Design and analysis of experiments
Lecture 6

Jakob G. Rasmussen
Department of Mathematics
Aalborg University
Denmark

Greaco-latin design

- The latin design showed us one way of designing an experiment with one main factor and two nuisance factors - in greaco-latin designs we insert one more nuisance factor.
- The levels of this factor is combined with the other factors, such that each level is combined with each level of the other factors exactly once.
- Example:

$A\alpha$	$B\beta$	$C\gamma$	$D\delta$
$B\gamma$	$A\delta$	$D\alpha$	$C\beta$
$C\delta$	$D\gamma$	$A\beta$	$B\alpha$
$D\beta$	$C\alpha$	$B\delta$	$A\gamma$

- Such designs can be made for the number of levels in all four factors $p > 2$ and $p \neq 6$.
- We need to make p^2 number of experiments, rather than p^4 for making all combinations once.

R

- R demo, part 1
- Exercise 1

Balanced incomplete block design

- Consider a problem with one main factor and one blocking factor - what do we do if there's no time/resources to do all combinations of levels in the two factors?
- BIBD: an incomplete block design, where any pair of treatments appear together an equal number of times.
- Numbers:
  - $a =$ number of treatments
  - $r =$ number of replicates
  - $b =$ number of blocks
  - $k =$ size of block
  - $\lambda =$ number occurrences of a pair in the same block
- A simple example ($a = 4, r = 3, b = 4, k = 3, \lambda = 2$):

A	A	A	B
B	B	C	C
C	D	D	D

Requirements

- All the numbers a, r, b, k, λ must be integers.
- The number of observations is $N = ar = bk$. Often a, b and k are given, hence $r = \frac{bk}{a}$ must be an integer.
- By considering the blocks in which a specific treatment occur, the remaining treatments in these blocks can be counted in two ways, giving $r(k-1) = \lambda(a-1)$. For example, A is repeated $r = 3$ times with $k - 1 = 2$ other treatments, but it also appears $\lambda = 2$ times with $a - 1 = 3$ other treatments.

A	A	A	B
B	B	C	C
C	D	D	D

  Hence $\lambda = r\frac{k-1}{a-1}$ must be an integer.
- These two conditions are necessary, but not sufficient. The simplest example with no BIBD is $a = 15$, $b = 21$, $k = 5$, $r = 7$, $\lambda = 2$.

A Classic Design

- For $a = 7$ treatments, $b = 7$ blocks of size $k = 3$, the conditions are fulfilled with $r = \frac{bk}{a} = 3$, $\lambda = r\frac{k-1}{a-1} = 1$.
- Let the treatments be A, B, C, D, E, F, G.
- Steps:
  1. We order the 3 blocks with A first and assume that B-C, D-E and F-G are together in these blocks:
  2. B and C must occur twice more and not together.
  3. D-F must occur together. We can assume this happens in a B block. The E-G must be together in the other B block. Then there is only one choice for the two C blocks.

A	A	A	B	B	C	C
B	D	F	D	E	D	E
C	E	G	F	G	G	F

R

- R demo, part 2
- Exercise

Types of data

- Qualitative, categorical (nominal):
  - Categories of data
  - Fx eye colors (blue, green, grey, brown)
- Qualitative, ordinal
  - Ordered data
  - Fx cloth sizes (S, M, L, XL)
- Quantitative, continuous (interval)
  - Data measured in real numbers
  - Fx heights

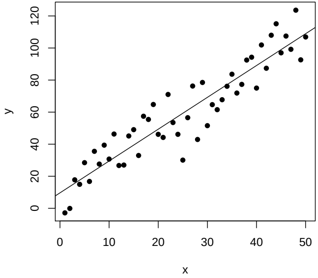
Explanatory variables	Response variable	
	categorical	continuous
categorical	contingency table	analysis of variance
continuous	e.g. logistic regression	regression analysis
both	?	analysis of covariance

Simple regression

- We start with simple regression, where we model a continuous response variable y as a function of a continuous explanatory variable x.
- Data: we have observed $((x_1, y_1)), \ldots, (x_n, y_n))$.
- Model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- Assumptions:
  - Normal distribution: $\epsilon_i \sim N(0, \sigma^2)$
  - Independence: ϵ_i are independent from each other



Simple regression - matrix notation

- Matrix notation: we can write this as $y = X\beta + \epsilon$, where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- X is called the design matrix - formulating other design matrices will give us other models.
- The matrix notation makes life easier, since we can cope with all regression models at the same time.

Multiple regression

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \epsilon_i$$

- Design matrix and parameter vector:

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ & & \vdots & \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Polynomial regression

- Model:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_k x_i^k + \epsilon_i$$

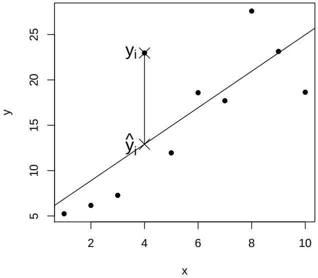
- Design matrix and parameter vector:

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ & & & \vdots & \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Fitting a straight line (or other function)

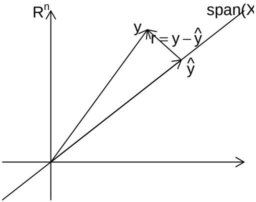
- "True" model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ (or generally: $y = X\beta + \epsilon$)
- Fitted model: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ (or generally: $\hat{y} = X\hat{\beta}$)
- Ordinary least squares (OLS):
  - We want to find values $\hat{\beta}_0$ and $\hat{\beta}_1$ (or generally: $\hat{\beta}$) such that the fit is as close to the data as possible
  - I.e. we minimize the sum of squares of the residuals $r_i = y_i - \hat{y}_i$ (i.e. the length of the residual vector $r = y - \hat{y}$):

$$SS_E = \sum r_i^2 = \sum (y_i - \hat{y}_i)^2 = |r|^2 = |y - \hat{y}|^2$$



Estimation

- We need to find $\hat{\beta}$, the estimate of β, by minimising the length of the residual vector $r = y - \hat{y}$.
- Two methods:
  - Differentiation - used in the book
  - Geometry - used here
- Consider y and $\hat{y}$ vectors in n-dimensional space $\mathbb{R}^n$ - $r = y - \hat{y}$ is smallest when $r \perp \text{cols}(X)$



Estimating β

- Since r is orthogonal to any x_j (the jth column of X) it is also orthogonal to any linear combination, i.e. $r \cdot (Xv) = 0$ for any vector $v \in \mathbb{R}^{k+1}$:

$$
\begin{aligned}
&0 = (Xv)^\top (y - X\hat{\beta}) = v^\top (X^\top y - X^\top X\hat{\beta}) \Rightarrow \\
&X^\top y - X^\top X\hat{\beta} = 0 \Rightarrow \\
&\hat{\beta} = (X^\top X)^{-1} X^\top y \Rightarrow \\
&\hat{y} = X\hat{\beta} = X(X^\top X)^{-1} X^\top y = Hy
\end{aligned}
$$

- H is called a projection matrix or hat-matrix.
- Parameter estimates:

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

- Fitted values:

$$\hat{y} = X(X^\top X)^{-1} X^\top y = Hy$$

R

- R demo, part 3
- Exercise 3