\beta - X\hat{\beta}||^2}{2\sigma^2}\right) \cdot \\ \frac{1}{(2\pi)^{(n-p)/2}\sigma^{(n-p)}} \exp\left(-\frac{||\mathbf{y} - X\hat{\beta}||^2}{2\sigma^2}\right) \end{array}$$

- Maximizing over β gives $\beta = \hat{\beta}$.
- $X\hat{\beta} \sim n(X\beta, (X'X)^{-1}\sigma^2)$
- The estimate of σ^2 from here is $\hat{\sigma^2} = \frac{||y - X\hat{\beta}||^2}{n-p}$

6.2 The model

We take as a baseline model

$$\mathbf{y} = X\beta + Z\mathbf{u} + \varepsilon$$

with

$$\begin{array}{rcl} \varepsilon & \sim & n(0, \sigma^2 R) \\ u & \sim & n(0, \sigma^2 G) \end{array}$$

usually $\sigma^2 R = \sigma^2 \mathbb{I}$ and sometimes $\sigma^2 G = \sigma_A^2 \mathbb{I}$ (i.e. $G = \frac{\sigma_A^2}{\sigma^2}\mathbb{I}$).
In particular

$$\begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} \sim n\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 G & 0 \\ 0 & \sigma^2 R \end{pmatrix} \right)$$

6.3 Best linear unbiased prediction

Best linear unbiased prediction, Henderson($\sim$1953) looked at the joint density of y and u and suggested maximizing this to estimate an predict u.
The joint density is obtained from

$$\begin{array}{rcl} \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \end{pmatrix} & = & \begin{pmatrix} X\beta + Z\mathbf{u} + \varepsilon \\ \mathbf{u} \end{pmatrix} \\ & = & \begin{pmatrix} X\beta \\ 0 \end{pmatrix} + \begin{pmatrix} Z & \mathbb{I} \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \varepsilon \end{pmatrix} \\ & \sim & n\left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} & \mathbb{I} \\ Z & \mathbb{I} \end{pmatrix}^T \begin{pmatrix} \sigma^2 G & 0 \\ 0 & \sigma^2 R \end{pmatrix} \begin{pmatrix} Z & \mathbb{I} \\ 0 & \mathbb{I} \end{pmatrix} \right) \end{array}$$

so $f_\beta(y,u) = \ldots$ see 4.1 in Robinson (http://hi.is/ gunnar/kennsla/lmm/robinson91.pdf)
Maximizing $f_\beta(y,u)$ over β and u leeds to the estimating equations:

$$X'R^{-1}X\hat{\beta} + X'R^{-1}Z\hat{u} = X'R^{-1}y$$
$$Z'R^{-1}X\hat{\beta} + (Z'R^{-1}Z + G^{-1})\hat{u} = Z'R^{-1}y$$

If R,G are unknown.

6.4 The likelihood

The likelihood is based only on the pdf of **y**. We know **y** is Gaussian and we can easily derive

$$\begin{aligned} \mathrm{E}[\mathbf{y}] &= X\beta \\ \mathrm{Var}[\mathbf{y}] &= \mathrm{Var}[Z\mathbf{u}] + \mathrm{Var}[\varepsilon] \\ &= \sigma^2 ZGZ' + \sigma^2 R \end{aligned}$$

so

$$\mathbf{y} \sim n(X\beta, \sigma^2(ZGZ' + R))$$

From this we can find the MLE for β as

$$\hat{\beta} = [X'(R + ZGZ')^{-1}X]^{-1}X'(R + ZGZ')^{-1}\mathbf{y}$$

Note that if we set

$$\hat{u} = \ldots \text{ see Robinson 1991, p.19}$$

then $\begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix}$ solves the estimation equation.

7 lmm with R

8 variations??2-way layout??