Introduction to General and Generalized Linear Models

Mixed effects models - Part I

Henrik Madsen
Poul Thyregod

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

January 2011

This lecture

- Introduction
- Gaussian mixed effects model
- One-way random effects model

Introduction

- We will at first consider a general class of models for the analysis of *grouped data*.
- It is assumed that the (possibly experimental) conditions within groups are the same, whereas the conditions vary between groups.
- For most experimental conditions it is reasonable to talk about *repetitions* for the observations within groups.

Introduction

We will initially represent the observations using the following table

Group	Observations
1	$Y_{11}, Y_{12}, \ldots, Y_{1n_1}$
2	$Y_{21}, Y_{22}, \ldots, Y_{2n_2}$
$\vdots$	$\vdots, \vdots$
k	$Y_{k1}, Y_{k2}, \ldots, Y_{kn_k}$

which corresponds to a so-called *classification* in k groups (cells) with $n_i, (i = 1, 2, \ldots, k)$ observations in each group.

Introduction

- If the same number of observations or repetitions is available for each of the k homogeneous *groups*, we say that the experiment is *balanced*.
- The grouping may be the result of one or several *factors*, and each set of factor levels defines a single (homogeneous) situation or *treatment*, often called a *cell*.
- The possible values of the factors are often called *levels*.
- If the factor is “sex”, the levels are “male” and “female”.
- For the so-called *factorial experiment* all the explanatory variables are *categorical*, and often called factors.
- We will also consider situations where such variables are combined with for instance continuous quantitative variables.

Introduction

Consider again a one way fixed effecs ANOVA model as in Example 3.2.

- Sometimes it might be more reasonable to consider the levels as an outcome of picking a number of groups in a *large population*, where only the variation between groups within this population is of interest and not the specific level for each group as for the fixed effects model.
- This leads to a simple *hierarchical model* where the levels are considered as random variables, and this gives rise to the so-called *random effects models*.
- Models containing both fixed and random effects are called *mixed effects models* or just *mixed models*.
- The hierarchical structure arises here from the fact that the so-called first stage model describes the observations given the random effects, and the second stage model is a model for these random effects.

Introduction

- In a general setting mixed models describes dependence between observations within and between groups by assuming the existence of one or more unobserved *latent variables* for each group of data.
- The latent variables are assumed to be random and hence referred to as *random effects*.
- Hence a mixed model consists of both fixed model parameters $\boldsymbol{\theta}$ and random effects $\boldsymbol{U}$, where the random effects are described by another model and hence another set of parameters describing the assumed distribution for the random effects.
- A key feature of mixed models is that, by introducing random effects in addition to fixed effects, they allow you to address multiple sources of variation, e.b. they allow you to take into account both within and between subject or group variation.

Gaussian mixed model

Definition (Gaussian mixed model)

A general formulation of the *Gaussian mixed model* is

$$\boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} \sim N(\boldsymbol{\mu}(\boldsymbol{\beta}, \boldsymbol{u}), \boldsymbol{\Sigma}(\boldsymbol{\beta})) \tag{1}$$

$$\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{\Psi}(\boldsymbol{\psi})) \tag{2}$$

where the dimension of $\boldsymbol{U}$ and therefore $\boldsymbol{\Psi}$ might be large, but $\boldsymbol{\psi}$ is generally small.

It is clearly seen that the model is also a so-called *hierarchical model* where (1) and (2) are the first and second stage model, respectively.

Gaussian mixed effects

It is seen that the conditional distribution of $\boldsymbol{Y}$ given the outcome, $\boldsymbol{u}$ of the *random effect* $\boldsymbol{U}$ is Gaussian.
Assuming that the random effects are scalar and independent and that the residuals within groups are independent, then

$$\begin{aligned} Y_{ij}|U_i = u_i &\sim N(\mu_{ij}(\boldsymbol{\beta}, u_i), \sigma^2) \\ U_i &\sim N(0, \sigma_u^2), i = 1, \dots, k; j = 1, \dots, n_i \end{aligned}$$

where k is the number of groups and n_i is the number of observations within group i.

Gaussian mixed effects

Due to the within group independence of the residuals this is equivalently written

$$\begin{aligned} \boldsymbol{Y}_i | U_i = u_i &\sim N(\boldsymbol{\mu}_i(\boldsymbol{\beta}, u_i), \sigma^2 \boldsymbol{I}) \\ U_i &\sim N(0, \sigma_u^2), \ i = 1, \dots, k \end{aligned}$$

Using the mean value function here, where the random effect is scalar, this case is equivalently written

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}, U_i) + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{\epsilon}_i | U_i = u_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}), \quad U_i \sim N(0, \sigma_u^2)$$

where $\boldsymbol{\mu}_i(\boldsymbol{\beta}, u_i)$ is the *mean value function*.

Gaussian mixed effects

If $\boldsymbol{\mu}(\boldsymbol{\beta}, \boldsymbol{U})$ is nonlinear in $\boldsymbol{\beta}$ then we have a *nonlinear mixed model*, whereas a model with the mean value function

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{U}$$

with $\boldsymbol{X}$ and $\boldsymbol{Z}$ denoting known matrices, is called a *linear mixed model*

Notice how the mixed effect linear model in is a linear combination of *fixed effects*, $\boldsymbol{X}\boldsymbol{\beta}$ and *random effects*, $\boldsymbol{Z}\boldsymbol{U}$.

One-way random effects model - example

Unprocessed (baled) wool contain varying amounts of fat and other impurities that need to be removed before further processing. The price - and the value of the baled wool depends on the amount of pure wool that is left after removal of fat and impurities. The purity of the baled wool is expressed as the mass percentage of pure wool in the baled wool.

As part of the assessment of different sampling plans for estimation of the purity of a shipment of several bales of wool has U.S.Customs Laboratory, Boston selected 7 bales at random from a shipment of Uruguyan wool, and from each bale, 4 samples were selected for analysis.

Data

	Bale no.						
Sample	1	2	3	4	5	6	7
1	52.33	56.99	54.64	54.90	59.89	57.76	60.27
2	56.26	58.69	57.48	60.08	57.76	59.68	60.30
3	62.86	58.20	59.29	58.72	60.26	59.58	61.09
4	50.46	57.35	57.51	55.61	57.53	58.08	61.45
Bale average	55.48	57.81	57.23	57.33	58.86	58.78	60.78

Table: The purity in % pure wool for 4 samples from each of 7 bales of Uruguyan wool.

Model with fixed effects

We could formulate a one-way model as discussed in Chapter 3:

$$\mathcal{H}_1 : Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2) \quad i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, n_i$$

and obtain the ANOVA table:

Variation	Sum of Squares		f	$s^2 = \mathrm{SS}/f$	F-value	Prob > F
Between bales	SSB	65.9628	6	10.9938	1.76	0.16
Within bales	SSE	131.4726	21	6.2606		
Total	SST	197.4348	27			

The test statistic for $\mathcal{H}_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ is $F = 10.99/6.26 = 1.76$.

Model with fixed effects

Such a model would be relevant, if we had selected seven specific bales, eg the bales with identification labels “AF37Q”, “HK983”, . . ., and “BB837”.

Thus, $i = 1$ would refer to bale “AF37Q”, and the probability distributions would refer to repeated sampling, but under such imaginative repeated sampling, $i = 1$ would always refer to this specific bale with label “AF37Q”.

Model with random effects

However, although there is not strong evidence against $\mathcal{H}_1$, we will not consider the bales to have the same purity. The idea behind the sampling was to *describe the variation* in the shipment, and the purity of the seven selected bales was not of interest in it self, but rather as representative for the variation in the shipment.

Therefore, instead of the *fixed effects* model in Chapter 3, we shall introduce a *random effects* model.

One-way model with random effects

Definition (One-way model with random effects)

Consider the random variables $Y_{ij}, i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n_i$
A *one-way random effects model* for Y_{ij} is a model such that

$$Y_{ij} = \mu + U_i + \epsilon_{ij},$$

with $\epsilon_{ij} \sim N(0, \sigma^2)$ and $U_i \sim N(0, \sigma_u^2)$, and where ϵ_{ij} are mutually independent, and also the U_i's are mutually independent, and finally the U_i's are independent of ϵ_{ij}.
We shall put

$$N = \sum_{i=1}^{k} n_i$$

When all groups are of the same size, $n_i = n$, we shall say that the model is *balanced*.

Parameters in the one-way random effects model

Parameters in the one-way random effects model

Consider a one-way random effects model as specified before. The (*fixed*) parameters of the model are $(\mu, \sigma^2, \sigma_u^2)$.
Sometimes, the *signal to noise ratio*

$$\gamma = \frac{\sigma_u^2}{\sigma^2}$$

is used instead of σ_u^2. Thus, the parameter γ expresses the inhomogeneity between groups in relation to the internal variation in the groups. We shall use the term *signal/noise ratio* for the parameter γ.

The one-way model as a hierarchical model

The one-way model as a hierarchical model

Putting $\mu_i = \mu + U_i$ we may formulate the one way random effects model as a *hierarchical model*, where we shall assume that

$$Y_{ij}|\mu_i \sim N(\mu_i, \sigma^2)$$

and in contrast to the *systematic/fixed model*, the bale level μ_i is modeled as a realization of a random variable,

$$\mu_i \sim N(\mu, \sigma_u^2),$$

where the μ_i's are assumed to be mutually independent, and Y_{ij} are *conditionally independent*, i.e. Y_{ij} are mutually independent in the conditional distribution of Y_{ij} for given μ_i.

Interpretation of the one-way random effect model

- The random effects model will be a reasonable model in situations where the interest is not restricted alone to the experimental conditions at hand, but where the experimental conditions rather are considered as representative for a larger collection (population) of varying experimental conditions, in principle selected at random from that population.

- The analysis of the systematic model puts emphasis on the assessment of the results in the individual groups, μ_i, and possible differences, $\mu_i - \mu_h$, between the results in specific groups, whereas the analysis of the random effects model aims at describing the variation between the groups, $\mathrm{Var}[\mu_i] = \sigma_u^2$.

Marginal and joint distributions

Theorem (Marginal distributions in the random effects model for one way analysis of variance)

The marginal distribution of Y_{ij} *is a normal distribution with*

$$\mathrm{E}[Y_{ij}] = \mu$$

$$\mathrm{Cov}[Y_{ij}, Y_{hl}] = \begin{cases} \sigma_u^2 + \sigma^2 & \textit{for } (i,j) = (h,l) \\ \sigma_u^2 & \textit{for } i = h, j \neq l \\ 0 & \textit{for } i \neq h \end{cases}$$

Observations from the same group

Observations from the same group are correlated

We note that there is a positive covariance between observations from the same group. This positive covariance expresses that observations within the same group will deviate in the same direction from the mean, μ, in the marginal distribution, viz. in the direction towards the group mean in question.

The coefficient of correlation,

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2} = \frac{\gamma}{1 + \gamma}$$

that describes the correlation within a group, is often termed the *intraclass correlation*.

Distribution of individual group averages

Distribution of individual group averages

We finally note that the simultaneous distribution of the group averages is characterized by

$$\mathrm{Cov}[\bar{Y}_{i\cdot}, \bar{Y}_{h\cdot}] = \begin{cases} \sigma_u^2 + \sigma^2/n_i & \text{for} \quad i = h \\ 0 & \text{otherwise} \end{cases}$$

That is, that the k group averages $\bar{Y}_{i\cdot}, i = 1, 2, \ldots, k$ are mutually independent, and that the variance of the group average

$$\mathrm{Var}[\bar{Y}_{i\cdot}] = \sigma_u^2 + \sigma^2/n_i = \sigma^2(\gamma + 1/n_i)$$

includes the variance of the random component, $\sigma_u^2 = \sigma^2\gamma$, as well as the effect of the residual variance on the group average.
Thus, an increase of the sample size in the individual groups will improve the precision by the determination of the group mean α_i,but the variation between the individual groupmeans is not reduced by this averaging.

Observation vector for a group

When we consider the set of observations corresponding to the i'th group as a n_i-dimensional column vector,

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix}$$

The set of observations $\boldsymbol{Y}_i, i = 1, 2, \ldots, k$ may be described as k *independent observations* of a n_i dimensional variable $\boldsymbol{Y}_i \sim \mathrm{N}_{n_i}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_{n_i} + \sigma_b^2 \boldsymbol{J}_{n_i})$, i.e. that the dispersion matrix for $\boldsymbol{Y}_i$ is

$$\begin{aligned} \boldsymbol{V}_i &= \mathrm{D}[Y_i] \\ &= \mathrm{E}[(\boldsymbol{Y}_i - \boldsymbol{\mu})(\boldsymbol{Y}_i - \boldsymbol{\mu})^T] \\ &= \begin{pmatrix} \sigma_b^2 + \sigma^2 & \sigma_b^2 & \ldots & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 + \sigma^2 & \ldots & \sigma_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_b^2 & \sigma_b^2 & \ldots & \sigma_b^2 + \sigma^2 \end{pmatrix} \end{aligned}$$

where $\boldsymbol{J}_{n_i}$ is a $n_i \times n_i$-dimensional matrix consisting solely by 1's.

Covariance structure for the whole set of observations

Covariance structure for the whole set of observations

If we organize all observations in one column, organized according to groups, we observe that the $N \times N$-dimensional dispersion matrix $\mathrm{D}[\boldsymbol{Y}]$ is

$$\boldsymbol{V} = \mathrm{D}[\boldsymbol{Y}] = \text{ Block diag}\{\boldsymbol{V}_i\}$$

This illustrates that observations from *different groups are independent*, whereas observations from the *same group are correlated*.

Test of hypothesis of homogeneity

$$\mathrm{SSE} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2$$

$$\overline{\overline{y}} = \sum_{i} n_i \overline{y}_{i\cdot} \Big/ N$$

$$\mathrm{SSB} = \sum_{i=1}^{k} n_i (\overline{y}_i - \overline{\overline{y}})^2$$

$$\mathrm{SST} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{\overline{y}})^2 = \mathrm{SSB} + \mathrm{SSE}$$

Furthermore we introduce the *shrinkage factor*

$$w_i(\gamma) = \frac{1}{1 + n_i \gamma}$$

the importance of $w_i(\gamma)$ is seen as $\mathrm{Var}[\overline{Y}_{i\cdot}] = \sigma^2 / w_i(\gamma)$.

Test of hypothesis of homogeneity

Under the random effects model, the hypothesis that the varying experimental conditions do not have an effect on the observed values, is formulated as

$$H_0 : \sigma_u^2 = 0.$$

The hypothesis is tested by comparing the variance ratio with the quantiles in a $F(k-1, N-k)$-distribution.

Test of the hypothesis of homogeneity in the random effects model

Theorem

For the one-way model as a hierarchical model the likelihood ratio test for the hypothesis has the test statistic

$$Z = \frac{\mathrm{SSB}\,/(k-1)}{\mathrm{SSE}\,/(N-k)}$$

Large values of z are considered as evidence against the hypothesis.

Under the hypothesis, Z will follow a $F(k-1, N-k)$-distribution. In the balanced case, $n_1 = n_2 = \ldots = n_k = n$, we can determine the distribution of Z also under the alternative hypothesis. In this case we have

$$Z \sim (1 + n\gamma)F(k-1, N-k).$$