

## Module 5: Hierarchical random effects

Anders Nielsen

[anielsen@dina.kvl.dk](mailto:anielsen@dina.kvl.dk)

## Main example: Lactase in piglets<sup>a</sup>

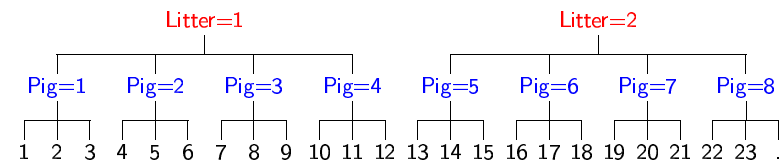
| Obs | litter | pig | reg | status | loglact | Obs | litter | pig | reg | status | loglact |
|-----|--------|-----|-----|--------|---------|-----|--------|-----|-----|--------|---------|
| 1   | 1      | 1   | 1   | 1      | 1.89537 | 31  | 3      | 11  | 1   | 2      | 1.58104 |
| 2   | 1      | 1   | 2   | 1      | 1.97046 | 32  | 3      | 11  | 2   | 2      | 1.52606 |
| 3   | 1      | 1   | 3   | 1      | 1.78255 | 33  | 3      | 11  | 3   | 2      | 1.65058 |
| 4   | 1      | 2   | 1   | 1      | 2.24496 | 34  | 4      | 12  | 1   | 1      | 1.97162 |
| 5   | 1      | 2   | 2   | 1      | 1.43413 | 35  | 4      | 12  | 2   | 1      | 2.11342 |
| 6   | 1      | 2   | 3   | 1      | 2.16905 | 36  | 4      | 12  | 3   | 1      | 2.51278 |
| 7   | 1      | 3   | 1   | 2      | 1.74222 | 37  | 4      | 13  | 1   | 1      | 2.06739 |
| 8   | 1      | 3   | 2   | 2      | 1.84277 | 38  | 4      | 13  | 2   | 1      | 2.25631 |
| 9   | 1      | 3   | 3   | 2      | 0.17479 | 39  | 4      | 13  | 3   | 1      | 1.79251 |
| 10  | 1      | 4   | 1   | 2      | 2.12704 | 40  | 5      | 14  | 1   | 1      | 1.93274 |
| 11  | 1      | 4   | 2   | 2      | 1.90954 | 41  | 5      | 14  | 2   | 1      | 1.82394 |
| 12  | 1      | 4   | 3   | 2      | 1.49492 | 42  | 5      | 14  | 3   | 1      | 1.23629 |
| 13  | 2      | 5   | 1   | 2      | 1.62897 | 43  | 5      | 15  | 1   | 2      | 2.07386 |
| 14  | 2      | 5   | 2   | 2      | 2.26642 | 44  | 5      | 15  | 2   | 2      | 1.96713 |
| 15  | 2      | 5   | 3   | 2      | 1.96763 | 45  | 5      | 15  | 3   | 2      | 0.47971 |
| 16  | 2      | 6   | 1   | 2      | 2.01948 | 46  | 5      | 16  | 1   | 2      | 2.01307 |
| 17  | 2      | 6   | 2   | 2      | 2.56443 | 47  | 5      | 16  | 2   | 2      | 1.85483 |
| 18  | 2      | 6   | 3   | 2      | 1.16387 | 48  | 5      | 16  | 3   | 2      | 2.18274 |
| 19  | 2      | 7   | 1   | 2      | 2.20681 | 49  | 5      | 17  | 1   | 2      | 2.86629 |
| 20  | 2      | 7   | 2   | 2      | 2.55652 | 50  | 5      | 17  | 2   | 2      | 2.71414 |
| 21  | 2      | 7   | 3   | 2      | 1.69358 | 51  | 5      | 17  | 3   | 2      | 1.60533 |
| 22  | 2      | 8   | 1   | 2      | 1.09186 | 52  | 5      | 18  | 1   | 2      | 1.97865 |
| 23  | 2      | 8   | 2   | 2      | 1.93091 | 53  | 5      | 18  | 2   | 2      | 1.93342 |
| 24  | 2      | 8   | 3   | 2      | .       | 54  | 5      | 18  | 3   | 2      | 0.74943 |
| 25  | 3      | 9   | 1   | 1      | 2.36462 | 55  | 5      | 19  | 1   | 2      | 2.89886 |
| 26  | 3      | 9   | 2   | 1      | 2.72261 | 56  | 5      | 19  | 2   | 2      | 2.88606 |
| 27  | 3      | 9   | 3   | 1      | 2.80336 | 57  | 5      | 19  | 3   | 2      | 2.20697 |
| 28  | 3      | 10  | 1   | 1      | 2.42834 | 58  | 5      | 20  | 1   | 2      | 1.87733 |
| 29  | 3      | 10  | 2   | 1      | 2.64971 | 59  | 5      | 20  | 2   | 2      | 1.70260 |
| 30  | 3      | 10  | 3   | 1      | 2.54788 | 60  | 5      | 20  | 3   | 2      | 1.11077 |

<sup>a</sup>Data kindly supplied by: Charlotte Reinhard Bjørnvad, Division of Animal Nutrition (RVAU)

## Aim of this module

- Present example where hierarchial error structure is natural
- Introduce one, two, three,... layer models
- Investigate the covariance structure
- See how to specify such structures in SAS

## The hierarchial structure of the lactase data set



- The first layer (counting from the bottom) is the individual measurements  $i = 1 \dots N$ .
- The second layer is the pigs
  - Pigs could be different, and the model should allow for this
  - If pigs are different, then two measurements on the same pig would be more similar than two from different pigs
- The third layer is litters
  - If litters are different two measurements from the same litter, but different pigs should be correlated, but maybe not as correlated as two from the same pig

(Imagine we had information about farm, area,...)

## One layer model

- The simplest hierarchical model:

$$y_i = \mu + \varepsilon_i, \quad \varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2)$$

- The mean value parameter(s) are estimated by:

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

- The single variance parameter is estimated by:

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{i=1}^N (y_i - \hat{\mu}_i)^2$$

Here is  $p$  the number of mean value parameters, and  $\hat{\mu}_i = \mathbf{X}\hat{\beta}$

5/11

present5.tex – May 1, 2003

## Two layer model

- The simplest *real* hierarchical model:

$$y_i = \mu + a(\text{sub}_i) + \varepsilon_i$$

where  $a(\text{sub}_i) \sim N(0, \sigma_a^2)$ , and  $\varepsilon_i \sim N(0, \sigma^2)$ , and all (both  $a$ 's and  $\varepsilon$ 's) are independent

- The covariance structure is:

$$\text{cov}(y_{i_1}, y_{i_2}) = \begin{cases} 0 & , \text{ if } \text{sub}_{i_1} \neq \text{sub}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_a^2 & , \text{ if } \text{sub}_{i_1} = \text{sub}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_a^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}$$

- Or the same expressed via the covariance matrix:

$$\mathbf{V} = \begin{pmatrix} \sigma_a^2 + \sigma^2 & \sigma_a^2 & \sigma_a^2 & 0 & 0 & 0 \\ \sigma_a^2 & \sigma_a^2 + \sigma^2 & \sigma_a^2 & 0 & 0 & 0 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_a^2 + \sigma^2 & \sigma_a^2 & \sigma_a^2 \\ 0 & 0 & 0 & \sigma_a^2 & \sigma_a^2 + \sigma^2 & \sigma_a^2 \\ 0 & 0 & 0 & \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma^2 \end{pmatrix}$$

7/11

present5.tex – May 1, 2003

## Lactase data with one layer model

Consider again the lactase data set, but ignore information about pig and litter

- A relevant one layer model could be:

$$y_i = \mu + \alpha(\text{status}_i) + \beta(\text{reg}_i) + \gamma(\text{status}_i, \text{reg}_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2)$$

- To specify this model in SAS, we write:

```
proc mixed method=REML;
  class status reg;
  model logLact = status reg status*reg;
run;
```

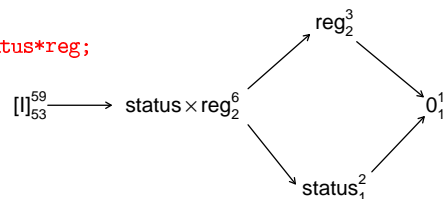
- Part of the SAS output is:

```
Covariance Parameter Estimates
Cov Parm      Estimate
Residual      0.2457
```

```
Fit Statistics -2 Res Log Likelihood      89.5
```

```
Type 3 Tests of Fixed Effects
```

| Effect     | Num DF | Den DF | F Value | Pr > F |
|------------|--------|--------|---------|--------|
| status     | 1      | 53     | 4.73    | 0.0341 |
| reg        | 2      | 53     | 3.11    | 0.0529 |
| status*reg | 2      | 53     | 2.86    | 0.0659 |



6/11

present5.tex – May 1, 2003

## Lactase data with two layer model

- Now we can include the information about pig:

$$y_i = \mu + \alpha(\text{status}_i) + \beta(\text{reg}_i) + \gamma(\text{status}_i, \text{reg}_i) + d(\text{pig}_i) + \varepsilon_i,$$

where  $d(\text{pig}_i) \sim N(0, \sigma_d^2)$ ,  $\varepsilon_i \sim N(0, \sigma^2)$ , and all independent.

- The SAS code for this model is:

```
proc mixed method=REML;
  class status pig reg;
  model logLact = status reg status*reg /ddfm=satterth;
  random pig;
run;
```

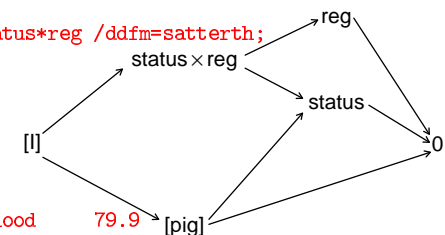
- Part of the SAS output is:

```
Covariance Parameter Estimates
Cov Parm      Estimate
pig            0.1119
Residual      0.1349
```

```
Fit Statistics -2 Res Log Likelihood      79.9
```

```
Type 3 Tests of Fixed Effects
```

| Effect     | Num DF | Den DF | F Value | Pr > F |
|------------|--------|--------|---------|--------|
| status     | 1      | 17.7   | 2.65    | 0.1211 |
| reg        | 2      | 34.9   | 6.09    | 0.0054 |
| status*reg | 2      | 34.9   | 5.63    | 0.0076 |



8/11

present5.tex – May 1, 2003

## Three or more layers

- The Basic three layer model is denoted:

$$y_i = \mu + a(\text{block}_i) + b(\text{sub}_i) + \varepsilon_i$$

Here  $a(\text{block}_i) \sim N(0, \sigma_a^2)$ ,  $b(\text{sub}_i) \sim N(0, \sigma_b^2)$ , and  $\varepsilon_i \sim N(0, \sigma^2)$  all independent.

- The interpretation of the variance parameters:

$\sigma^2$  is the variance between observations from the same subject

$\sigma_b^2$  is the variance between subjects within the same block

$\sigma_a^2$  is the variance between blocks

- The covariance structure:

$$\text{cov}(y_{i_1}, y_{i_2}) = \begin{cases} 0 & , \text{ if } \text{block}_{i_1} \neq \text{block}_{i_2} \\ \sigma_a^2 & , \text{ if } \text{block}_{i_1} = \text{block}_{i_2} \text{ and } \text{sub}_{i_1} \neq \text{sub}_{i_2} \\ \sigma_a^2 + \sigma_b^2 & , \text{ if } \text{sub}_{i_1} = \text{sub}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_a^2 + \sigma_b^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}$$

## Comparing different variance structures

- Compare two models  $A$  and  $B$ , where  $B$  is a *sub-model* of  $A$

- We use the restricted/residual likelihood ratio test:

1. Compute the test value:

$$G_{A \rightarrow B} = 2\ell_{re}^{(B)} - 2\ell_{re}^{(A)}$$

2. Lookup the P-value in a  $\chi_{df}^2$  statistical table, where  $df = \dim(A) - \dim(B)$

### For the lactase data

| Model                        | $2\ell_{re}$ | G-value                     | df | P-value                        |
|------------------------------|--------------|-----------------------------|----|--------------------------------|
| (A) Pig and litter included  | 79.6         | $G_{A \rightarrow B} = 0.3$ | 1  | $P_{A \rightarrow B} = 0.5839$ |
| (B) Only pig included        | 79.9         | $G_{B \rightarrow C} = 9.6$ | 1  | $P_{B \rightarrow C} = 0.0019$ |
| (C) Independent observations | 89.5         |                             |    |                                |

## Lactase data with three layer model

- The model with both litter and pig included:

$$y_i = \mu + \alpha(\text{status}_i) + \beta(\text{reg}_i) + \gamma(\text{status}_i, \text{reg}_i) + d(\text{litter}_i) + e(\text{pig}_i) + \varepsilon_i$$

where  $d(\text{litter}_i) \sim N(0, \sigma_d^2)$ ,  $e(\text{pig}_i) \sim N(0, \sigma_e^2)$ ,  $\varepsilon_i \sim N(0, \sigma^2)$ , all independent.

- The SAS code for this model is:

```
proc mixed method=REML nobound;
  class litter status pig reg;
  model logLact = status reg status*reg /ddfm=satterth;
  random litter pig;
run;
```

- Part of the SAS output is:

```
Covariance Parameter Estimates
Cov Parm      Estimate
litter        -0.01718
pig           0.1261
Residual      0.1348
```

```
Fit Statistics -2 Res Log Likelihood      79.6
```

```
Type 3 Tests of Fixed Effects
```

```
Effect      Num    Den    F Value    Pr > F
status      1      9.02    2.62    0.1400
reg         2      35     6.11    0.0053
status*reg  2      35     5.65    0.0075
```

