2.7 General linear models

General linear models

The term **general linear models** is used for Gaussian models with any combination of categorical and continuous explanatory variables.

The factors can be

- **crossed:** there are observations for each combination of levels of the factors (see two factors ANOVA)
- **nested:** the combinations of factors are different

Example — nested factors

Two-factor nested design, comparison of **two drugs** A_1 tested in hospital B_1, B_2 and B_3 and A_2 tested in hospital B_4 and B_5.

	Drug A_1			Drug A_2	
Hospitals	B_1	B_2	B_3	B_4	B_5
Responses	Y_{111}	Y_{121}	Y_{131}	Y_{241}	Y_{251}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	Y_{11n_1}	Y_{12n_2}	Y_{13n_3}	Y_{24n_4}	Y_{25n_5}

We want to *compare the effects of the two drugs* and *possible differences among hospitals* using the same drug.

The saturated model is

$$\mathbb{E}(\mathrm{Y}_{jkl}) = \mu + \alpha_1 + \alpha_2 + (\alpha\beta)_{11} + (\alpha\beta)_{12} + (\alpha\beta)_{13} + (\alpha\beta)_{24} + (\alpha\beta)_{25}$$

under constraints

- $\alpha_1 = 0$
- $(\alpha\beta)_{11} = 0$
- $(\alpha\beta)_{24} = 0$

Hospitals 1, 2 and 3 can be only compared within drug A_1 and hospitals 4 and 5 can be only compared within drug A_2.

This model is not different from other Gaussian models

- Response variable is assumed to be normally distributed
- Response and explanatory variables are assumed to be linearly related
- The variance σ^2 is constant
- The responses are assumed to be independent

These assumption must be checked by looking at the **residuals**.

If the assumption of normality is not plausible, a **transformation** can be used.

A very used transformation is the **Box-Cox transformation**:

$$y^* = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases} \tag{2.7.1}$$

- if $\lambda = 1$, y is unchanged (except for a location shift which does not influence inference)
- if $\lambda = \frac{1}{2}$, the transformation is the square root
- if $\lambda = -1$, the transformation is the reciprocal
- if $\lambda = 0$, the transformation is the logarithm

The value of λ which produces the **"most normal" distribution** can be estimated by the method of **maximum likelihood**.