Course topics (tentative)

- linear mixed models
- likelihood-based inference
- generalized linear mixed models
- computational methods
- estimating equations (depending on time)
- Bayesian inference and Monte Carlo methods (depending on time)

The role of random effects

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

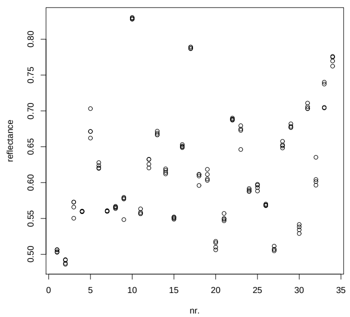
February 8, 2012

Outline for today

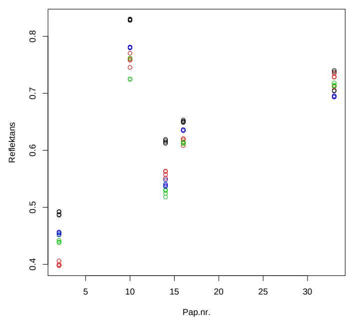
- examples of data sets.
- analysis of variance
- multivariate normal distribution
- linear mixed models

Reflectance (colour) measurements for samples of cardboard (egg trays)

Four replications at same position on each cardboard



For five cardboards: four replications at five positions at each cardboard

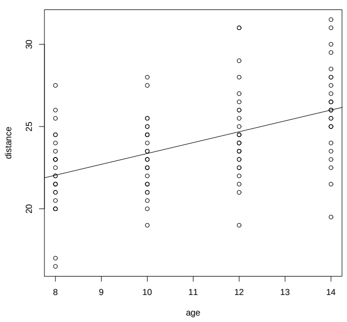


Colour variation between/within cardboards ?

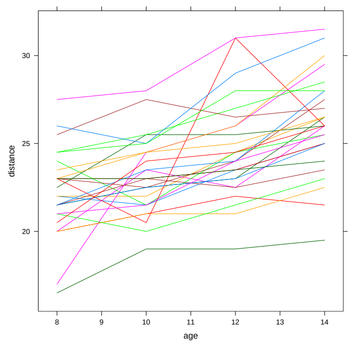
Orthodontic growth curves

Distance between pituitary and the pterygomaxillary fissure for children of age 8-14

Distance versus age:



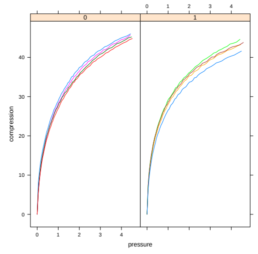
Distance versus age grouped according to child



Different intercepts for different children

Compression of mats for cows

Compression vs. pressure for two brands of mats



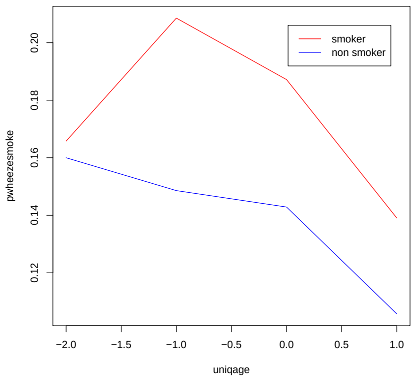
Non-linear relation

$$y = \frac{ab + cx^d}{b + x^d},$$

Random variation between mats of same brand, small measurement noise.

Wheezing

Probability of wheezing (astma) in relation to age and smoking habits of mother:

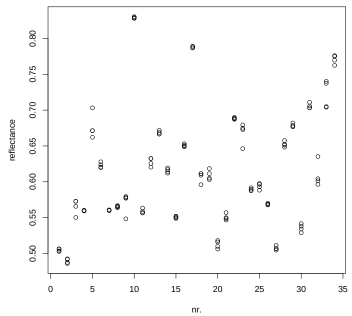


Original data binary: wheezed or not for each of 4 years for each child.

Correlation between measurements for the same child ?

Model for reflectances: one-way anova

Four replications on each cardboard



Models:

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

or

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where μ and α_i are fixed unknown parameters and ϵ_{ij} stochastic noise or

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

where U_i are random variables

Which is most relevant ?

The role of random effects

Quantify sources of variation (e.g. quality control): is pulp for paper production too heterogeneous ?

Decomposition of variance: $\mathbb{V}\mathrm{ar} Y_{ij} = \mathbb{V}\mathrm{ar} U_i + \mathbb{V}\mathrm{ar} \epsilon_{ij} = \sigma^2 + \tau^2$

Covariances:

$$\mathbb{C}\mathrm{ov}[Y_{ij}, Y_{i'j'}] = \begin{cases} 0 & i \neq i' \\ \mathbb{V}\mathrm{ar} U_i & i = i', j \neq j' \\ \mathbb{V}\mathrm{ar} U_i + \mathbb{V}\mathrm{ar} \epsilon_{ij} & i = i', j = j' \end{cases}$$

Correlations:

$$\mathbb{C}\mathrm{orr}[Y_{ij}, Y_{i'j'}] = \begin{cases} 0 & i \neq i' \\ \sigma^2/(\sigma^2 + \tau^2) & i = i', j \neq j' \\ 1 & i = i', j = j' \end{cases}$$

That is, observations for same cardboard are correlated !

Implications: evaluation of uncertainty

Correct evaluation of uncertainty of estimates of fixed effects: suppose we wish to estimate $\mu = \mathbb{E} Y_{ij}$. Due to correlation, observations on same cardboard to some extent redundant.

Model ignoring variation between cardboards

$$Y_{ij} = \mu + \epsilon_{ij}, i = 1, \dots, m, j = 1, \dots, k$$

$$\mathbb{V}\mathrm{ar} \epsilon_{ij} = \sigma^2 + \tau^2$$

$$\mathbb{V}\mathrm{ar} \bar{Y}_{\cdot\cdot} = \frac{\sigma^2 + \tau^2}{mk}$$

Model with random cardboard effects

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

$$\mathbb{V}\mathrm{ar} U_i = \sigma^2, \quad \mathbb{V}\mathrm{ar} \epsilon_{ij} = \tau^2$$

$$\mathbb{V}\mathrm{ar} \bar{Y}_{\cdot\cdot} = \frac{\sigma^2}{m} + \frac{\tau^2}{mk}$$

With first model, variance is underestimated !

Break

Show results regarding variances on two previous slides.

Two levels of random effects

For five cardboards we have 4 replications at 4 positions.

Hierarchical model (nested random effects)

$$Y_{ipj} = \mu + U_i + U_{ip} + \epsilon_{ipj}$$

$$\mathbb{V}\mathrm{ar} Y_{ipj} = \sigma^2 + \omega^2 + \tau^2$$

Covariance structure for nested random effects model

$$Y_{ipj} = \mu + U_i + U_{ip} + \epsilon_{ipj}$$

$$\mathbb{C}\mathrm{ov}(Y_{ipj}, Y_{lqk}) = \begin{cases} 0 & i \neq l \\ \sigma^2 & i = l, p \neq q \text{ same card} \\ \sigma^2 + \omega^2 & i = l, p = q \text{ same card and pos.} \\ \sigma^2 + \omega^2 + \tau^2 & i = 1, p = q, k = j \quad (\mathbb{V}\mathrm{ar} Y_{ipj}) \end{cases}$$

Correlation structure for nested random effects model

$$Y_{ipj} = \mu + U_i + U_{ip} + \epsilon_{ipj}$$

$$\mathbb{C}\mathrm{orr}(Y_{ipj}, Y_{lqk}) = \begin{cases} 0 & i \neq l \\ \frac{\sigma^2}{\sigma^2+\omega^2+\tau^2} & i = l, p \neq q \\ \frac{\sigma^2+\omega^2}{\sigma^2+\omega^2+\tau^2} & i = l, p = q \\ 1 & i = 1, p = q, k = j \end{cases}$$

Model for longitudinal growth data

$$Y_{ij} = \xi_i + \eta_i x_{ij} + \zeta_{ij} + \epsilon_{ij}$$

i: child, j: time.

Random intercepts and slopes ?

Correlated error ζ_{ij} ? e.g. AR(1)

$$\zeta_{ij} = \phi\zeta_{i(j-1)} + \nu_i$$

Multivariate normal distribution

Let $\mu \in \mathbb{R}^p$ and Σ a $p \times p$ symmetric and positive semidefinite $p \times p$ matrix.

Spectral decomposition of Σ:

$$\Sigma = U\Lambda U^{\mathsf{T}}$$

where U orthonormal matrix (columns=eigen vectors) and Λ diagonal matrix of eigen values.

Definition: a p-variate random $p \times 1$ vector Y is p-variate normal $N_p(\mu, \Sigma)$ if Y is distributed as

$$\mu + U\Lambda^{1/2}Z$$

where $Z = (Z_1, \ldots, Z_n)$ is a vector of independent standard normal random varriables.

$N_p(\mu, \Sigma)$ uniquely determined by μ and Σ.

Geometric interpretation and PCA

$\Lambda^{1/2}$: scaling. U rotation. I.e. Y scaled and rotated Z.

Let v_i ith eigen vector. Then $v_i^\mathsf{T} Y$ ith principal component with variance λ_i.

Principal components are independent. Since $\lambda_1 > \lambda_2, \ldots, \lambda_p$, $v_1^\mathsf{T} Y$ explains most of the variance in Y ($\sum_i \mathbb{V}\mathrm{ar} Y_i = \sum_i \lambda_i$).

v_i is called loading vector for ith PC.

Equivalent definitions:

Definition: a random $p \times 1$ vector Y is p-variate normal with mean μ and covariance matrix Σ if $a^\mathsf{T} Y$ is univariate normal with mean $a^\mathsf{T}\mu$ and variance $a^\mathsf{T}\Sigma a$ for any $a \in \mathbb{R}^p$.

Definition: a random $p \times 1$ vector Y is p-variate normal with mean μ and covariance matrix Σ if Y has characteristic function $L_Y(t) = \mathbb{E}\exp(it^\mathsf{T} Y) = \exp(it^\mathsf{T}\mu - \frac{1}{2}t^\mathsf{T}\Sigma t^\mathsf{T})$.

NB: since $\mathbb{V}\mathrm{ar} a^\mathsf{T} Y = a^\mathsf{T}\Sigma a \geq 0$ it follows that Σ must be positive semi-definite.

From the definition it follows easily that

$$Y \sim N_p(\mu, \Sigma) \Rightarrow AY \sim N_m(A\mu, A\Sigma A^\mathsf{T})$$

for any $m \times p$ matrix A.

Break

Show last result on previous slide.

Density of multivariate normal

Suppose Z_i are independent standard normal.

Then $Z = (Z_1, \ldots, Z_p) \sim N_p(0, I)$ with joint density

$$f_Z(z_1, \ldots, z_p) = (2\pi)^{n/2} \exp(-\|z\|^2/2)$$

Suppose further that $Y \sim N_p(\mu, \Sigma)$ where Σ *positive definite*. Then $\Sigma = LL^\mathsf{T}$ for some invertible matrix L (Cholesky or spectral decomposition, Jiang, B.5).

Thus $Y \sim \mu + LZ$ and Jacobian of transformation is $|L| = |\Sigma|^{1/2}$. By multivariate transformation theorem

$$f_Y(y_1, \ldots, y_p) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp(-\frac{1}{2}(y-\mu)^\mathsf{T}\Sigma^{-1}(y-\mu))$$

Density if Σ not positive definite

Suppose Σ has rank $r < p$. Then $\lambda_{r+1} = \cdots = \lambda_p = 0$ and Y lives on subspace $L = \text{span}\{v_1, \ldots, v_r\} \subset \mathbb{R}^p$. Possible to define density function on L.

Exercises

1. execute the script `basic.R` (on the webpage) to get acquainted with basic operations in R.
2. compute $\mathbb{V}\text{ar}\bar{Y}_{\cdot\cdot}$ for one way ANOVA.
3. fit linear models for the orthodontic growth curves with subject specific intercepts. Draw histograms of the fitted intercepts (can be extracted using `coef()`). Check residuals from the model.
4. compute covariance and correlation structure of observations from linear models with random intercepts or random slopes:

   $$Y_{ij} = \alpha + U_i + \beta x_{ij} + \epsilon_{ij} \quad Y_{ij} = \alpha + V_i x_{ij} + \epsilon_{ij}$$

   where the U_i and V_i are independent $N(0, \sigma^2)$. What can you say about the variance structure of Y_{ij} ? Consider also the model with both random intercepts and slopes.

More exercises

5. show
   $$Y \sim N_p(\mu, \Sigma) \Rightarrow AY \sim N_m(A\mu, A\Sigma A^\mathsf{T})$$
6. In Bayesian statistics the following often used as a ‘smoothing prior’:
   $$f(x_1, \ldots, x_n) \propto \exp(-\frac{1}{2}\sum_{i=2}^{n}(x_i - x_{i-1})^2)$$
   Find Q playing the role as Σ^{-1} so that the above is of the form of a multivariate Gaussian density. Is Q invertible ? Can you find a ‘square-root’ of Q ?
7. exercises 1.1, 1.2 and 1.3 at page 48 in Jiang.
8. The Laplace transform (moment generating function) of a univariate $N(\xi, \tau^2)$ random variable is $M(t) = \exp(t\xi + t^2\tau^2/2)$. Use this to compute the first four moments and central moments of a normal distribution.