Statistical Data Analysis

Jana de Wiljes
wiljes@uni-potsdam.de
www.dewiljes-lab.com

31. Oktober 2022

Universität Potsdam

Linear regression

Model for simple linear regression

Model:

$$Y_i = f(X_i, \beta) + \epsilon_i, \quad i = 1, \dots, n \tag{1}$$

where ϵ_i are iid with $\mathbb{E}[\epsilon_i] = 0$ and $Var(\epsilon_i) = \sigma^2$

Data: it is possible to observe realisations

$$(y_i, x_i) \quad i = 1, \dots, n \tag{2}$$

Goal: estimate parameters β of the function to obtain approximative $f(x, \hat{\beta})$

Note: note that f approximates $\mathbb{E}[Y_i | X_i]$

Model for simple linear regression

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, \dots, n \tag{3}$$

where ϵ_i are iid with $\mathbb{E}[\epsilon_i] = 0$ and $Var(\epsilon_i) = \sigma^2$

Data:

$$(y_i, x_i) \quad i = 1, \dots, n \tag{4}$$

Goal: estimate $f(x, \hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 x$

The Ordinary Multiple Linear Regression Model

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \cdots + \beta_p X_{i,p} + \epsilon_i, \quad i = 1, \ldots, n \quad (5)$$

where ϵ_i are iid with $\mathbb{E}[\epsilon_i] = 0$ and $Var(\epsilon_i) = \sigma^2$

Data:

$$(y_i, x_i) \quad i = 1, \ldots, n \quad (6)$$

Goal: estimate $\hat{f}(x_1, \ldots, x_p, \hat{\beta}_1, \ldots, \hat{\beta}_p) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots, \hat{\beta}_p x_p$

Multivariate Random Variables

Def: Let $\mathbf{X}$ be a vector of (univariate) random variables, i.e., $\mathbf{X} = (X_1, \ldots, X_p)^\top$ with $\mathbb{E}[X_i] = \mu_i$. $\mathbf{X}$ is called a multivariate random variable and we denote $\mathbb{E}[\mathbf{X}] = \mu$

Note:

- Variance $\text{Var}(X_i) = \mathbb{E}[(X_i - \mathbb{E}(X_i))^2] = \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_i - \mathbb{E}(X_i))]$
- Covariance $\text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]$

Least squares Estimation: minimize the sum of squared errors

Least squares estimation: minimize the sum of squared errors

$$L(\beta) = \sum_{i=1}^{N} (y_i - x_i^\top \beta_i)^2 = \sum_{i=1}^{N} \epsilon_i^2 = \epsilon^\top \epsilon \tag{7}$$

with respect to $\beta \in \mathbb{R}^{p+1}$

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \epsilon_i^2 = \epsilon^\top \epsilon = (Y - X\beta)^\top (Y - X\beta) \tag{8}$$

$$Y = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{\in \mathbb{R}^{N \times 1}}, \quad X = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \\ 1 & x_{N1} & \dots & x_{Np} \end{bmatrix}}_{\in \mathbb{R}^{N \times p+1}}, \beta = \underbrace{\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}}_{\in \mathbb{R}^{p+1 \times 1}}, \epsilon = \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix}}_{\in \mathbb{R}^{N \times 1}},$$

Covariance

Def: The covariance of the multivariate random variable $\mathbf{X}$ is defined by

$$\Sigma := \mathrm{Cov}(\mathbf{X}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top] \tag{9}$$

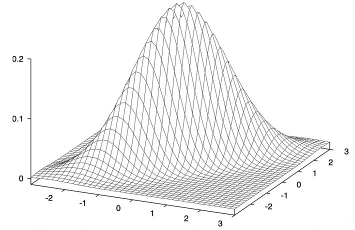
Example:

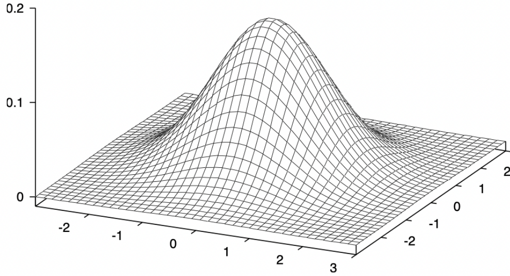
$$\Sigma = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{pmatrix} \tag{10}$$

Properties of Σ:

- quadractic
- symmetric
- positive-semidefinite

Multivariate Normal Distribution





$$\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma) \tag{11}$$

Positive semi-definite

Lemma: Let $\mathbf{B}$ be an $n \times (p+1)$ matrix. Then the matrix $\mathbf{B}^\top\mathbf{B}$ is symmetric and postive semi-definite. It is positive definite, if $\mathbf{B}$ has full column rank. Then, besides $\mathbf{B}^\top\mathbf{B}$ also $\mathbf{B}\mathbf{B}^\top$ is postive semi-definite.

LS-estimator

Theorem: The LS-estimator of the unknown parameters β is

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \tag{12}$$

if $\mathbf{X}$ has full column rank p+1.

Proof:

Proof

Positive semi-definite

Proposition: The hat-matrix $\mathbf{H} = (h_{ij})_{1 \leq i,j, \leq b}$ has the followi-ng properties:

1. $\mathbf{H}$ is symmetric
2. $\mathbf{H}$ is idempotent, i.e., $\mathbf{HH} = \mathbf{H}$
3. $rk(\mathbf{H}) = tr(\mathbf{H}) = p + 1$
4. $0 \leq h_{ii} \leq 1, \quad \forall i = 1, \ldots, n$
5. the matrix $\mathbf{I}_n - \mathbf{H}$ is also symmetric and idempotent with $rk(\mathbf{I}_n - \mathbf{H}) = n - p - 1$

ML-estimator

Theorem: The ML-estimator of the unknown parameters σ^2 is $\hat{\sigma}^2_{ML} = \frac{\hat{\epsilon}\hat{\epsilon}}{n}$ with $\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}$.

Proof

ML-estimator

Proposition: For the ML-estimator $\hat{\sigma}^2_{ML}$ of σ^2 the following property holds:

$$\mathbb{E}[\sigma^2_{ML}] = \frac{n - p - 1}{n}\sigma^2 \tag{13}$$

Proof

Adjusted estimator

Proposition: The adjusted estimator

$$\hat{\sigma}^2_{ad} = \frac{\hat{\epsilon}\hat{\epsilon}}{n - p - 1} \tag{14}$$

of the unknown parameter σ^2 can be written as

$$\hat{\sigma}^2_{ad} = \frac{\mathbf{y}^\top \mathbf{y} - \hat{\beta}^\top \mathbf{X}^\top \mathbf{y}}{n - p - 1} \tag{15}$$

Proof

ML estimator

Proposition: The LS-estimator $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is equivalent to the ML-estimator based on maximization of the log-likelihood

$$l(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \tag{16}$$

Proof

LS estimator

Proposition: The LS-estimator $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$ and the REML-estimator $\hat{\sigma}^2 = \frac{1}{n-p-1}\hat{\epsilon}^\top \hat{\epsilon}$ the following properties hold:

1. $\mathbb{E}[\hat{\beta}] = \beta$, $\mathrm{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$
2. $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$

Linear estimator

Def: A linear estimator has the form

$$\hat{\beta}^L = \mathbf{b} + \mathbf{A}\mathbf{y} \tag{17}$$

where $\mathbf{b} \in \mathbb{R}^{(p+1)\times 1}$ and $\mathbf{A} \in \mathbb{R}^{(p+1)\times n}$.

Example: The LS-estimator:

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y} \tag{18}$$

is a linear estimator with $\mathbf{b} = \mathbf{0}$ and $\mathbf{A} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$