Linear models and linear mixed models

STAT3306/7306

Dr. Loïc Yengo
(slides Dr. Florian Rohart)

The Institute for Molecular Bioscience

19/08/19

Outline

1. Linear Models

2. Linear Mixed Models

Outline

Simple Linear Regression

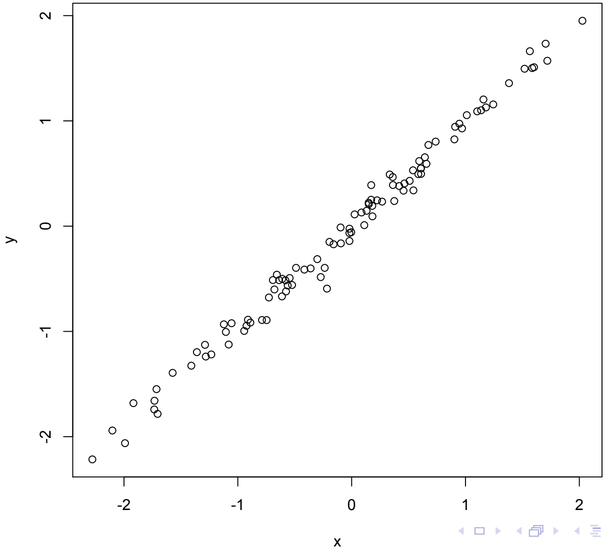
Considering observing n samples from a simple linear model with only a single unknown slope parameter $\beta \in \mathbb{R}$,

$$y_i = x_i\beta + e_i, \qquad i = 1, \ldots, n$$

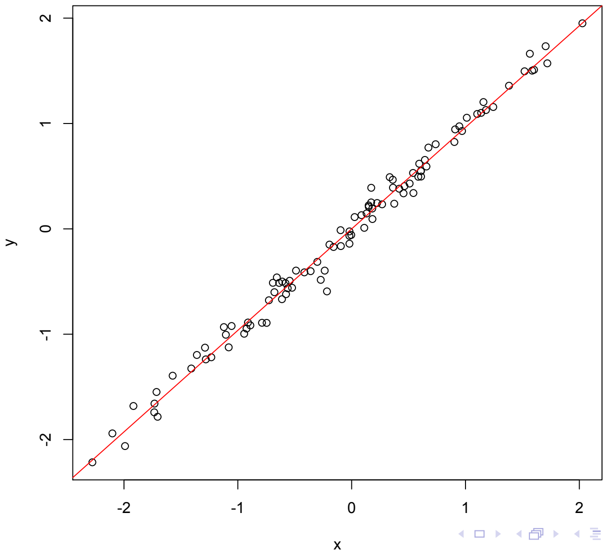
This is probably the simplest linear model.

- x_i are fixed and known quantities.
- y_i are observed and known quantities.
- we want to estimate β
- e_i are some noise, usually assumed Gaussian

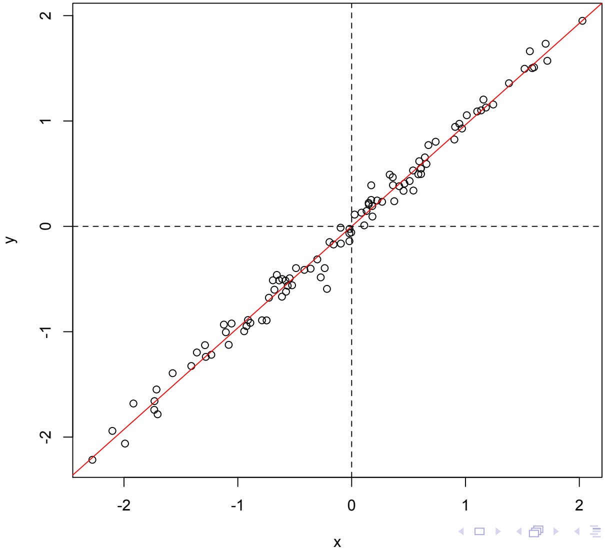
Example



Example



Example



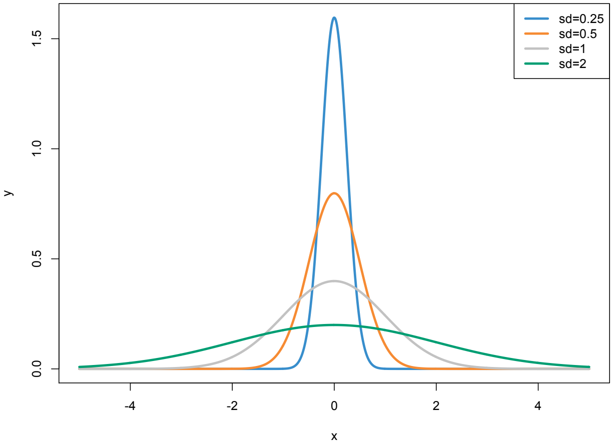
The error terms e_i are assumed to be independent and identically distributed (i.i.d) random variables with a normal density function: $e_i \sim \mathcal{N}(0, \sigma^2)$

$$\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

for some unknown variance $\sigma^2 > 0$.

$\boldsymbol{I}$ identity matrix of size $n \times n$,
$\mathbf{0}$ is a n-vector of 0s.
$\mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}) = \text{MVN}(\mathbf{0}, \sigma^2 \boldsymbol{I})$

Normal distribution



Density of a normally distributed random variable

The density function of a normally distributed random variable with mean μ and variance σ^2 is given by:

$$f(z;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\}$$

for $z, \mu \in \mathbb{R}$, and $\sigma > 0$.

The first term is a normalisation factor so that the density sums to 1.
The important part is:

$$f(z;\mu,\sigma^2) \;\propto \exp\left\{-0.5\left(\frac{z-\mu}{\sigma}\right)^2\right\}$$

$\propto$: proportional to

A multiple regression is a typical linear model,

$$\begin{aligned} y_i &= x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ip}\beta_p + e_i \\ &= \boldsymbol{X}_i^\top \boldsymbol{\beta} + e_i \end{aligned}$$

where $\boldsymbol{X}_i$ is a $p \times 1$ vector of measurements and $\boldsymbol{\beta}$ is a vector of p parameters to estimate.

Matrix form of a linear regression

The model can also be written in a matrix form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

where

- $\boldsymbol{y}$ vector of n observed dependent values
- $\boldsymbol{X}$ observations of the variables in the assumed linear model, $n \times p$ matrix
- $\boldsymbol{\beta}$ vector of p unknown parameters to estimate
- $\boldsymbol{e}$ vector of residuals (deviation from the model fit), $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. Usually, assumed independent and identically distributed (i.i.d), $\boldsymbol{e} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$

Example

Suppose we have 3 variables in a multiple regression, with four (y, x) vectors of observations.

$$y_i = \mu + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + e_i$$

In a matrix form, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{pmatrix}, \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$$

Details of both the experimental design and the observed values of the predictor variables all reside solely in X.

Ordinary Least Squares (OLS) - solution to $y = X\beta + e$

OLS

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{b}\in\mathbb{R}^p}\{||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}||_2^2\} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

- The variance-covariance estimate for $\widehat{\boldsymbol{\beta}}$ is

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\sigma_e^2$$

  The ij-th element gives the covariance between the estimated of β_i and β_j.

$\text{OLS}(\beta) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$

proof

$$\begin{aligned}
||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}||_2^2 &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\
&= (\boldsymbol{y}^\top - \boldsymbol{b}^\top \boldsymbol{X}^\top)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\
&= \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{b} - \boldsymbol{b}^\top \boldsymbol{X}^\top \boldsymbol{y} + \boldsymbol{b}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{b} \\
&= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{b} + \boldsymbol{b}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{b}
\end{aligned}$$

$$\frac{\partial \left(||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 \right)}{\partial \boldsymbol{\beta}} = 0 \iff -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = 0 \iff \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y}$$

$\boldsymbol{X}^\top \boldsymbol{X}$ needs to be invertible

R example

```
## Fitting a linear model using lm
> lm(longley$Employed ~ ., data = longley)
Coefficients:
 (Intercept)   GNP.deflator          GNP     Unemployed   Armed.Forces
  -3.482e+03      1.506e-02   -3.582e-02     -2.020e-02     -1.033e-02
  Population           Year
  -5.110e-02      1.829e+00

## Estimating beta with the formula
> X = cbind("intercept" = 1, longley[,1:6])# add the intercept
> Y = longley[,"Employed"]

> beta = solve( t(X) %*% X ) %*%  t(X) %*% Y
> beta
                       [,1]
intercept     -3.482259e+03
GNP.deflator   1.506187e-02
GNP           -3.581918e-02
Unemployed    -2.020230e-02
Armed.Forces  -1.033227e-02
Population    -5.110411e-02
Year           1.829151e+00
```

Properties: OLS = BLUE

In the case of a linear model where the residuals are homoscedastic (equal variance), uncorrelated and have expectation zero, the **OLS** estimator is also the **Best Linear Unbiased Estimator (BLUE)**, i.e the OLS estimator has the lowest variance among all the unbiased estimators.

Unbiased estimator

An estimator $\widehat{\theta}$ of θ is unbiased if and only if $E(\widehat{\theta}) = \theta$, where E denotes the expectation

Likelihood (Sample of Normal Variables)

$\{x_1, \dots, x_n\}$ is a realisation of $\{X_1, \dots, X_n\}$ where $X_i \sim \mathcal{N}(\mu, \sigma^2)$, i.i.d.

The likelihood of $\{x_1, \dots, x_n\}$ is

$$
\begin{aligned}
L(\mu, \sigma^2) = L(\mu, \sigma^2; x_1, \dots, x_n) &= Pr\left((X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)\right) \\
{}^{a} &= \prod_{i=1}^{n} Pr(X_i = x_i) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}
\end{aligned}
$$

[a] because independent and identically distributed

The log-likelihood is

$$\ell(\mu, \sigma^2) = \log(L(\mu, \sigma^2)) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Likelihood (Linear Regression Model)

Consider the regression model:

$$y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + e_i$$

We have $y_i \sim \mathcal{N}(\boldsymbol{X}_i^\top \boldsymbol{\beta}, \sigma^2)$

Likelihood (Linear Regression Model)

Consider the regression model:

$$y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + e_i$$

We have $y_i \sim \mathcal{N}(\boldsymbol{X}_i^\top \boldsymbol{\beta}, \sigma^2)$

The likelihood of $\{x_1, \dots, x_n\}$ is

$$\begin{aligned} L(\mu, \sigma^2) = L(\mu, \sigma^2; x_1, \dots, x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \end{aligned}$$

Likelihood (Linear Regression Model)

Consider the regression model:

$$y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + e_i$$

We have $y_i \sim \mathcal{N}(\boldsymbol{X}_i^\top \boldsymbol{\beta}, \sigma^2)$

The likelihood of $\{y_1, \ldots, y_n\}$ is

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) = L(\boldsymbol{\beta}, \sigma^2; y_1, \ldots, y_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{\sigma} \right)^2 \right\} \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta} \right)^2 \right\} \end{aligned}$$

Likelihood (Linear Regression Model)

Consider the regression model:

$$y_i = \boldsymbol{X}_i^\top \boldsymbol{\beta} + e_i$$

We have $y_i \sim \mathcal{N}(\boldsymbol{X}_i^\top \boldsymbol{\beta}, \sigma^2)$

The likelihood of $\{y_1, \dots, y_n\}$ is

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) = L(\boldsymbol{\beta}, \sigma^2; y_1, \dots, y_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}}{\sigma}\right)^2\right\} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2\right\} \end{aligned}$$

The log-likelihood is

$$\ell(\boldsymbol{\beta}, \sigma^2) = \log(L(\boldsymbol{\beta}, \sigma^2)) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2$$

Definition (Maximum Likelihood estimator)

A Maximum Likelihood Estimator (MLE) $\hat{\theta}$ of $\theta \in \Theta$ is a solution to the maximisation problem:

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \; L(\theta) \quad \text{or equivalently} \quad \widehat{\theta} = \arg\max_{\theta \in \Theta} \; \ell(\theta)$$

To obtain MLE, we solve the partial derivatives $\frac{\partial \ell(\theta)}{\partial \theta_j}$, $\quad j = 1, 2, \ldots,$
For the sample $\{x_1, \ldots, x_n\}$,

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Definition (Maximum Likelihood estimator)

A Maximum Likelihood Estimator (MLE) $\hat{\theta}$ of $\theta \in \Theta$ is a solution to the maximisation problem:

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} L(\theta) \quad \text{or equivalently} \quad \widehat{\theta} = \arg\max_{\theta \in \Theta} \ell(\theta)$$

To obtain MLE, we solve the partial derivatives $\frac{\partial \ell(\theta)}{\partial \theta_j}, \quad j = 1, 2, \ldots,$
For the sample $\{x_1, \ldots, x_n\}$,

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \quad \textit{and} \quad \frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

Definition (Maximum Likelihood estimator)

A Maximum Likelihood Estimator (MLE) $\hat{\theta}$ of $\theta \in \Theta$ is a solution to the maximisation problem:

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} \; L(\theta) \quad \text{or equivalently} \quad \widehat{\theta} = \arg\max_{\theta \in \Theta} \; \ell(\theta)$$

To obtain MLE, we solve the partial derivatives $\frac{\partial \ell(\theta)}{\partial \theta_j}$, $\quad j = 1, 2, \ldots,$
For the sample $\{x_1, \ldots, x_n\}$,

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \;\; \textit{and} \quad \frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\begin{cases} \dfrac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = 0 \\ \dfrac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = 0 \end{cases} \Longleftrightarrow \begin{cases} \widehat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} x_i \\ \widehat{\sigma^2} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$

$\bar{x}$ is the mean of $(x_1, \ldots, x_n)$

Exercise

Calculate MLE of β and σ^2 for a linear regression model.
Verify that MLE is also OLS.

Exercise

Calculate MLE of β and σ^2 for a linear regression model.
Verify that MLE is also OLS.

What now?
We want to be able to test whether a model is better than another based on the log-likelihood (goodness of fit).

For instance, I estimated a model with 10 parameters, but maybe some of them are irrelevant. I want to assess whether a model with only 9 parameters is still a good fit to the data.

Likelihood ratio test

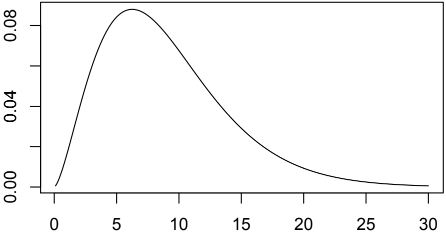
Reminder (?) on hypothesis testing - Example

Example: We want to test whether the population mean μ is 0.

$$H_0 : \mu = 0 \text{ against } H_1 : \mu \neq 0.$$

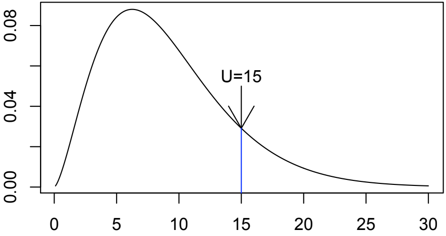
We have access to 10 observations $x_1, \ldots, x_{10}$ and we use the sample mean $\bar{x}$ to test the null hypothesis H_0

Reminder (?) on hypothesis testing - General Framework



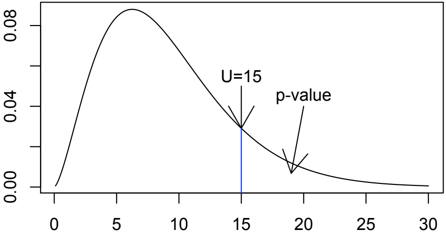
H_0 against H_1, tested with the statistic U, which is following a known distribution

Reminder (?) on hypothesis testing - General Framework



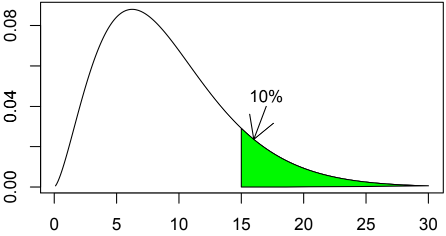
H_0 against H_1, tested with the statistic U, which is following a known distribution

Reminder (?) on hypothesis testing - General Framework



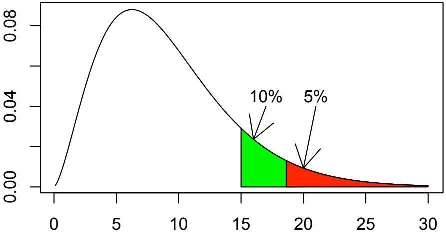
H_0 against H_1, tested with the statistic U, which is following a known distribution

Reminder (?) on hypothesis testing - General Framework



H_0 against H_1, tested with the statistic U, which is following a known distribution

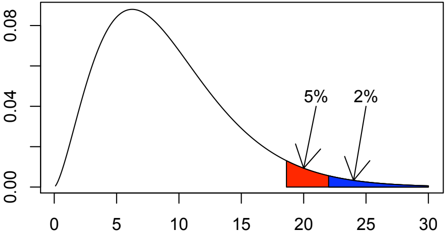
Reminder (?) on hypothesis testing - General Framework



H_0 against H_1, tested with the statistic U, which is following a known distribution

if *pvalue* $> \alpha$, then H_0 is accepted

Reminder (?) on hypothesis testing - General Framework



H_0 against H_1, tested with the statistic U, which is following a known distribution

if *pvalue* $> \alpha$, then H_0 is accepted
if *pvalue* $< \alpha$, then H_0 is rejected

The lower the p-value is, the stronger we reject H_0.
Critical region $U > \alpha$; $\alpha =$ type I error.

Likelihood Ratio Test (LRT)

For some subset $\Theta_0 \subset \Theta$,

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \notin \Theta_0$$

The MLE of θ solves: $\widehat{\theta} = \arg\max_{\theta \in \Theta} L(\theta)$

The MLE of θ under H_0 solves: $\widehat{\theta}_0 = \arg\max_{\theta \in \Theta_0} L(\theta)$

Definition (Likelihood Ratio)

The likelihood ratio for testing H_0 vs H_1 is defined as

$$\Lambda = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}$$

- $0 < \Lambda \leq 1$
- Higher values of Λ are evidence in favour of H_0
- Lower values of Λ are evidence against H_0
- Critical region: $\{x \mid \Lambda \leq \lambda_0\}$ where $0 \leq \lambda_0 \leq 1$.

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

Example

Full model $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$
Sub-model $y_i = x_{i1}\beta_1 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$,

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

Example

Full model $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$
Sub-model $y_i = x_{i1}\beta_1 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$,

$$\Theta_0 =$$

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

Example

Full model $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$
Sub-model $y_i = x_{i1}\beta_1 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$,

$$\Theta_0 = \left\{\sigma^2, \boldsymbol{\beta} : \beta_2 = 0\right\}, \qquad H_0 : \boldsymbol{\beta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\beta} \notin \Theta_0$$

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

Example

Full model $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$

Sub-model $y_i = x_{i1}\beta_1 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$,

$$\Theta_0 = \left\{\sigma^2, \boldsymbol{\beta} : \beta_2 = 0\right\}, \qquad H_0 : \boldsymbol{\beta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\beta} \notin \Theta_0$$

$$\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}^2) - \frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}\left(y_i - x_{i1}\widehat{\beta}_1 - x_{i2}\widehat{\beta}_2\right)^2$$

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

Example

Full model $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$

Sub-model $y_i = x_{i1}\beta_1 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$,

$$\Theta_0 = \left\{\sigma^2, \boldsymbol{\beta} : \beta_2 = 0\right\}, \qquad H_0 : \boldsymbol{\beta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\beta} \notin \Theta_0$$

$$\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}^2) - \frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}\left(y_i - x_{i1}\widehat{\beta}_1 - x_{i2}\widehat{\beta}_2\right)^2$$

$$\ell(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}_0^2) - \frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^{n}\left(y_i - x_{i1}\widehat{\beta}_{01}\right)^2$$

log-LRT

Definition (log-Likelihood Ratio)

The log-likelihood ratio for testing H_0 vs H_1 is defined as

$$-2 \times log(\Lambda) = -2 \times log\left[\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right] = 2\left[\ell(\widehat{\theta}) - \ell(\widehat{\theta}_0)\right]$$

Example

Full model $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2)$
Sub-model $y_i = x_{i1}\beta_1 + e_i$, which is solved by $(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)$,

$$\Theta_0 = \left\{\sigma^2, \boldsymbol{\beta} : \beta_2 = 0\right\}, \qquad H_0 : \boldsymbol{\beta} \in \Theta_0 \text{ against } H_1 : \boldsymbol{\beta} \notin \Theta_0$$

$$\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}^2) - \frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}\left(y_i - x_{i1}\widehat{\beta}_1 - x_{i2}\widehat{\beta}_2\right)^2$$

$$\ell(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}_0^2) - \frac{1}{2\widehat{\sigma}_0^2}\sum_{i=1}^{n}\left(y_i - x_{i1}\widehat{\beta}_{01}\right)^2$$

$$-2 \times log(\Lambda) = 2\left[\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) - \ell(\widehat{\boldsymbol{\beta}}_0, \widehat{\sigma}_0^2)\right]$$

Generalised Least Squares

GLS

Residuals are heteroscedastic and/or dependent, $\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{V})$. The linear model becomes

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}, \mathrm{Var}(\boldsymbol{e}) = \sigma_e^2 \boldsymbol{R}$$

- OLS: special case of GLS, where $\mathrm{Var}(\boldsymbol{e}) = \sigma_e^2 \boldsymbol{I}$.
- The GLS estimate is $\text{GLS}(\beta) = (\boldsymbol{X}^\top \boldsymbol{R}^{-1} \boldsymbol{X})^{-1} \boldsymbol{R}^{-1} \boldsymbol{X}^\top \boldsymbol{y}$
- The variance-covariance of the estimated model parameters is given by

$$\mathrm{Var}(\widehat{\beta}) = (\boldsymbol{X}^\top \boldsymbol{R}^{-1} \boldsymbol{X})^{-1} \sigma_e^2$$

Exercise: how do you get the GLS estimate?

The trick is to pre-multiply $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}$ by $\boldsymbol{R}^{-1/2}$:
$\boldsymbol{R}^{-1/2} y = \boldsymbol{R}^{-1/2} \boldsymbol{X}\beta + \boldsymbol{R}^{-1/2} \boldsymbol{e}$
$z = \boldsymbol{Z}\beta + \boldsymbol{f}$ with $\boldsymbol{f} \sim \mathcal{N}(0, \boldsymbol{I}\sigma_e^2)$, and then apply an OLS.

Summary OLS vs GLS

	OLS	GLS
Assumed distribution of residuals	$\boldsymbol{e} \sim (\boldsymbol{0}, \sigma_e^2 \boldsymbol{I})$	$\boldsymbol{e} \sim (\boldsymbol{0}, \boldsymbol{V})$
Least-squares estimator of $\boldsymbol{\beta}$	$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$	$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y}$
$\mathrm{Var}(\widehat{\boldsymbol{\beta}})$	$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \sigma_e^2$	$(\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1}$
Predicted values, $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$	$\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$	$\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y}$
$\mathrm{Var}(\widehat{\boldsymbol{y}})$	$\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \sigma_e^2$	$\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top$

Problems with High-Dimensional Linear Model

high-dimension, $n < p$

Same model as before

$$y = X\beta + e$$

where X $n \times p$ matrix, and $p > n$, or $p >> n$

Example

```
## small high-dimensional dataset, 3 * 6
> X
        ENSG00000000003 ENSG00000000005 ENSG00000000419 ENSG00000000457 ENSG00000000460 ENSG00000000938
sample1       0.8656802       0.2445878       0.9027015       0.3773931       0.3773931     0.008697272
sample2       0.8561306       0.1807828       0.8853417       0.4156058       0.4156058     0.098042092
sample3       0.8870595       0.1840356       0.8915388       0.4016337       0.4016337     0.082282120
```

$X^\top X$ is not invertible.

```
> solve(t(X)%*% X)
 Error in solve.default(t(X) %*% X) :
   Lapack routine dgesv: system is exactly singular: U[5,5] = 0
```

We can estimate at most n parameters. Here we have more parameters ($p = 6$) to estimate than observations ($n = 3$). We have lost identifiability: no unique β (we can find several β solution to $X^\top X\beta = X^\top y$)

Solutions to High-Dimensional Linear Model

- add a constraint to the optimisation problem - like Lasso (ℓ^1), Ridge (ℓ^2), Elastic net (mixed ℓ^1 and ℓ^2).

Lasso (Tibshirani 1996)

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{b} \in \mathbb{R}^p, ||\boldsymbol{b}||_1 < \lambda}{\arg\min} \{||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}||_2^2\},$$

where $\lambda > 0$ is the penalty or regularisation parameter, and control the amount of shrinkage (and the number of non zero coefficients in $\boldsymbol{b}$)

- one parameter at a time

Marginal regression, $\boldsymbol{y} = \boldsymbol{X}_j \beta_j + \boldsymbol{e}$

For all $j \in \{1, 2, \dots, p\}$

$$\widehat{\beta}_j = \underset{b \in \mathbb{R}}{\arg\min} \{||\boldsymbol{y} - \boldsymbol{X}_j b||_2^2\} = (\boldsymbol{X}_j^\top \boldsymbol{X}_j)^{-1} \boldsymbol{X}_j^\top \boldsymbol{y}$$

Outline

Fixed effects vs random effects

Factor effects are either fixed or random.

- Fixed: The levels in the study represent all levels of interest
- Random: The levels in the study represent only a sample of the levels of interest. Levels are considered to be drawn from an infinite population of levels.

What do you think?

Gender, year to year variation in rainfall at a location, school

We generally speak of estimating fixed factors (BLUE) and predicting random effects (BLUP - Best linear unbiased Predictor).

Mixed models (MM)

Mixed models (MM) contain both fixed and random factors

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

where

- $\boldsymbol{y}$ vector of observed dependent values, with mean $E(\boldsymbol{y}) = \boldsymbol{X}\beta$
- β vector of unknown parameters to estimate (fixed effects)
- $\boldsymbol{u}$ vector of unknown random effects, with mean $E(\boldsymbol{u}) = 0$ and variance-covariance $\mathrm{Var}(\boldsymbol{u}) = \boldsymbol{G}$
- $\boldsymbol{e}$ vector of residuals, with mean $E(\boldsymbol{e}) = 0$ and variance-covariance $\mathrm{Var}(\boldsymbol{e}) = \boldsymbol{R}$
- $\boldsymbol{X}$ and $\boldsymbol{Z}$ are design matrices

Example

Suppose we have 3 variables in a multiple regression, with four (y, x) vectors of observations.

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \beta_1 \\ \beta_2 \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \\ 1 & x_{51} & x_{52} \end{pmatrix}, \boldsymbol{Z} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \boldsymbol{u} \sim \mathcal{N}_3(\boldsymbol{0}, \sigma_u^2 \boldsymbol{I}_3)$$

$$\boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}, \boldsymbol{e} \sim \mathcal{N}_5(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I}_3)$$

We estimate $\boldsymbol{\beta}$ and we predict $\boldsymbol{u}$ (we do not want the values u_1, u_2, u_3 but we want σ_u^2).

k effects, two possibilities:

- treated as fixed effects: we lose k degrees of freedom.
- treated as random effects from $\mathcal{N}(0, \sigma^2)$: only one degree of freedom is lost (estimating the variance) and we can then predict the values of the k realisations.

MME

Mixed Model Equations (MME)

$$\begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^\top \boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^\top \boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^\top \boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}$$

The solutions to the MME are the best linear unbiased estimates ($\widehat{\boldsymbol{\beta}}$, BLUE) and predictors ($\widehat{\boldsymbol{u}}$, BLUP) for $\boldsymbol{\beta}$ and $\boldsymbol{u}$.

Usually solved by an EM algorithm (Expectation-Maximisation).

Note that $\widehat{\boldsymbol{\beta}}$ is the GLS estimate from the marginal model: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ with $\boldsymbol{e} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{V})$ and $\boldsymbol{V} = \boldsymbol{ZGZ}^\top + \boldsymbol{R}$

Problems with High-Dimensional Linear Mixed Model

high-dimension, $n < p$

Same model as before

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

where $\boldsymbol{X}$ $n \times p$ matrix, and $p > n$, or $p \gg n$

Solution to the problem,

- add constraints to the optimisation problem
- one parameter at a time

Summary

- Least Squares

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{b}\in\mathbb{R}^p}{\arg\min}\{||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}||_2^2\}$$

	OLS	**GLS**
Assumed distribution of residuals	$\boldsymbol{e} \sim (\boldsymbol{0}, \sigma_e^2 \boldsymbol{I})$	$\boldsymbol{e} \sim (\boldsymbol{0}, \boldsymbol{V})$
Least-squares estimator of $\boldsymbol{\beta}$	$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$	$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{y}$
$\mathrm{Var}(\widehat{\boldsymbol{\beta}})$	$(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\sigma_e^2$	$(\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$

- Likelihood ratio test

$$\Lambda = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}$$

- Fixed vs Random effects
  - Fixed: The levels in the study represent all levels of interest, e.g. gender
  - Random: The levels are considered to be drawn from an infinite population of levels, e.g. a batch