Part I: Linear Models

Chapter I.4

Linear Models – Introduction & Estimation

Topics

To be discussed...

- Definition of a linear model
- Examples: simple linear regression, one-way ANOVA & ANCOVA
- Least squares estimators & properties
- Identifiability & estimability

A preliminary note

I.4.1 Remark

There are many aspects that can be discussed in the framework of linear models:

- estimation concepts
- probabilistic background
- model assumptions & diagnostics
  - normality, independence, heteroscedasticity, outliers
- testing
- model selection
- prediction
- multiple comparisons
- random effects
- experimental design
- ...

In the following, we will consider only a selection of topics with a focus on introductory probabilistic and inferential aspects.

I.4.2 Definition (Linear Model (LM))

Let $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} P$ with $E\varepsilon_1 = 0$ and $\operatorname{Var} \varepsilon_1 = \sigma^2 > 0$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)' \in \Theta \subseteq \mathbb{R}^p$ be a parameter space and $B \in \mathbb{R}^{n \times p}$ be a (known) matrix.

Then, we call

- $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ **linear model (LM)**,
- $\mathbf{Y} = (Y_1, \dots, Y_n)'$ random vector of observations,
- B design matrix,
- $\boldsymbol{\beta}$ parameter vector
- $\boldsymbol{\varepsilon}$ error term.

I.4.3 Definition (Normal Linear Model (NoLM))

A LM as in Definition I.4.2 with normally distributed error terms, that is, $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$ with $\sigma^2 > 0$, is called **normal linear model (NoLM)**.

I.4.4 Remark

- The model is called linear because $g(\boldsymbol{\beta}) = B\boldsymbol{\beta}$ is a linear function of the parameter $\boldsymbol{\beta}$.
- $\mathbf{Y} = (Y_1, \dots, Y_n)'$ are observed $(\mathbf{Y} = \mathbf{y})$.
- $\boldsymbol{\varepsilon}$ is unobservable.
- The parameters $\boldsymbol{\beta} \in \Theta = \mathbb{R}^p$ and $\sigma^2 > 0$ are supposed unknown. σ^2 may be considered as nuisance parameter, that is, it is not of primary interest. However, we will also provide estimators for σ^2.

  The (complete) parameter space is given by $\widetilde{\Theta} = \Theta \times (0, \infty)$.
- The model can also be discussed under the assumption of correlated errors, that is, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$ has a covariance $\mathrm{Cov}\,(\boldsymbol{\varepsilon}) = \sigma^2 V$ with a (known) matrix $V \in \mathbb{R}^{n \times n}_{>0}$.
- In our model, the design matrix B is a fixed matrix. In the literature, LMs with random design matrix are also discussed (see, e.g., Rencher & Schaalje 2008, Chapter 10).
- For simplicity, LM are often discussed under the following assumptions:
  - $n \geqslant p$
  - the design matrix B has maximal rank p.

  These assumptions guarantee that $B'B$ is a regular matrix so that its inverse exists.

I.4.5 Remark

For a NoLM as in Definition I.4.3, we have

- For $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $\sigma^2 > 0$ (or $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 I_n)$): $\boldsymbol{\varepsilon}$ has density function

$$f(\mathbf{t}) = \prod_{j=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-t_j^2/(2\sigma^2)} \right] = \frac{1}{(\sqrt{2\pi})^n \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{t}\|^2 \right), \quad \mathbf{t} \in \mathbb{R}^n.$$

- Using Theorem I.2.11, we get:

$$\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon} \sim N_n(B\boldsymbol{\beta}, \sigma^2 I_n)$$

so that

$$f^{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(\sqrt{2\pi})^n \sigma^n} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - B\boldsymbol{\beta}\|^2 \right), \quad \mathbf{y} \in \mathbb{R}^n.$$

- In particular (even without a normal distribution assumption):

$$E\mathbf{Y} = B\boldsymbol{\beta}, \quad \text{Cov}\,(\mathbf{Y}) = \sigma^2 I_n.$$

☞ Inference for the mean $B\boldsymbol{\beta}$ and the variance σ^2.

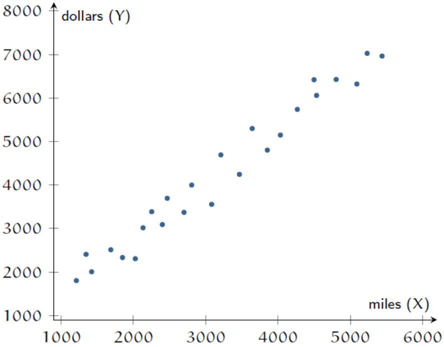
Remark

- The random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ is the *response* of interest.
- The design matrix B contains the *explanatory variables* $\mathbf{X}_j$, $j = 1, \ldots, m$ $(m \geqslant 1)$, with $\mathbf{X}_j = (X_{1j}, \ldots, X_{nj})'$ (usually $p = m + 1$).
- The responses are random to be modeled while <u>here</u> the explanatory variables are fixed $\mathbf{X}_j = \mathbf{x}_j$, $j = 1, \ldots, m$.
- The type of the explanatory variables specifies the type of the linear model.
  If the explanatory variables are
  - all quantitative ☞ linear regression
  - all qualitative ☞ analysis of variance (ANOVA)
  - some are quantitative and other qualitative ☞ analysis of covariance (ACOVA)
- Main goals
  - Assess the dependence of response on explanatory variables
  - Estimation and specification of models for the underlying *response distribution*

Some Data

I.4.6 Example

Consider the following data where X denotes the number of miles traveled by a credit card holder and Y denotes the charges (in US$). The credit card company suspects that the charges increase with the number of traveled miles.



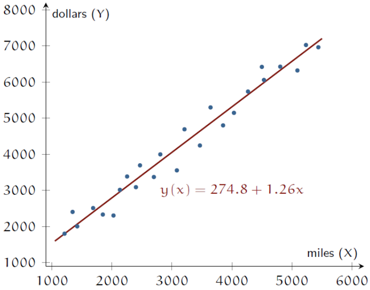
Passenger	Miles (X)	Dollars (Y)
1	1211	1802
2	1345	2405
3	1422	2005
4	1687	2511
5	1849	2332
6	2026	2305
7	2133	3016
8	2253	3385
9	2400	3090
10	2468	3694
11	2699	3371
12	2806	3998
13	3082	3555
14	3209	4692
15	3466	4244
16	3643	5298
17	3852	4801
18	4033	5147
19	4267	5738
20	4498	6420
21	4533	6059
22	4804	6426
23	5090	6321
24	5233	7026
25	5439	6964

I.4.6 Example (simple linear regression as LM)

Let $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $\sigma^2 > 0$, $\beta_0, \beta_1 \in \mathbb{R}$ be a parameter and $x_1, \ldots, x_n \in \mathbb{R}$. Then

$$Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j, \quad j = 1, \ldots, n,$$

forms a LM $\mathbf{Y} = B\boldsymbol{\beta} + \varepsilon$ with parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and design matrix $B = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = [\mathbb{1}_n \mid \mathbf{x}] \in \mathbb{R}^{n \times 2}$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)'$.



I.4.7 Example

Wine prices in Finger Lake area (New York): Cayuga Lake, Keuka Lake, Seneca Lake [a]

Price	Location	Price	Location	Price	Location
123	Seneca	80	Cayuga	52	Seneca
112	Seneca	118	Keuka	95	Seneca
151	Seneca	100	Cayuga	66	Seneca
143	Keuka	110	Seneca	115	Seneca
70	Seneca	78	Seneca	103	Seneca
92	Seneca	118	Seneca	131	Keuka
115	Cayuga	128	Keuka	135	Cayuga
100	Cayuga	75	Seneca	105	Seneca
90	Seneca	72	Cayuga	138	Keuka
93	Seneca	106	Seneca	59	Seneca
97	Seneca	130	Cayuga	76	Cayuga
88	Cayuga	60	Seneca	85	Keuka

[a] Map of the Oswego River drainage basin, with the Seneca River highlighted. Karl Musser/CC BY-SA 3.0.

I.4.7 Example (one factorial analysis of variance as LM)

Let $n_1, n_2, n_3 \in \mathbb{N}$, $(\varepsilon_{ij})_{i=1,\ldots,3,j=1,\ldots,n_i} \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $\sigma^2 > 0$, and $\mu, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ be parameters. Then,

$$\begin{aligned} Y_{1j} &= \mu + \alpha_1 + \varepsilon_{1j}, \quad j = 1, \ldots, n_1, \\ Y_{2j} &= \mu + \alpha_2 + \varepsilon_{2j}, \quad j = 1, \ldots, n_2, \\ Y_{3j} &= \mu + \alpha_3 + \varepsilon_{3j}, \quad j = 1, \ldots, n_3, \end{aligned}$$

forms a LM $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with parameter vector $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3)'$, design matrix

$$B = \begin{pmatrix} 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{bmatrix} \mathbb{1}_{n_1} & \mathbb{1}_{n_1} & 0 & 0 \\ \mathbb{1}_{n_2} & 0 & \mathbb{1}_{n_2} & 0 \\ \mathbb{1}_{n_3} & 0 & 0 & \mathbb{1}_{n_3} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2+n_3)\times 4}$$

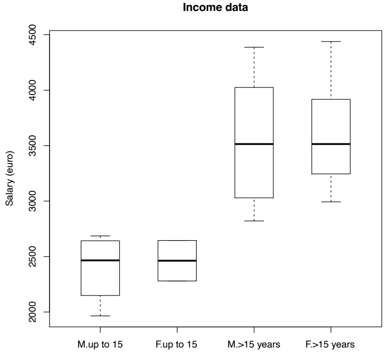
and $\mathbf{Y} = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, Y_{31}, \ldots, Y_{3n_3})'$. Notice $\operatorname{rank}(B) = 3 < 4$.

α_i are called **treatment effects**, μ is called **overall mean**.

Example (analysis of covariance - ANCOVA)

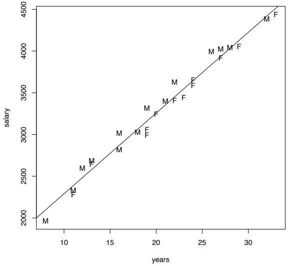
A company claims that the salaries of its employes of a certain sector depend only on their seniority (in years) and men are not better paid than women. The data of the $n = 26$ employes of this sector are given below. Note that the *mean salary* for men and women are 3306.115 and 3393.667 euro, respectively, while the box-plot does not seem to support a gender bias with respect to seniority ($>$15 years in company).

	years in company	gender	salary
i	(x_{i1})	(x_{i2})	(y_i)
1	21	M	3400
2	27	F	3917
3	13	M	2686
4	12	M	2598
$\vdots$	$\vdots$	$\vdots$	$\vdots$
24	11	M	2333
25	23	F	3440
26	19	F	2993



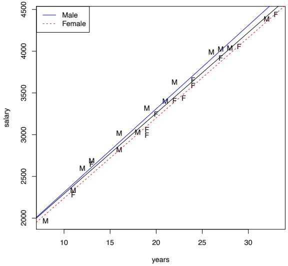
Example (ANCOVA - continues...)

Regressing salaries on seniority (in years):



$\hat{\mu}_i = 1323.16 + 96.73x_{i1}, \quad i = 1, \ldots, 26$

$R^2 = 0.978$



M $(n_1 = 14)$**:** $\hat{\mu}_i = 1310.48 + 99.96x_{i1}$

$R^2 = 0.982$

F $(n_2 = 12)$ **:** $\hat{\mu}_i = 1276.94 + 96.22x_{i1}$

$R^2 = 0.989$

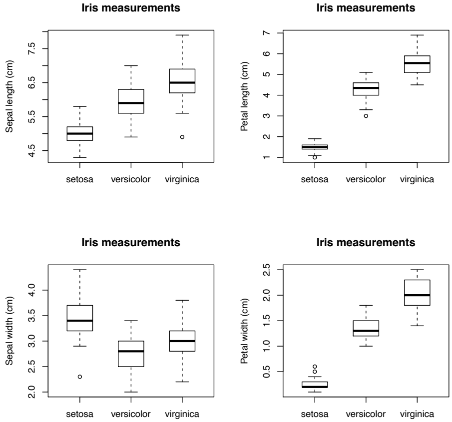
ancova $(n = 26)$**:**

$\hat{\mu}_i = 1338.65 + 98.49x_{i1} - 111.77 \cdot \mathrm{I}(x_{i2} = \mathrm{F})$

$R^2 = 0.985$

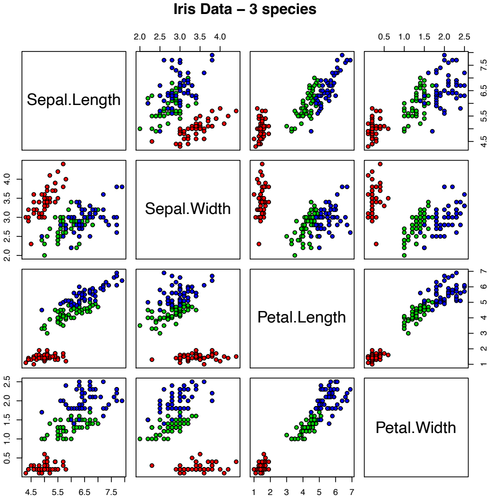
Example (Iris data set of Fisher)

This famous data set gives the measurements (in cm) of the variables *sepal length* and *width* and *petal length* and *width*, respectively, for 50 flowers from each of 3 *species* of iris (sample size $n = 150$).



Example (Iris data: continues...)

Pairwise scatterplots for the Iris data set:



I.4.8 Definition

Estimator : generalized rule that produces estimate of a given quantity based on observed data

Estimate : a realization of estimator using specific values

Let $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ be a LM and $\mathbf{y}$ be a realisation of $\mathbf{Y}$. Then:

- A solution $\beta^* = \beta^*(\mathbf{y})$ of the minimization problem

$$\psi(\boldsymbol{\beta}) = \|\mathbf{y} - B\boldsymbol{\beta}\|^2 \longrightarrow \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \tag{I.1}$$

  is called **least-squares-estimate**.
- $\beta^*(\mathbf{Y})$ is called **least-squares-estimator (LSE)** for $\boldsymbol{\beta}$.
- For $A \in \mathbb{R}^{k \times p}$, a LSE of $A\boldsymbol{\beta}$ is defined as $A\boldsymbol{\beta}^*(\mathbf{Y})$.

I.4.9 Theorem (LSE in LM)

Given a LM with design matrix $B \in \mathbb{R}^{n \times p}$, a LSE $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\mathbf{y})$ has to satisfy the normal equations

generalized inverse is used if the design matrix doesn't have full rank or it isn't regular matrix (det(B) > 0)

$$B'\mathbf{y} = B'B\widehat{\boldsymbol{\beta}}.$$

hat : estimator or estimate, hier is estimator

If B^+ and $(B'B)^+$ denote the Moore-Penrose inverse of B and $B'B$, respectively, a solution is given by

LSE based on moore-penrose inverse

$$\widehat{\boldsymbol{\beta}}^+ = B^+\mathbf{y} = (B'B)^+B'\mathbf{y} \quad \text{with } \psi(\widehat{\boldsymbol{\beta}}^+) = \mathbf{y}'(I_n - B(B'B)^+B')\mathbf{y}.$$

The set of all LSEs is given by $\{\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^+ + (I_p - B^+B)\mathbf{z} \mid \mathbf{z} \in \mathbb{R}^p\}$.

I.4.10 Corollary (decomposition formula)

Let $\widehat{\beta}$ be a LSE in a LM with design matrix B. Then, for any $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\psi(\boldsymbol{\beta}) = \|\mathbf{y} - B\widehat{\boldsymbol{\beta}}\|^2 + \|B(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\|^2 = \psi(\widehat{\boldsymbol{\beta}}) + \|B(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\|^2 \tag{I.2}$$

I.4.11 Remark

Let $Q = B(B'B)^+B'$ be the (unique) orthogonal projector on $\mathrm{Im}(B)$ (see Theorem I.1.10).

- Notice $Q = Q'$ and $Q^2 = B(B'B)^+B'B(B'B)^+B' = B(B'B)^+B' = Q$.
- $P = I_n - Q$ satisfies $P = P'$ and $P^2 = P$, too.
- $\mathrm{rank}(Q) = \mathrm{rank}(B)$, $\mathrm{rank}(P) = n - \mathrm{rank}(B)$.
- From Theorem I.4.9, we get that an estimator $\widehat{\boldsymbol{\beta}}$ is an LSE of $\boldsymbol{\beta}$ iff $B\widehat{\boldsymbol{\beta}} = Q\mathbf{Y}$ (show equivalence of equations by properties given in Theorem I.1.10).

If a matrix symmetric and idempotent, its eigenvalues are 0 or 1.

In the following, we will use the orthogonal projectors Q and P as defined in Remark I.4.11, that is,

$$Q = B(B'B)^+B', \quad P = I_n - Q = I_n - B(B'B)^+B'.$$

I.4.12 Corollary (LSE in LM)

regular matrix

Given a LM with design matrix B satisfying $\det(B'B) > 0$, the unique LSE is given by

$$\widehat{\beta} = (B'B)^{-1}B'\mathbf{Y}.$$

I.4.13 Remark

Kernel(B) = 0

In case the LSE is unique, we write for short $\widehat{\beta}$ instead of $\widehat{\beta}^{+}$ for the LSE based on the Moore-Penrose inverse matrix.

Say that the system $Ax=b$ Ax=b
has a solution. Denote such a solution by x0.x0.
Let xh xh(the $h$hstands for "homogenous") denote a nonzero solution to the system $Ax=0$
Ax=0(called the "homogenous system"). It is true that x0x0 and $x0+xh$
x0+xh are two separate solutions to the system $Ax=b$.
Ax=b.
So, if there is any nonzero vector in ker$A$,
kerA, then $Ax=b$ Ax=b has two separate solutions, meaning that it cannot be invertible.

Identifiability

I.4.14 Remark

Given a NoLM $\mathbf{Y} = B\boldsymbol{\beta} + \varepsilon$, the distributional assumption reads

$$\mathscr{P} = \{N_p(B\boldsymbol{\beta}, \sigma^2 I_p) \mid \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 > 0\}.$$

This rises the question of **identifiability**.Here, this means that the probability distribution is uniquely specified by the parameter:

$$B\boldsymbol{\beta}_1 = B\boldsymbol{\beta}_2 \quad \Longrightarrow \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2.$$

If B has maximal rank then $B'B$ is a regular matrix. Thus, for $\boldsymbol{\beta}_i$ with $B\boldsymbol{\beta}_1 = B\boldsymbol{\beta}_2$, we get:

$$\boldsymbol{\beta}_1 = (B'B)^{-1}B'B\boldsymbol{\beta}_1 = (B'B)^{-1}B'B\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2.$$

the mean value of the distribution is same

I.4.15 Example

Consider a LM with design matrix $B = \mathbb{1}_{n\times 2}$. Then, with $\Theta = \mathbb{R}^2$, we have that $\mathrm{Ker}(B) = \{\lambda\binom{1}{-1} \mid \lambda \in \mathbb{R}\}$. Therefore, each parameter $\boldsymbol{\beta} \in \mathrm{Ker}(B)$ yields the same probability distribution, that is, $N_n(0, \sigma^2 I_n)$. As a consequence, it is not possible to estimate $\boldsymbol{\beta}$ meaningful from the data.

It is important to determine which parameters can be estimated and which are not.
Those we can estimate are called identifiable. Linear functions of the identifiable parameters are called estimable

I.4.16 Definition

Let $\mathbf{Y} \sim P_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \Theta$, be a statistical model and $E\mathbf{Y} = h(\boldsymbol{\beta})$ with a known function h. Then, the parameter $\boldsymbol{\beta}$ is called **identifiable** if for any $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, the equation $h(\boldsymbol{\beta}_1) = h(\boldsymbol{\beta}_2)$ implies $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

If $\boldsymbol{\beta}$ is identifiable, then the parametrization $h(\boldsymbol{\beta})$ is called identifiable. A function $g(\boldsymbol{\beta})$ is called identifiable if $h(\boldsymbol{\beta}_1) = h(\boldsymbol{\beta}_2)$ implies $g(\boldsymbol{\beta}_1) = g(\boldsymbol{\beta}_2)$

I.4.17 Remark

Given a LM with design matrix $B \in \mathbb{R}^{n \times p}$, identifiability of $\boldsymbol{\beta}$ means that

$$h(\boldsymbol{\beta}) = B\boldsymbol{\beta}$$

is identifiable. Hence, $\text{Ker}(B) = \{0\}$, which is equivalent to $\text{rank}(B) = p$. This implies that $B'B$ has rank p and, thus, is a regular matrix.

I.4.18 Theorem

Consider a LM with design matrix $B \in \mathbb{R}^{n \times p}$ and $g : \mathbb{R}^p \to \mathbb{R}^k$. Then, $g(\boldsymbol{\beta})$ is identifiable if and only if g is a function of $B\boldsymbol{\beta}$, that is, it exists a function $\nu : \mathbb{R}^n \to \mathbb{R}^k$ such that $g(\boldsymbol{\beta}) = \nu(B\boldsymbol{\beta})$.

Proof: see Christensen (2011, p.19)

In view of Theorem I.4.18, we introduce the following concept.

I.4.19 Definition

Consider a LM with design matrix $B \in \mathbb{R}^{n\times p}$ and let $C \in \mathbb{R}^{p\times k}$ be a matrix. Then, the function $g(\boldsymbol{\beta}) = C'\boldsymbol{\beta}$ is called (linear) **estimable** if $C' = Q'B$ for some matrix $Q \in \mathbb{R}^{n\times k}$.

I.4.20 Remark

- From a statistical point of view, it makes only sense to consider estimates of identifiable functions. Otherwise, it is not clear what you are really estimating.

  Given a LM, we have with Theorem I.4.18 that $C'\boldsymbol{\beta}$ is identifiable if $C'\boldsymbol{\beta}$ is a linear function of $B\boldsymbol{\beta}$. Hence, a matrix $Q \in \mathbb{R}^{n\times k}$ must exist with $C'\boldsymbol{\beta} = Q'B\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$. This leads to the condition given in Definition I.4.19.
- For $k = 1$, the above definition reads: For $\mathbf{c} \in \mathbb{R}^p$, $\mathbf{c}'\boldsymbol{\beta}$ is called (linear) **estimable** if $\mathbf{q} \in \mathbb{R}^n$ exists with $\mathbf{q}'B = \mathbf{c}'$.

I.4.21 Example

Consider the simple regression model I.4.6 with design matrix $B' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \in \mathbb{R}^{2\times n}$ such that $\operatorname{rank}(B) = 2$ and let $\mathbf{c} \in \mathbb{R}^2$. Then, $\mathbf{c}'\boldsymbol{\beta}$ is estimable if $\mathbf{c} = B'\mathbf{q}$ for some $\mathbf{q} \in \mathbb{R}^n$, that is, it exists $q \in \mathbb{R}^n$ with

$$c_1 = \sum_{i=1}^n q_i, \quad c_2 = \sum_{i=1}^n x_i q_i.$$

Since $\operatorname{rank}(B) = 2$, we have that $\operatorname{Im}(B') = \mathbb{R}^2$. Therefore, for any $\mathbf{c} \in \mathbb{R}^2$, such a $\mathbf{q}$ exists. Notice that $\mathbf{q}$ is not unique in general.

Thus, for instance, the following functions are (linear) estimable:

- $\mathbf{c}'\boldsymbol{\beta} = \beta_0$ (choose $\mathbf{c}' = (1, 0)$)
- $\mathbf{c}'\boldsymbol{\beta} = \beta_1$ (choose $\mathbf{c}' = (0, 1)$)
- $\mathbf{c}'\boldsymbol{\beta} = \beta_0 + \beta_1 x$ (choose $\mathbf{c}' = (1, x)$ with $x \in \mathbb{R}$ arbitrarily)

LSE of $\mathbf{c}'\mathrm{B}\boldsymbol{\beta}$

I.4.22 Corollary

Given a LM and $\mathbf{c} \in \mathbb{R}^n$, the unique LSE of $\mathbf{c}'\mathrm{B}\boldsymbol{\beta}$ is given by $\mathbf{c}'\mathrm{B}(\mathrm{B}'\mathrm{B})^+\mathrm{B}'\mathbf{Y}$.

Proof

From Remark I.4.11, we know that any LSE $\widehat{\boldsymbol{\beta}}$ has to satisfy the equation $\mathrm{B}\widehat{\boldsymbol{\beta}} = \mathrm{Q}\mathbf{Y} = \mathrm{B}(\mathrm{B}'\mathrm{B})^+\mathrm{B}'\mathbf{Y}$. This implies

$$\mathbf{c}'\mathrm{B}\boldsymbol{\beta} = \mathbf{c}'\mathrm{Q}\mathbf{Y} = \mathbf{c}'\mathrm{B}(\mathrm{B}'\mathrm{B})^+\mathrm{B}'\mathbf{Y}$$

showing that the LSE of $\mathbf{c}'\mathrm{B}\boldsymbol{\beta}$ is unique.

I.4.23 Remark

As shown in Christensen (2011, Theorem 2.2.4), a LSE of $\mathbf{c}'\boldsymbol{\beta}$ is unique iff $\mathbf{c}'\boldsymbol{\beta}$ is estimable, that is, $\mathbf{c}' = \mathbf{q}'\mathrm{B}$ for some $\mathbf{q} \in \mathbb{R}^n$.

I.4.24 Theorem (see Theorem I.4.9)

Given a LM with design matrix B and $\text{rank}(B) = r < n$, LSE $\widehat{\beta}$, and $P = I_n - B(B'B)^+B'$. Then, $\widehat{\sigma^2} = \frac{1}{n-r}\psi(\widehat{\beta}) = \frac{1}{n-r}\mathbf{Y}'P\mathbf{Y}$ is an unbiased estimator of σ^2.

unbiased estimator of a parameter is an estimator whose expected value is equal to the parameter

Lemma 3.3

Let Y be a random vector with EY = μ and Cov (Y) = Σ and A ∈ Rp×p.
Then, EY ′AY = trace(AΣ) + μ′Aμ.

Properties of the LSE $\widehat{\beta}^+$

I.4.25 Theorem

Given a LM with design matrix B, the LSE $\widehat{\beta}^+$ has the following properties:

1. $E\widehat{\beta}^+ = B^+B\beta$, Cov $(\widehat{\beta}^+) = \sigma^2(B'B)^+$
2. Under a NoLM, $\widehat{\beta}^+ \sim N_p(B^+B\beta, \sigma^2(B'B)^+)$

If the design matrix B has $\text{rank}(B) = p$, then the (unique) LSE $\widehat{\beta}^+$ is an unbiased estimator for β, that is,

$E\widehat{\beta}^+ = \beta$ and the Moore-Penrose inverse $(B'B)^+$ equals the inverse matrix $(B'B)^{-1}$.

In particular, we get under a NoLM, $\widehat{\beta} \sim N_p(\beta, \sigma^2(B'B)^{-1})$.

E(Y) = B^+B ß
this doesn't equal ß in general
In order to have E(y) = ß, we need
B^+B = B, => B^+ = B^-1