Course topics

- random effects
- linear mixed models
- analysis of variance
- frequentist likelihood-based inference (MLE and REML)
- prediction
- Bayesian inference

The role of random effects

Rasmus Waagepetersen
Department of Mathematics
Aalborg University
Denmark

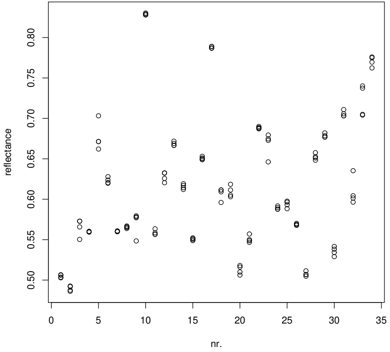
February 19, 2024

Outline for today

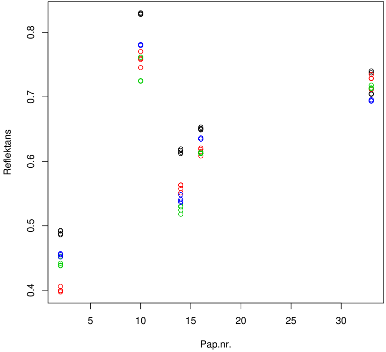
- examples of data sets.
- analysis of variance
- multivariate normal distribution
- density for multivariate normal distribution

Reflectance (colour) measurements for samples of cardboard (egg trays) (project at Department of Biotechnology, Chemistry and Environmental Engineering)

Four replications at same position on each cardboard



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

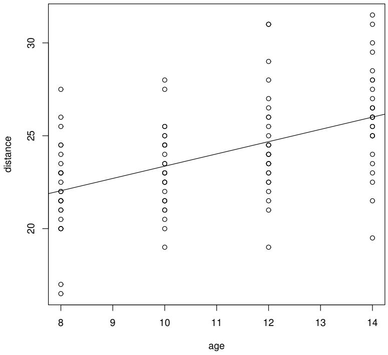


Colour variation between/within cardboards ?

Orthodontic growth curves

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

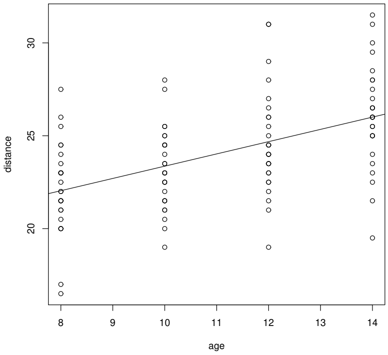
Distance versus age:



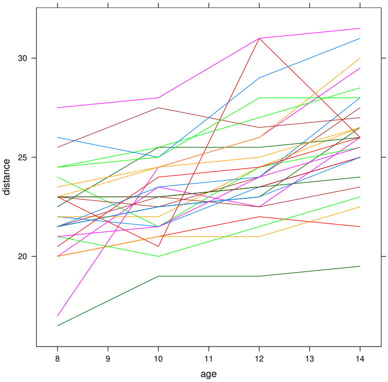
Orthodontic growth curves

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

Distance versus age:



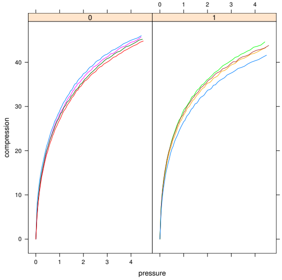
Distance versus age grouped according to child



Different intercepts for different children

Compression of mats for cows

Compression vs. pressure for two brands of mats



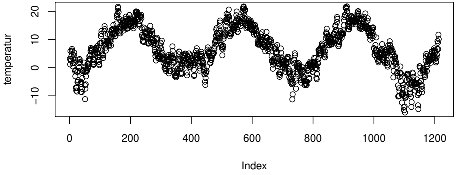
Non-linear relation

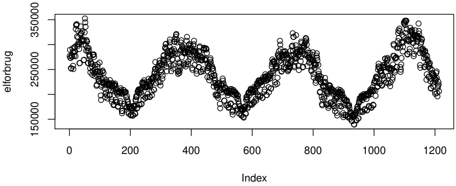
$$y = \frac{ab + cx^d}{b + x^d},$$

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

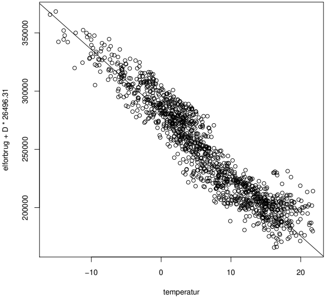
Electricity consumption and temperature in Sweden

Temperature and electricity consumption:

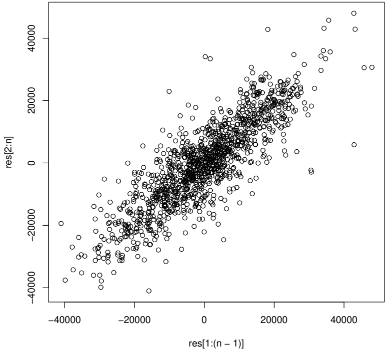




Consumption versus temperature and residuals from linear regression:



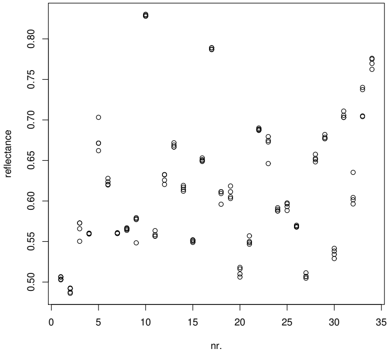
Linear dependence between consumption and temperature



Serial correlation

Model for reflectances: one-way anova

Four replications on each cardboard

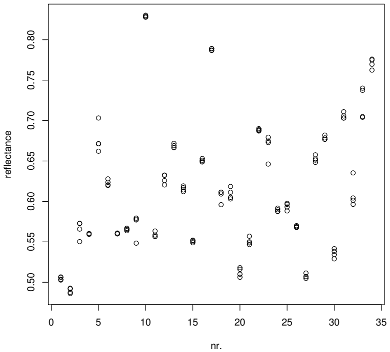


Models:

$$Y_{ij} = \xi + \epsilon_{ij}$$
$$i = 1, \ldots, k \quad j = 1, \ldots, m$$

Model for reflectances: one-way anova

Four replications on each cardboard



Models:

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

$$i = 1, \ldots, k \quad j = 1, \ldots, m$$

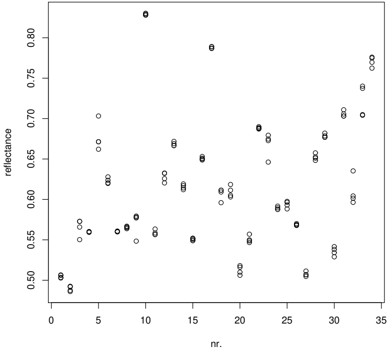
or

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

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

Model for reflectances: one-way anova

Four replications on each cardboard



Models:

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

$$i = 1, \ldots, k \quad j = 1, \ldots, m$$

or

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

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

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

where U_i are zero-mean random variables independent of each other and of ϵ_{ij}

Which is most relevant ?

One role of random effects: parsimonious and population relevant models

With fixed effects α_i: many parameters (ξ, σ^2, $\alpha_2, \dots, \alpha_{34}$). Parameters $\alpha_2, \dots, \alpha_{34}$ not interesting as they just represent intercepts for specific card boards which are individually not of interest.

With random effects: just three parameters (ξ, $\sigma^2 = \mathbb{V}\mathrm{ar}\epsilon_{ij}$ and $\tau^2 = \mathbb{V}\mathrm{ar} U_i$).

Hence parsimonious model. Variance parameters interesting for several reasons.

Second role of random effects: quantify sources of variation

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

With random effects model

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

assuming independent U_i and ϵ_{ij}, we have decomposition of variance:

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

Hence we can quantify variation between (τ^2) cardboard pieces and within (σ^2) cardboard.

Ratio $\gamma = \tau^2/\sigma^2$ is ‘signal to noise’.

Proportion of variance

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

is called *intra-class correlation*.

High proportion of between cardboard variance leads to high correlation (next slide).

Third role: modeling of covariance and correlation

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:

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

That is, observations for same cardboard are correlated !

Correct modeling of correlation is important for correct evaluation of uncertainty.

Fourth role: correct evalution of uncertainty

Suppose we wish to estimate $\xi = \mathbb{E} Y_{ij}$. Due to correlation, observations on same cardboard to some extent redundant.

Estimate is empirical average $\hat{\xi} = \bar{Y}_{\cdot\cdot}$. Evaluation of $\mathbb{V}\mathrm{ar}\bar{Y}_{\cdot\cdot}$:

Model erroneously ignoring variation between cardboards

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

$$\mathbb{V}\mathrm{ar}\epsilon_{ij} = \sigma^2_{\text{total}}[= \sigma^2 + \tau^2]$$

Naive variance expression is

$$\mathbb{V}\mathrm{ar}\bar{Y}_{\cdot\cdot} = \frac{\sigma^2_{\text{total}}}{n} \left[= \frac{\sigma^2 + \tau^2}{km} \right]$$

Correct model with random cardboard effects

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

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

Correct variance expression is

$$\mathbb{V}\mathrm{ar}\bar{Y}_{\cdot\cdot} = \frac{\tau^2}{k} + \frac{\sigma^2}{km} \qquad (1)$$

With first model, variance is underestimated !

For $\mathbb{V}\mathrm{ar}\bar{Y}_{\cdot\cdot} \to 0$ is it enough that $km \to \infty$?

Balanced one-way ANOVA (analysis of variance)

Decomposition of empirical variance/sums of squares ($i = 1, \ldots, k$, $j = 1, \ldots, m$):

$$SST = \sum_{ij} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2 = \sum_{ij} (Y_{ij} - \bar{Y}_{i\cdot})^2 + m \sum_{i} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = SSE + SSB$$

Expected sums of squares:

$$\mathbb{E} SSE = k(m-1)\sigma^2$$

$$\mathbb{E} SSB = m(k-1)\tau^2 + (k-1)\sigma^2$$

Moment-based estimates:

$$\hat{\sigma}^2 = \frac{SSE}{k(m-1)} \quad \hat{\tau}^2 = \frac{SSB/(k-1) - \hat{\sigma}^2}{m}$$

NB: $\hat{\tau}^2$ may be negative.

Slightly less nice formulae in the unbalanced case (Thm 5.3 in M&T)

Design of experiment - one-way ANOVA

Suppose Y_{ij} is outcome of analysis of jth sample in ith lab, $\tau^2 = 2$ variance between labs and $\sigma^2 = 1$ measurement variance.

Suppose we want to analyze in total 100 samples. What is then the optimal number of labs that makes $\mathbb{V}\mathrm{ar}\,\bar{Y}_{\cdot\cdot}$ minimal?

Suppose instead we have available 5000 kr., there is an initial cost of 200 kr. for each lab and subsequently 10 kr. for the analysis of each sample. What are then the optimal numbers k of labs and m of samples per lab that give the smallest $\mathbb{V}\mathrm{ar}\,\bar{Y}_{\cdot\cdot}$?

Exercise !

Two levels of random effects

For five cardboards we have 4 replications at 4 positions.

Hierarchical model (nested random effects)

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

where the U_i, U_{ip} and ϵ_{ipj} assumed to be zero-mean and independent.

Decomposition of variance:

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

Covariance structure for nested random effects model

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

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

Correlation structure for nested random effects model

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

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

Model for longitudinal growth data

$$Y_{ij} = a_i + b_i x_{ij} + \epsilon_{ij}$$

i: child, j: time.

Random intercepts and slopes ?

Correlated error ϵ_{ij} ? e.g. AR(1)

$$\epsilon_{ij} = \phi \epsilon_{i(j-1)} + \nu_{ij}$$

Summary - role of random effects

Random effects models are useful for:

- quantifying different sources of variation
- appropriate modelling of correlation ($\Rightarrow$ correct evalution of uncertainty of parameter estimates)
- sometimes individual subject specific effects not of interest - the population variation of effects more interesting
- more parsimonious models (replace many systematic effects by just one variance)

Likelihood based inference

Need probability density

Normal distribution allows to build tractable and (reasonably) flexible models

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:

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

where O 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

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

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

Geometric interpretation and PCA (principal component analysis)

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

Conversely, starting with Y: Let v_i ith eigen vector in O. Then $v_i^\mathsf{T}(Y-\mu)=\sqrt{\lambda_i}Z_i$ (projection on v_i) ith principal component with variance λ_i.

Principal components are independent (uncorrelated if Y not normal).

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

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 $\phi(t) = \mathbb{E}\exp(it^\mathsf{T} Y) = \exp(it^\mathsf{T}\mu - t^\mathsf{T}\Sigma t/2)$.

From the last 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.

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.

NB: the distribution of a random vector is uniquely determined by the characteristic function.

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

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

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).

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

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

Density if Σ not positive definite

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

Possible to define density function on L with respect to "r"-dimensional Lebesgue measure on L.

It is of the form:

$$(2\pi)^{-r/2}(\prod_{i=1}^{r} \lambda_i)^{-1/2} \exp[-\frac{1}{2}(y-\mu)^{\mathsf{T}}\Sigma^{-}(y-\mu)]$$

where

$$\Sigma^{-} = O\text{diag}(\lambda_1^{-1}, \ldots, \lambda_r^{-1}, 0, \ldots, 0)O^{\mathsf{T}} \quad \text{and} \quad \prod_{i=1}^{r} \lambda_i$$

are generalized inverse and generalized determinant of Σ.

(we will see several examples of this during the course)

2D example with Σ rank deficient

Let Z be standard normal and

$$X_1 = Z \quad X_2 = -Z$$

Then (X_1, X_2) belong to the line $y = -x$. Also X_1, X_2 satisfy the linear constraint $X_1 + X_2 = 0$.

The covariance matrix is

$$\Sigma = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

which has eigen values $2, 0$ and eigen vectors $(-\sqrt{1/2}, \sqrt{1/2})^{\mathsf{T}}$ and $(-\sqrt{1/2}, -\sqrt{1/2})^{\mathsf{T}}$.

The generalized determinant is 2.

Exercises

1. Show results regarding covariances and correlations in one-way ANOVA (slide 16).
2. compute $\mathbb{V}\mathrm{ar}\bar{Y}_{\cdot\cdot}$ for one way ANOVA (equation (1)).
3. compute expectations of SSB and SSE in one-way ANOVA.
4. 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. Also fit model with common intercept and plot residuals against subject.
5. compute covariance and correlation structure of observations from linear models with random intercepts and random slopes:

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

where $U_i \sim N(0, \tau_U^2)$ and $V_i \sim N(0, \tau_V^2)$ and $\mathbb{C}\mathrm{orr}(U_i, V_i) = \rho$ (while (U_i, V_i) and (U_i', V_i') independent when $i \neq i'$). What can you say about the variance structure of Y_{ij} ?

More exercises

6. show

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

7. show that the three definitions of multivariate normal distribution are equivalent
8. solve the design of experiment problem regarding optimal choice of numbers of laboratories and samples for the one-way ANOVA

One more exercise

9. In Bayesian statistics the following pairwise-difference density is often used as a ‘smoothing prior’ (promotes similar values of y_i and y_{i-1}):

$$f(y_1, \ldots, y_n) \propto \exp(-\frac{1}{2}\sum_{i=2}^{n}(y_i - y_{i-1})^2)$$

   9.1 Find a matrix D such that $\sum_{i=2}^{n}(y_i - y_{i-1})^2 = y^\top D^\top D y$.
   9.2 Use the previous question of find Q playing the role as Σ^{-1} so that the above is of the form of a multivariate Gaussian density. Is Q invertible ?
   9.3 Find the eigen space for the eigen value 0.
   9.4 The pairwise-difference density can be viewed as a density on an $n-1$-dimensional subspace L. Characterize L.