Notes and examples: The linear model

math612.0 A1: From numbers through algebra to calculus and linear algebra

Gunnar Stefansson (editor) with contributions from very many students

March 7, 2022

Simple linear regression in R

To test the effect of one variable on another, simple linear regression may be applied. The fitted model may be expressed as $y = \alpha + \hat{\beta}x$, where α is a constant, $\hat{\beta}$ is the estimated coefficient, and x is the explanatory variable.

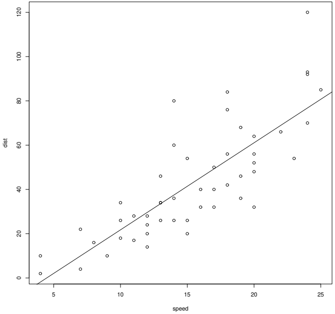

Figure: Example taken from R of a fitted model using linear regression.

Multiple linear regression

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Formally, the model for multiple linear regression, given n observations, is

$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_p x_{i,p} + e_i$ for $i = 1, 2, \ldots, n$

As always, we view the data, y_i as observations of random variables, so another way to describe the same model is

$Y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_p x_{i,p} + \epsilon_i$ for $i = 1, 2, \ldots, n$,

and we note that the x-values are just numbers and are usually assumed to be without any measurement error.

The one-way model

The one-way ANOVA model is of the form:

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

or

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

Random effects in the one-way layout

The simplest random effects model is the one-way layout, commonly written in the form

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

where $j = 1, \ldots, J$ and $i = 1, \ldots, I$.
Normally one also assumes $\epsilon_{ij} \sim n(0, \sigma_A^2)$, $\alpha_i \sim n(0, \sigma_A^2)$, and that all these random variables are independent.
Note that we have stopped making a distinction in notation between random variables and measurements (the y-values are just random variables when distributions occur).

Linear mixed effects models (lmm)

The simplest mixed effects model is

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

where $\mu, \alpha_1, \alpha_2, \ldots, \alpha_i$ are unknown constants,
$\beta_j \sim n(0, \sigma_\beta^2)$
$\epsilon_{ij} \sim n(0, \sigma^2)$
(β_j and ϵ_{ij} independent).

Maximum likelihood estimation in lmm

The likelihood function for the unknown parameters $L(\boldsymbol{\beta}, \sigma_A^2, \sigma^2)$ is

$$\frac{1}{(2\pi)^{n/2}\left|\Sigma_y\right|^{n/2}} e^{-1/2(\mathrm{y}-X\boldsymbol{\beta})'\Sigma_y^{-1}(\mathrm{y}-X\boldsymbol{\beta})}$$

where $\Sigma_y = \sigma_A^2 ZZ' + \sigma^2 I$.

Maximising L over $\boldsymbol{\beta}, \sigma_A^2, \sigma^2$ gives the variance components and the fixed effects. May also need $\hat{\mathrm{u}}$, this is normally done using BLUP.