Part I: Linear Models

Chapter I.5

Regression Models – Estimation

Topics

To be discussed...

- Simple linear regression
- Multiple linear regression
- Least-squares estimation
- Polynomial regression and general regression

Simple linear regression (s. Example I.4.6)

Consider the model of simple linear regression with $\operatorname{rank}(B) = 2$ with normally distributed error terms. Furthermore,

- $s_{xx} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2, s_{yy} = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \overline{y})^2$ denote the **samples variances**
- $s_{xy} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y})$ denotes the **sample covariance**.

Then, using Theorem I.4.25, the LSEs $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have the properties:

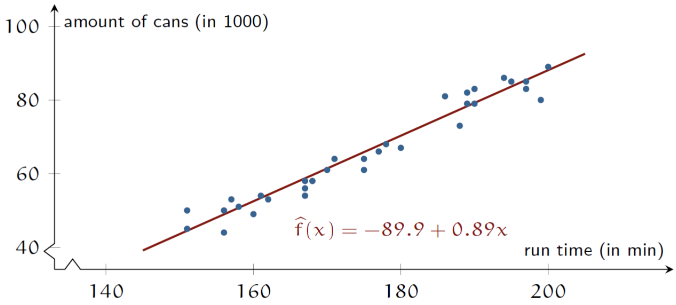
1. $\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \sigma^2\Sigma_0\right)$ mit $\Sigma_0 = (B'B)^{-1} = \frac{1}{(n-1)s_{xx}}\begin{pmatrix} \frac{n-1}{n}s_{xx} + \overline{x}^2 & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$,
2. $\widehat{Y}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x \sim N\left(\beta_0 + \beta_1 x, \frac{\sigma^2}{(n-1)s_{xx}}\left(\frac{n-1}{n}s_{xx} + (x - \overline{x})^2\right)\right), x \in \mathbb{R}.$ (**regression line**)

I.5.1 Example

Consider a bottling plant for cans. 32 measurements yield the following data with first component `run time` (in min) and second component `amount of cans` (in 1 000):

(189,82)	(189,79)	(180,67)	(199,80)	(197,83)	(186,81)	(200,89)	(162,53)	(195,85)	(197,85)	(158,51)	(194,86)
(157,53)	(188,73)	(168,58)	(167,54)	(175,64)	(151,45)	(175,61)	(156,44)	(190,79)	(160,49)	(190,83)	(170,61)
		(151,50)	(177,66)	(156,50)	(167,56)	(171,64)	(161,54)	(178,68)	(167,58)		

regression line



Simple linear regression (s. Example I.4.6)

From

$$\frac{1}{\sqrt{\mathbf{c}'(B'B)^{-1}\mathbf{c}}} \cdot \frac{\mathbf{c}'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sqrt{\|\mathbf{Y}-B\widehat{\boldsymbol{\beta}}\|^2/(n-p)}} \sim t(n-p)$$

(see Theorem I.4.32 ❹), we can construct confidence intervals und hypothesis tests by choosing the vector $\mathbf{c}$. For instance,

- $\mathbf{c}' = (1, 0)$ ☞ β_0
- $\mathbf{c}' = (0, 1)$ ☞ β_1
- $\mathbf{c}' = (1, x)$, $x \in \mathbb{R}$ ☞ $f(x) = \beta_0 + \beta_1 x$

I.5.2 Example

Using the above construction, a $(1-\alpha)$-level confidence interval for $f(x) = \beta_0 + \beta_1 x$, $x \in \mathbb{R}$ fixed, is given by

$$\left[\widehat{f(x)} - t_{1-\alpha/2}(n-2)\left(\frac{1}{n} + \frac{(x-\overline{x})^2}{n \cdot s_x^2}\right)\widehat{\sigma}, \widehat{f(x)} + t_{1-\alpha/2}(n-2)\left(\frac{1}{n} + \frac{(x-\overline{x})^2}{n \cdot s_x^2}\right)\widehat{\sigma}\right],$$

where $t_{1-\alpha/2}(n-2)$ denotes the $(1-\alpha/2)$-quantile of the t-distribution with $n-2$ degrees of freedom, $p = 2$, and $\widehat{\sigma}^2 = \|\mathbf{Y} - B\widehat{\boldsymbol{\beta}}\|^2/(n-2)$.

Multiple linear regression

- **more** explanatory variables $x_1, \dots, x_m$ and **one** dependent variable Y
- **Regression function**:

$$f(x_1, \dots, x_m) = \beta_0 + \sum_{i=1}^{m} \beta_i x_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \tag{I.5}$$

- $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $\sigma^2 > 0$
- Estimates are based on observations: $(y_1, x_{11}, \dots, x_{1m}), \dots, (y_n, x_{n1}, \dots, x_{nm})$

I.5.3 Definition

A LM $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\beta} = (\beta_0, \dots, \beta_m)'$, and design matrix

$$B = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \in \mathbb{R}^{n \times (m+1)} \tag{I.6}$$

is called **multiple linear regression model**.

Multiple linear regression

I.5.4 Corollary

Let $n \geqslant m+1$ and suppose B has **maximal rank**, i.e. $\operatorname{rank}(B) = m+1 (= p)$. Then, the LSE in the multiple linear regression model is given by

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

with B as in (I.6).

I.5.5 Remark

- All statements true for a LM are valid in the model of multiple linear regression.
- Further regression models can be interpreted as multiple linear regression:

  e.g., polynomial regression models: $f(x) = \sum_{i=0}^{m} \beta_i x^i$, $x \in \mathbb{R}$, by defining $x_i = x^i$, $i = 1, \ldots, m$, with $x \in \mathbb{R}$ in (I.5).

Quadratic regression as multiple linear regression

For $Y = a + bX + cX^2 + \varepsilon$, we get with $X_1 = X$ **and** $X_2 = X^2$:

$$\boxed{\text{☞}\, Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \quad \beta_0, \beta_1, \beta_2 \in \mathbb{R}}$$

I.5.6 Representation

- Let $B = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}$. For $n \geqslant 3$ and $\operatorname{rank}(B) = 3$, we get $\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = (B'B)^{-1}B'\mathbf{Y}$.
- Notice that $\operatorname{rank}(B) = 3$ iff the observations are measured at a minimum of three different points x_i. This can be easily proven by elementary row operations (see Remark I.5.9).

Quadratic regression

I.5.7 Example

Consider a quadratic regression model with

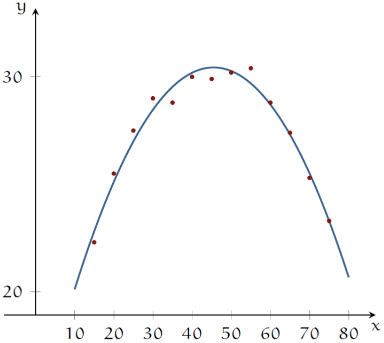
$$y = f(x) = \beta_0 + \beta_1 x + \beta_2 x^2, \quad x \in \mathbb{R},$$

and the data (speed in miles per hour, mileage in miles per gallon)

speed (X)	mileage (Y)
15	22.3
20	25.5
25	27.5
30	29.0
35	28.8
40	30.0
45	29.9
50	30.2
55	30.4
60	28.8
65	27.4
70	25.3
75	23.3

The LSE for the regression function is given by

$$\widehat{f}(x) = -0.0082x^2 + 0.746x + 13.47, \quad x \in \mathbb{R}.$$



Polynomial regression

I.5.8 Polynomial regression model (with regression function of grade m)

Consider

- $x_1, \ldots, x_n \in \mathbb{R}$
- Design matrix $B_{n,m} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{pmatrix}$ with $n \geqslant m+1$
- $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)' \in \mathbb{R}^{m+1}$ (unknown parameter)
- $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2),\ \sigma^2 > 0$

The LM $Y = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with regression function

$$\boxed{f(x) = \sum_{j=0}^{m} \beta_j x^j}$$

is called **polynomial regression model** (with regression function of grade m).

I.5.9 Remark

Using the notation of Definition I.5.8, the matrix $B_{m+1,m}$ is called **Vandermonde** matrix. Then:

- $\det B_{m+1,m} = \prod_{1 \leqslant i < j \leqslant m} (x_j - x_i)$.
- For $n \geqslant m+1$, $B_{n,m} = \begin{bmatrix} B_{m+1,m} \\ * \end{bmatrix}$. Hence, $B_{n,m}$ has full rank if $|\{x_1, \ldots, x_n\}| \geqslant m+1$.
- **Criterion**: For a polynomial regression with grade m, it is recommended to measure at at least $m+1$ different values in order to have a full rank design matrix.
  - For a simple linear regression, you need at least two different points.
  - For a quadratic regression, you need at least three different points, etc.

General regression model

I.5.10 General regression model (with known functions of $f_0, \dots, f_m$)

Consider

- $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^p$
- $f_j : \mathbb{R}^p \longrightarrow \mathbb{R}$, $0 \leqslant j \leqslant m$ (known functions)
- $\boldsymbol{\beta} = (\beta_0, \dots, \beta_m)' \in \mathbb{R}^{m+1}$ (unknown parameter)
- $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $\sigma^2 > 0$

The LM $Y = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with regression function

$$f(\boldsymbol{x}) = \sum_{j=0}^{m} \beta_j f_j(\boldsymbol{x})$$

is called **general regression model** (with known functions $f_0, \dots, f_m$).

General regression model

Matrix representation of general regression model

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \underbrace{\begin{pmatrix} f_0(\boldsymbol{x}_1) & f_1(\boldsymbol{x}_1) & f_2(\boldsymbol{x}_1) & \dots & f_m(\boldsymbol{x}_1) \\ f_0(\boldsymbol{x}_2) & f_1(\boldsymbol{x}_2) & f_2(\boldsymbol{x}_2) & \dots & f_m(\boldsymbol{x}_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ f_0(\boldsymbol{x}_n) & f_1(\boldsymbol{x}_n) & f_2(\boldsymbol{x}_n) & \dots & f_m(\boldsymbol{x}_n) \end{pmatrix}}_{= B} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

☞ Handling as for multiple linear regression

Part I: Linear Models

Chapter I.5

Regression Models – Testing

Topics

To be discussed...

- Analysis of variance
- Simple linear regression – F-test
- Multiple linear regression – F-test
- Linear hypotheses
- Test on a subset of the parameters
- Testing general linear hypothesis
- Confidence region for $\boldsymbol{\beta}$

Variance decomposition: Simple linear regression

I.5.11 Theorem

Let $B = [\mathbb{1}_n \mid \boldsymbol{x}] \in \mathbb{R}^{n\times 2}$ with $\boldsymbol{x} \in \mathbb{R}^n$, $\operatorname{rank}(B) = 2$, and $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Then:

1. $\mathsf{SST} = \mathsf{SSR} + \mathsf{SSE}$ with
   - $\mathsf{SST} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$ total sum of squares
   - $\mathsf{SSR} = \sum_{i=1}^{n}(\widehat{\beta}_0 + \widehat{\beta}_1 x_i - \overline{Y})^2$ sum of squares regression
   - $\mathsf{SSE} = \sum_{i=1}^{n}(Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i))^2 = \|\boldsymbol{Y} - B\widehat{\boldsymbol{\beta}}\|^2 = \psi(\widehat{\boldsymbol{\beta}})$ sum of squares error
2. $\mathsf{SSR}, \mathsf{SSE}$ are stochastically independent.
3. $\mathsf{SST}/\sigma^2 \sim \chi^2(n-1)$
4. $\mathsf{SSR}/\sigma^2 \sim \chi^2(1)$, $\mathsf{SSE}/\sigma^2 \sim \chi^2(n-2)$
5. $\dfrac{\mathsf{SSR}}{\mathsf{SSE}/(n-2)} \sim \mathsf{F}(1, n-2)$

ANOVA-table for simple linear regression

Theorem I.5.11 motivates the following summary:

Source of variation	degrees of freedom	sum of squares
regression	1	SSR
residual	$n-2$	SSE
total	$n-1$	SST

- Using the Testing Procedure I.4.40, that is, $F = \frac{\mathsf{SSR}}{\mathsf{SSE}/(n-2)} \sim \mathsf{F}(1, n-2)$, the model fit can be checked according to a given level $\alpha \in (0, 1)$ (comparison with quantile of F-distribution):

$$\text{Reject } H_0 \text{ if } \frac{\mathsf{SSR}}{\mathsf{SSE}/(n-2)} > \mathsf{F}_{1-\alpha}(1, n-2)$$

- The procedure compares the model $Y = \beta_0 + \beta_1 x + \varepsilon$ with the model $Y = \beta_0 + \varepsilon$, that is, it tests the null hypothesis

$$\boxed{H_0 : \beta_1 = 0.}$$

Variance decomposition: Multiple linear regression

I.5.12 Theorem

Let $B = (\mathbb{1}_n \mid B_1) \in \mathbb{R}^{n\times(m+1)}$ with $\operatorname{rank}(B) = m+1$, $J_0 = \mathbb{1}_{n\times n}/n$, $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $\overline{\mathbf{Y}} = \overline{Y}\mathbb{1}_n$ and $Q = B(B'B)^{-1}B'$. Then:

- $QJ_0 = J_0Q = J_0$, $(I_n - Q)(Q - J_0) = 0$
- $Q - J_0$ is an orthogonal projector.
- $\mathsf{SST} = \mathsf{SSR} + \mathsf{SSE}$ with
  - $\mathsf{SST} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \|\mathbf{Y} - \overline{\mathbf{Y}}\|^2 = \mathbf{Y}'(I_n - J_0)\mathbf{Y} = \mathbf{Y}'E_n\mathbf{Y}$
  - $\mathsf{SSR} = \|B\widehat{\boldsymbol{\beta}} - \overline{\mathbf{Y}}\|^2 = \mathbf{Y}'(Q - J_0)\mathbf{Y}$
  - $\mathsf{SSE} = \|\mathbf{Y} - B\widehat{\boldsymbol{\beta}}\|^2 = \mathbf{Y}'(I_n - Q)\mathbf{Y}$
- $\mathsf{SSR}, \mathsf{SSE}$ are stochastically independent.
- $\mathsf{SST}/\sigma^2 \sim \chi^2(n-1)$
- $\mathsf{SSR}/\sigma^2 \sim \chi^2(m)$, $\mathsf{SSE}/\sigma^2 \sim \chi^2(n-m-1)$
- $\dfrac{\mathsf{SSR}/m}{\mathsf{SSE}/(n-m-1)} \sim F(m,n-m-1)$

ANOVA-table for multiple linear regression

Theorem I.5.11 motivates the following summary:

Source of variation	degrees of freedom	sum of squares
regression	m	SSR
residual	$n-m-1$	SSE
total	$n-1$	SST

- Using the Testing Procedure I.4.40, that is, $\frac{\text{SSR}/m}{\text{SSE}/(n-m-1)} \sim F(m, n-m-1)$, the model fit can be checked according to a given level $\alpha \in (0,1)$ (comparison with quantile of F-distribution):

$$\text{Reject } H_0 \text{ if } \frac{\text{SSR}/m}{\text{SSE}/(n-m-1)} > F_{1-\alpha}(m, n-m-1)$$

- The procedure compares the model $Y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \varepsilon$ with the model $Y = \beta_0 + \varepsilon$, that is, it tests the null hypothesis

$$\boxed{H_0 : \beta_1 = \cdots = \beta_m = 0.}$$

Further linear hypotheses

I.5.13 Remark

- The previous framework investigates the problem

$$H_0 : \beta_1 = \cdots = \beta_m = 0 \quad \longleftrightarrow \quad H_1 : \text{ it exists } j \in \{1, \dots, m\} \text{ with } \beta_j \neq 0.$$

- What about other testing problems like, e.g., (for some fixed $k \in \{0, \dots, m\}$)
  1. $H_0 : \beta_k = 0 \quad \longleftrightarrow \quad H_1 : \beta_k \neq 0,$
  2. $H_0 : \beta_1 = \beta_k = 0 \quad \longleftrightarrow \quad H_1 : \text{ it exists } j \in \{1, k\} \text{ with } \beta_j \neq 0,$
  3. $H_0 : \beta_1 = \beta_m \quad \longleftrightarrow \quad H_1 : \beta_1 \neq \beta_m$?

Such hypotheses can be written as a linear hypothesis with a matrix $K \in \mathbb{R}^{q \times p}$ with $p = m + 1$ and $1 \leqslant q \leqslant p$, that is,

$$\boxed{H_0 : K\boldsymbol{\beta} = \boldsymbol{\delta} \quad \longleftrightarrow \quad H_1 : K\boldsymbol{\beta} \neq \boldsymbol{\delta}} \tag{I.7}$$

with a given $\boldsymbol{\delta} \in \mathbb{R}^q$.

- We continue with a test on a subset of the parameters (1 and 2 are special cases).

Test on a subset of the parameters

I.5.14 Remark

Consider a partition of $\boldsymbol{\beta} = \binom{\boldsymbol{\beta}_1}{\boldsymbol{\beta}_2}$ with $\boldsymbol{\beta}_i \in \mathbb{R}^{m_i}$, $m_i \geqslant 1$, $i = 1, 2$, $m_1 + m_2 = m + 1$ (otherwise rearrange the components) and the testing problem

$$H_0 : \boldsymbol{\beta}_2 = 0 \quad \longleftrightarrow H_1 : \boldsymbol{\beta}_2 \neq 0. \tag{I.8}$$

Partitioning the design matrix $B = [B_1 \mid B_2]$ similarly, the LM reads

$$\mathbf{Y} = [B_1 \mid B_2] \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \varepsilon = B_1 \boldsymbol{\beta}_1 + B_2 \boldsymbol{\beta}_2 + \varepsilon. \tag{I.9}$$

Then, we have under $H_0 : \boldsymbol{\beta}_2 = 0$ the reduced model

$$\mathbf{Y} = B_1 \boldsymbol{\beta}_1 + \varepsilon. \tag{I.10}$$

Notice that $\hat{\boldsymbol{\beta}}_1$ and $\hat{\sigma}^2$ from the reduced model will normally be different from the corresponding estimators of the full model.

Test on a subset of the parameters

I.5.15 Lemma

Let $Q = B(B'B)^{-1}B'$ and $Q_1 = B_1(B_1'B_1)^{-1}B_1'$. Considering eq. (I.9) and (I.10), we get the decomposition

$$\text{SST} = \text{SSE} + \text{SS}(\beta_2 \mid \beta_1) + \text{SSR (reduced)}$$

where

- $\text{SST} = \mathbf{Y}'E_n\mathbf{Y}$
- $\text{SSE} = \mathbf{Y}'(I_n - Q)\mathbf{Y}$
- $\text{SSR (reduced)} = \mathbf{Y}'(Q_1 - \frac{1}{n}\mathbb{1}_{n \times n})\mathbf{Y}$
- $\text{SS}(\beta_2 \mid \beta_1) = \mathbf{Y}'(Q - Q_1)\mathbf{Y}$ denotes the 'extra' regression sum of squares due to β_2 after adjusting for β_1. Notice that $\text{SS}(\beta_2 \mid \beta_1) = \text{SSR (full)} - \text{SSR (reduced)}$.

Furthermore, the matrices $I_n - Q, Q - Q_1$ are orthogonal projectors satisfying $QQ_1 = Q_1Q = Q_1$.

I.5.16 Proof

The result follows from Theorem I.4.39 since $\text{Im}(B_1) \subsetneq \text{Im}(B)$ by the definition of B_1.

Test on a subset of the parameters

Using Theorem I.4.39 and Lemma I.5.15, we can construct a test procedure for the testing problem (I.8).

I.5.17 Testing procedure (F-test)

Given a LM with design matrix B and the testing problem (I.8), let $Q = B(B'B)^{-1}B'$ and $Q_1 = B_1(B_1'B_1)^{-1}B_1'$ as above and $m + 1 = \text{rank}(B), m_1 = m + 1 - m_2 = \text{rank}(B_1)$.

Then, an α-level statistical test for H_0 is given by the decision rule

$$\text{Reject } H_0 \text{ if } \frac{\mathbf{Y}'(Q - Q_1)\mathbf{Y}/m_2}{\mathbf{Y}'(I_n - Q)\mathbf{Y}/(n - m - 1)} > F_{1-\alpha}(m_2, n - m - 1)$$

where $F_{1-\alpha}(m_2, n - m - 1)$ denotes the $(1 - \alpha)$-quantile of the $F(m_2, n - m - 1)$-distribution.

I.5.18 Example

Suppose we want to test

$$H_0 : \beta_k = 0 \quad \longleftrightarrow \quad H_1 : \beta_k \neq 0$$

for some given $k \in \{0, \dots, m\}$ in the multiple linear regression model I.5.3. Then, using Theorem I.4.32 ④, an appropriate test statistic is given by

$$T_k = \frac{\widehat{\beta}_k}{\sqrt{((B'B)^{-1})_{kk} \|\mathbf{Y} - B\widehat{\boldsymbol{\beta}}\|^2/(n-m-1)}} \sim t(n-m-1)$$

The corresponding decision rule reads

$$\boxed{\text{Reject } H_0 \text{ if } |T_k| > t_{1-\alpha/2}(n-m-1),}$$

where $t_{1-\alpha/2}(n-m-1)$ is the $(1-\alpha/2)$-quantile of the $t(n-m-1)$-distribution.

Notice that this test is equivalent to the F-test given in I.5.17 ($m_2 = 1!$). Clearly, the test procedure can be easily adapted to the testing problem

$$H_0 : \beta_k = \delta \quad \longleftrightarrow \quad H_1 : \beta_k \neq \delta$$

with $\delta \in \mathbb{R}$.

Decomposition of multiple linear regression model

Consider a multiple linear regression model as in Definition I.5.3 and write the design matrix as block matrix

$$B = [\mathbf{b}_1 \mid \mathbf{b}_2 \mid \cdots \mid \mathbf{b}_{m+1}], \quad B_j = [\mathbf{b}_1 \mid \mathbf{b}_2 \mid \cdots \mid \mathbf{b}_j], \quad j \in \{1, \ldots, m+1\}.$$

For simplicity, assume that $\operatorname{rank}(B) = m + 1$, that is, the columns of B are linearly independent. Furthermore, we consider only models with intercept, that is, a constant β_0 is included in the model. Thus, $\mathbf{b}_1 = \mathbb{1}_n$.

Let $\boldsymbol{\beta}^{(k)} = (\beta_0, \ldots, \beta_{k-1})'$. We consider the (nested) multiple linear regression models

- $\mathbf{Y} = B_k \boldsymbol{\beta}^{(k)} + \boldsymbol{\varepsilon}$, $k = 1, \ldots, m+1$

Then, $\operatorname{Im}(B_1) \subsetneq \operatorname{Im}(B_2) \subsetneq \cdots \subsetneq \operatorname{Im}(B_{m+1}) = \operatorname{Im}(B)$. Hence, the corresponding orthogonal projectors Q_j satisfy the relations

- $Q_j Q_{j-1} = Q_{j-1} Q_j = Q_{j-1}$, $j = 2, \ldots, m+1$, where $Q_{m+1} = Q$.

Decomposition of multiple linear regression model

Therefore, for the random vectors $Q_j\mathbf{Y}$ we have $Q_j\mathbf{Y} \sim N_n(Q_jB\boldsymbol{\beta}, \sigma^2 Q_j)$, $j = 1, \dots, m+1$, and we get the decomposition

$$\begin{aligned}
\mathsf{SST} = \mathbf{Y}'E_n\mathbf{Y} &= \mathbf{Y}'(I_n - Q_{j+1})\mathbf{Y} + \mathbf{Y}'(Q_{j+1} - Q_j)\mathbf{Y} + \mathbf{Y}'(Q_j - \tfrac{1}{n}\mathbb{1}_{n\times n})\mathbf{Y} \\
&= \mathbf{Y}'(I_n - Q_{j+1})\mathbf{Y} + \mathbf{Y}'(Q_{j+1} - Q_j)\mathbf{Y} + \mathbf{Y}'(Q_j - Q_1)\mathbf{Y} \\
&= \mathbf{Y}'(I_n - Q_{j+1})\mathbf{Y} + \sum_{i=1}^{j} \mathbf{Y}'(Q_{i+1} - Q_i)\mathbf{Y} \\
&= \mathsf{SSE}_{j+1} + \sum_{i=1}^{j} \mathsf{SSR}(\beta_{i+1} \mid \beta_1, \dots, \beta_i).
\end{aligned}$$

Thus, the 'full' model $\mathbf{Y} = B_{j+1}\boldsymbol{\beta}^{(j+1)} + \boldsymbol{\varepsilon}$ and the 'reduced' model $\mathbf{Y} = B_j\boldsymbol{\beta}^{(j)} + \boldsymbol{\varepsilon}$ can be compared by an F-test as in Theorem I.5.17 and I.4.40 using the (independent) statistics

$$\mathbf{Y}'(I_n - Q_{j+1})\mathbf{Y}, \quad \mathbf{Y}'(Q_{j+1} - Q_j)\mathbf{Y}.$$

The distributions of the quadratic forms as well as the ratios (under H_0 or H_1) can be taken from Theorem I.4.39.

Testing general linear hypothesis

I.5.19 Assumption

For simplicity, we assume in the following that $\operatorname{rank}(B) = m + 1 < n$ which ensures, e.g., a unique LSE for β and the existence of the inverse matrix $(B'B)^{-1}$.

Furthermore, we assume that $K \in \mathbb{R}^{q\times(m+1)}$ satisfies $\operatorname{rank}(K) = q \leqslant m + 1$.

I.5.20 Problem

We consider the testing problem (I.7) with $\delta = 0$, that ist,

$$\boxed{H_0 : K\beta = 0 \quad \longleftrightarrow \quad H_1 : K\beta \neq 0} \tag{I.11}$$

Notice that, by choosing $K = [0_{m_2\times m_1} \mid I_{m_2}]$, the null hypothesis $H_0 : \beta_2 = 0$ results.

Testing general linear hypothesis

I.5.21 Theorem

Consider the NoLM $Y = B\boldsymbol{\beta} + \varepsilon$ with $\boldsymbol{\beta} \in \Theta = \mathbb{R}^{m+1}$ and $\sigma^2 > 0$ unknown. Suppose B has full rank $m + 1 \leqslant n$ and let $\operatorname{rank}(K) = q \leqslant m + 1$. Then:

1. $\widehat{\boldsymbol{\beta}}_K = K\widehat{\boldsymbol{\beta}} \sim N_q(K\boldsymbol{\beta}, \sigma^2 K(B'B)^1 K')$.
2. $\mathrm{SSR}/\sigma^2 = \widehat{\boldsymbol{\beta}}_K'(K(B'B)^{-1}K')^{-1}\widehat{\boldsymbol{\beta}}_K/\sigma^2 \sim \chi^2(q, \lambda)$ with non-centrality parameter $\lambda = \boldsymbol{\beta}'K'(K(B'B)^{-1}K')^{-1}K\boldsymbol{\beta}/(2\sigma^2)$
3. $\mathrm{SSE}/\sigma^2 = \mathbf{Y}'(I_n - B(B'B)^{-1}B')\mathbf{Y}/\sigma^2 \sim \chi^2(n - m - 1)$.
4. SSR and SSE are independent.
5. Let $F = \dfrac{\mathrm{SSR}/q}{\mathrm{SSE}/(n - m - 1)}$. Then
   - If $H_0 : K\boldsymbol{\beta} = 0$ is false then $F \sim F(q, n - m - 1, \lambda)$ with λ as in 2.
   - If $H_0 : K\boldsymbol{\beta} = 0$ is true then $F \sim F(q, n - m - 1)$

Testing general linear hypothesis

Using Theorem I.4.39, we can construct a test procedure for the testing problem (I.11).

I.5.22 Testing procedure (F-test)

Given the situation in Theorem I.5.21 and the testing problem (I.11). Then, an α-level statistical test for H_0 is given by the decision rule

$$\text{Reject } H_0 \text{ if } F = \frac{SSR/q}{SSE/(n-m-1)} > F_{1-\alpha}(q, n-m-1)$$

where $F_{1-\alpha}(1, n-m-1)$ denotes the $(1-\alpha)$-quantile of the $F(q, n-m-1)$-distribution.

I.5.23 Remark

- The approach can also be applied to an arbitrary LM.
- Special cases for K as in (I.11) are given by, e.g.,
  - $H_0 : \beta_k = 0$: $K = e'_{k,m+1}$.
  - $H_0 : \beta_k = 0, k \in S \subseteq \{0, \ldots, m\}$: $K = [e'_{k,m+1}, k \in S]$.
  - $H_0 : \beta_0 = \beta_1$: $K = e'_{1,m+1} - e'_{2,m+1}$

Testing general linear hypothesis

I.5.24 Remark

The F-test in Procedure I.5.22 for the general linear hypothesis

$$H_0 : K\boldsymbol{\beta} = 0$$

is a full-reduced-model test, that is, the reduced model

$$\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{with } K\boldsymbol{\beta} = 0$$

is considered.

Using Lagrange multipliers, it can be shown that the LSE $\widehat{\boldsymbol{\beta}}_{K,*}$ of $\boldsymbol{\beta}$ subject to the constraint $K\boldsymbol{\beta} = 0$ is given by

$$\widehat{\boldsymbol{\beta}}_{K,*} = \left(I_{m+1} - (B'B)^{-1}K'(K(B'B)^{-1}K')^{-1}K\right)\widehat{\boldsymbol{\beta}}$$

where $\widehat{\boldsymbol{\beta}}$ is the LSE in the full model (cf. Rencher, Schaalje 2008).

Testing $H_0 : K\beta = \delta$

The testing problem (I.7) with arbitrary δ can also be handled.

I.5.25 Theorem

Consider the NoLM $Y = B\beta + \varepsilon$ with $\beta \in \Theta = \mathbb{R}^{m+1}$ and $\sigma^2 > 0$ unknown. Suppose B has full rank $m + 1 \leqslant n$ and let $\operatorname{rank}(K) = q \leqslant m + 1$. Then:

1. $\widehat{\beta}_K - \delta = K\widehat{\beta} - \delta \sim N_q(K\beta - \delta, \sigma^2 K(B'B)^1 K')$
2. $SSR/\sigma^2 = (\widehat{\beta}_K - \delta)'(K(B'B)^{-1}K')^{-1}(\widehat{\beta}_K - \delta)/\sigma^2 \sim \chi^2(q, \lambda)$ with non-centrality parameter $\lambda = (K\beta - \delta)'(K(B'B)^{-1}K')^{-1}(K\beta - \delta)/(2\sigma^2)$
3. $SSE/\sigma^2 = \mathbf{Y}'(I_n - B(B'B)^{-1}B')\mathbf{Y}/\sigma^2 \sim \chi^2(n - m - 1)$.
4. SSR and SSE are independent.
5. Let $F = \dfrac{SSR/q}{SSE/(n - m - 1)}$. Then
   - If $H_0 : K\beta = \delta$ is false then $F \sim F(q, n - m - 1, \lambda)$ with λ as in 2.
   - If $H_0 : K\beta = \delta$ is true then $F \sim F(q, n - m - 1)$

I.5.26 Remark

Using the results of Theorem I.5.25, a test for the testing problem (I.7) can be constructed as in Procedure I.5.22.

Confidence region for β

The results of Theorem I.5.25 can be applied to construct a confidence region for β.

I.5.27 Theorem

Consider the NoLM $Y = B\beta + \varepsilon$ with $\beta \in \Theta = \mathbb{R}^{m+1}$ and $\sigma^2 > 0$ unknown. Then, the (random) ellipsoid

$$\Big\{\beta \in \mathbb{R}^{m+1} \;\Big|\; \frac{\|B(\widehat{\beta} - \beta)\|^2/(m+1)}{\|Y - B\widehat{\beta}\|^2/(n-m-1)} \leqslant F_{1-\alpha}(m+1, n-m-1)\Big\}$$

forms a $(1-\alpha)$-confidence region for the parameter β.

Confidence intervals for a single component β_k, $k \in \{0, \ldots, m\}$, are obtained from the property

$$\frac{1}{\sqrt{(B'B)^{-1}_{kk}}} \cdot \frac{\hat{\beta}_k - \beta_k}{\sqrt{\|Y - B\widehat{\beta}\|^2/(n-m-1)}} \sim t(n-m-1).$$

I.5.28 Remark

Similarly, one can obtain confidence intervals for linear combinations $c'\beta$. Choosing $c' = (1, x')$, one gets a confidence interval for the regression function at x (see also Example I.5.2 for the regression function in a simple linear regression).

Part I: Linear Models

Chapter I.5

Regression Models – Model Validation & Diagnostics

Topics

To be discussed...

- Residuals
- Hat matrix
- Outliers
- Influential observations and leverage

Variable selection

Consider a multiple linear regression model with m explanatory variables as in Definition I.5.3 and regression function as in (I.5):

$$f(x_1, \dots, x_m) = \beta_0 + \sum_{i=1}^{m} \beta_i x_i.$$

I.5.29 Remark

Problem

- Which explanatory variables $x_1, \dots, x_m$ are important for a regression on y?
- Which model fits best the data?

- Notice that there are 2^m different models so that a complete analysis is not adequate in most cases!
- Thus, we need methods to compare the fit of the models in order to identify (a set of) appropriate models which will be considered for a more detailed analysis.
- In the following, we present some ideas on model diagnostics. Further details are provided in Part II of the lecture (see also Christensen (2011, Chapter 14), James et al. (2013, Chapter 6)).

Residuals

I.5.30 Definition

Consider a LM $Y = B\boldsymbol{\beta} + \varepsilon$ as in Definition I.4.2 with $\boldsymbol{\beta} \in \Theta = \mathbb{R}^{m+1}$ and $\sigma^2 > 0$ unknown. Let $\widehat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$. Then

- $\widehat{\mathbf{Y}} = B\widehat{\boldsymbol{\beta}}$ is called vector of **predicted values**.
- $\widehat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - B\widehat{\boldsymbol{\beta}}$ is called the **residual vector**.
- $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$ are called **residuals**.

I.5.31 Theorem

Consider a LM $Y = B\boldsymbol{\beta} + \varepsilon$ as in Definition I.4.2 with $\boldsymbol{\beta} \in \Theta = \mathbb{R}^{m+1}$ and $\sigma^2 > 0$ unknown. Suppose B has full rank $m + 1 \leqslant n$. Furthermore, let $\widehat{\boldsymbol{\beta}}$ be the LSE of $\boldsymbol{\beta}$. Then,

1. $\widehat{\mathbf{Y}} = B\widehat{\boldsymbol{\beta}} = Q\mathbf{Y}$ with $Q = B(B'B)^{-1}B'$.
2. $\widehat{\boldsymbol{\varepsilon}} = (I_n - Q)\mathbf{Y} = (I_n - Q)\boldsymbol{\varepsilon}$.
3. $\widehat{\boldsymbol{\varepsilon}}'\mathbf{Y} = \mathbf{Y}'(I_n - Q)\mathbf{Y}$, $\widehat{\boldsymbol{\varepsilon}}'\widehat{\mathbf{Y}} = 0$, $B'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}$, $\overline{\widehat{\boldsymbol{\varepsilon}}} = \frac{1}{n}\mathbb{1}_n'\widehat{\boldsymbol{\varepsilon}} = 0$.
4. $E\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}$, Cov $(\widehat{\boldsymbol{\varepsilon}}) = \sigma^2(I_n - Q)$.
5. Cov $(\widehat{\mathbf{Y}}, \widehat{\boldsymbol{\varepsilon}}) = \mathbf{0}_{n\times n}$, Cov $(\mathbf{Y}, \widehat{\boldsymbol{\varepsilon}}) = \sigma^2(I_n - Q)$.
6. Given a NoLM, $\widehat{\boldsymbol{\varepsilon}} \sim N_n(\mathbf{0}, \sigma^2(I_n - Q))$.

Hat matrix

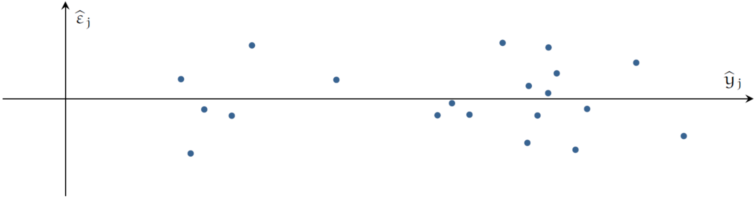
I.5.32 Remark

- $Q = B(B'B)^{-1}B'$ is called the **hat matrix** since it transforms the observed values $\mathbf{Y}$ to the predicted values $\widehat{\mathbf{Y}} = Q\mathbf{Y}$.
- Notice that Q is an orthogonal projector, that is, $Q^2 = Q, Q' = Q$.
- The sample correlation
$$r_{\widehat{\varepsilon},\mathbf{Y}} = \frac{\widehat{\varepsilon}'(\mathbf{Y} - \overline{\mathbf{Y}}\mathbb{1}_n)}{\sqrt{\|\widehat{\varepsilon}\|^2\|\mathbf{Y} - \overline{\mathbf{Y}}\mathbb{1}_n\|^2}} = \frac{\widehat{\varepsilon}'\mathbf{Y}}{\sqrt{\|\widehat{\varepsilon}\|^2\|\mathbf{Y} - \overline{\mathbf{Y}}\mathbb{1}_n\|^2}}$$
is (mostly) positive since $\widehat{\varepsilon}'\mathbf{Y} = \|(I_n - Q)\mathbf{Y}\|^2$.
- The sample correlation
$$r_{\widehat{\varepsilon},\widehat{\mathbf{Y}}} = \frac{\widehat{\varepsilon}'(\widehat{\mathbf{Y}} - \overline{\widehat{\mathbf{Y}}}\mathbb{1}_n)}{\sqrt{\|\widehat{\varepsilon}\|^2\|\widehat{\mathbf{Y}} - \overline{\widehat{\mathbf{Y}}}\mathbb{1}_n\|^2}} = \frac{\widehat{\varepsilon}'\widehat{\mathbf{Y}}}{\sqrt{\|\widehat{\varepsilon}\|^2\|\widehat{\mathbf{Y}} - \overline{\widehat{\mathbf{Y}}}\mathbb{1}_n\|^2}}$$
is equal to zero, that is, $r_{\widehat{\varepsilon},\widehat{\mathbf{Y}}} = 0$ by Theorem I.5.31 ❸. Therefore, $\widehat{\varepsilon}$ and $\widehat{\mathbf{Y}}$ are empirically uncorrelated.

Residual plot

I.5.33 Remark

- If the model and attendant assumptions are correct, then a plot of the residuals versus the predicted values, $(\widehat{y}_j, \widehat{\varepsilon}_j)$, $j = 1, \ldots, n$, should show no systematic pattern.
- This **residual plot** is therefore useful for checking the model. A typical residual plot is shown below.



Centered form of the multiple linear regression model

I.5.34 Definition

Consider the multiple linear regression model as in Definition I.5.3, that is, $\mathbf{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with

$$B = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} = [\mathbb{1}_n \mid B_1] \in \mathbb{R}^{n\times(m+1)}$$

The **centered form** of the multiple linear regression model is given by

$$\mathbf{Y} = \alpha\mathbb{1}_n + B_c\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$$

where $\alpha = \beta_0 + \sum_{j=1}^m \beta_j\overline{x}_j$ and $\gamma_j = \beta_j$, $\overline{x}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}$, $j = 1, \dots, m$, and the design matrix is given by $B_c = E_nB_1$.

Centered form of the multiple linear regression model

I.5.35 Remark

Notice that the representations $Y = B\boldsymbol{\beta} + \varepsilon$ and $\mathbf{Y} = \alpha\mathbb{1}_n + B_c\boldsymbol{\gamma} + \varepsilon$ of the LM are equivalent since $\text{Im}(B) = \text{Im}([\mathbb{1}_n \mid B_c])$. Thus,

- $\widehat{\alpha} = \overline{Y}$, $\widehat{\gamma} = (B_c'B_c)^{-1}B_c'\mathbf{Y}$
- $\widehat{\mathbf{Y}} = \widehat{\alpha}\mathbb{1}_n + B_c\widehat{\boldsymbol{\gamma}}$ with corresponding hat matrix $Q_c = B_c(B_c'B_c)^{-1}B_c'$
- ☞ $\widehat{\mathbf{Y}} = \widehat{\alpha}\mathbb{1}_n + B_c(B_c'B_c)^{-1}B_c'\mathbf{Y} = \frac{1}{n}\mathbb{1}_{n\times n}\mathbf{Y} + Q_c\mathbf{Y} = \left(\frac{1}{n}\mathbb{1}_{n\times n}Y + Q_c\right)\mathbf{Y}$
- ☞ $Q = \frac{1}{n}\mathbb{1}_{n\times n} + Q_c$

I.5.36 Theorem

Let B satisfy $\text{rank}(B) = m + 1 \leqslant n$. The hat matrix $Q = (q_{ij})_{i,j} = B(B'B)^{-1}B'$ has the following properties:

1. $q_{ii} = \frac{1}{n} + (Q_c)_{ii}$, $i = 1, \ldots, n$
2. $\frac{1}{n} \leqslant q_{ii} \leqslant 1$, $i = 1, \ldots, n$,
3. $-\frac{1}{2} \leqslant q_{ij} \leqslant \frac{1}{2}$ for all $i \neq j$
4. $\text{trace}(Q) = \sum_{i=1}^{n} q_{ii} = m + 1$

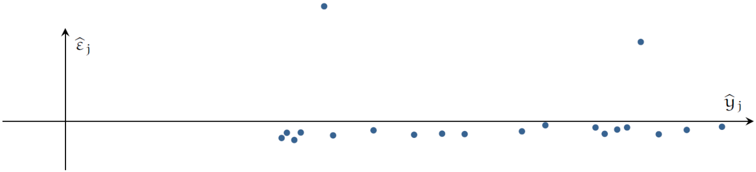
Outliers

I.5.37 Remark

In some cases, the model appears to be correct for most of the data, but one residual is much larger (in absolute value) than the others. Such an outlier

- may be due to an error in recording
- may be from another population
- may simply be an unusual observation from the assumed distribution.

One approach to check for outliers is the residual plot.



Notice that the variances of the residuals $\operatorname{Var}(\widehat{\varepsilon}_j) = \sigma^2(1 - q_{jj})$ are not constant. In particular, they will be small if $\frac{1}{n} \leqslant q_{jj} \leqslant 1$ is close to 1. Thus, an outlier may be masked due to the choice of the measurement points x (see Rencher, Schaalje 2008)!

Outliers

I.5.38 Remark

In order to take into account the variances of the residuals, scaling is an option

1. Consider $\widehat{\varepsilon}_j/(\sigma\sqrt{1-q_{jj}})$ and the **studentized residuals**

$$\widehat{r}_j = \frac{\widehat{\varepsilon}_j}{\sqrt{(1-q_{jj})\mathsf{SSE}/(n-m-1)}} \tag{I.12}$$

   where $\mathsf{SSE} = \|\mathbf{Y} - B\widehat{\boldsymbol{\beta}}\|^2$. Consider the residual plot $(\widehat{\mathbf{y}}_j, \widehat{r}_j)$, $j = 1, \ldots, n$. This eliminates the location effect due to q_{jj}.

2. A variation of this studentized residual excludes the j-th observation from the estimation of the variance. That is, use $\mathsf{SSE}_{[j]} = \mathbf{Y}_{[j]}'(I_{n-1} - B_{[j]}(B_{[j]}'B_{[j]})^{-1}B_{[j]}')\mathbf{Y}_{[j]}$ instead of SSE where [j] indicates the deletion of the respective components:

$$\widehat{r}_j^* = \frac{\widehat{\varepsilon}_j}{\sqrt{(1-q_{jj})\mathsf{SSE}_{[j]}/(n-m-2)}}$$

Further options can be found, e.g., in Rencher, Schaalje (2008).

Influential observation and leverage

I.5.39 Remark

Previously, we searched for outliers that do not fit the model. An observation is called **influential** if its deletion has a substantial effect on the estimates $\widehat{\boldsymbol{\beta}}$ and $B\widehat{\boldsymbol{\beta}}$.

The influence of an observation is measured as follows. From $\widehat{\mathbf{y}} = Q\mathbf{y}$, we get

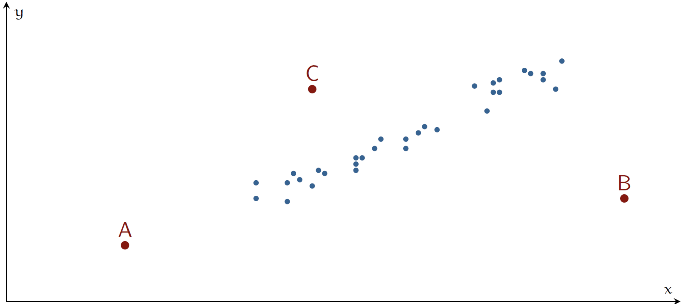
$$\widehat{y}_i = \sum_{j=1}^{n} q_{ij} y_j = q_{ii} y_i + \sum_{j \neq i}^{n} q_{ij} y_j \quad i = 1, \dots, n.$$

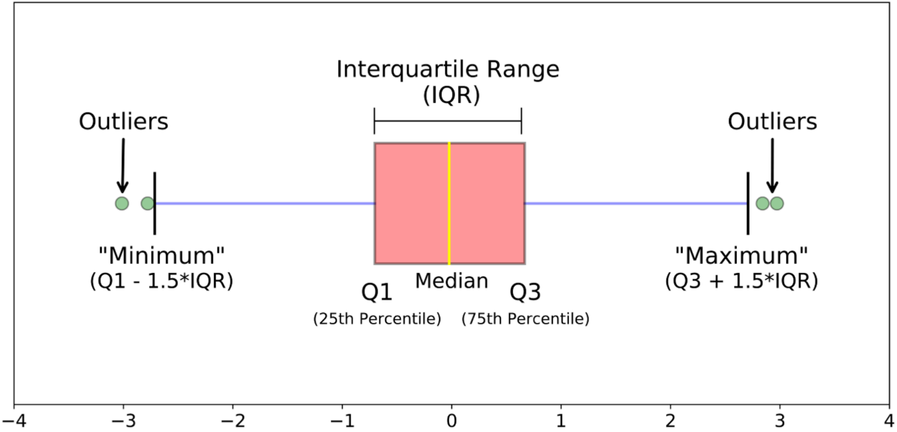
- If q_{ii} is close to 1, then q_{ij}, $j \neq i$, are small and y_i contributes much more to the $\widehat{y}_i$ than the other y's. Recall from the proof of Theorem I.5.36 that $1 = q_{ii} + \sum_{j \neq i} \frac{q_{ij}^2}{q_{ii}}$.
- Thus, q_{ii} is called the **leverage** of y_i.
- Points with high leverage have high potential for influencing regression results. In general, if an observation $(y_i, \widetilde{\mathbf{x}}_i)$ has a value of q_{ii} near 1, then the estimated regression equation will be close to y_i; that is, $\widehat{\varepsilon}_i = y_i - \widehat{y}_i$ will be small.

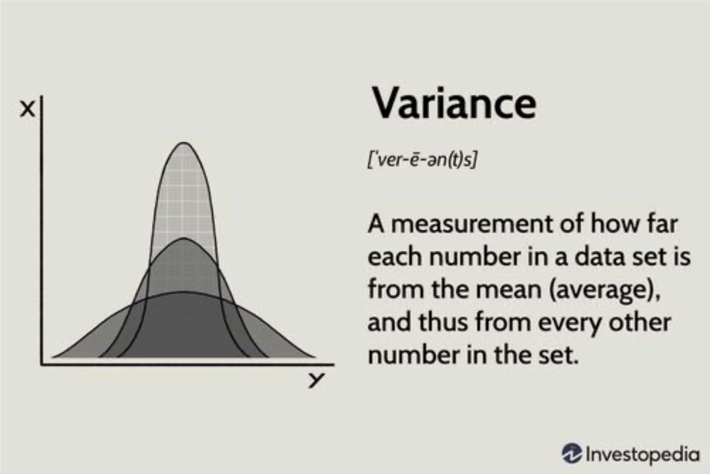
Influential observation and leverage

I.5.40 Example

- Point A, B may be considered as outliers in x-direction.
- Point B, C may be considered as outliers in y-direction.
- Point A may reduce the variance of the estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$.
- Point B will drastically alter the fitted regression line.




Interquartile Range
(IQR)
Outliers
Outliers
"Minimum"
(Q1 - 1.5*IQR)
Q1
(25th Percentile)
Median
Q3
(75th Percentile)
"Maximum"
(Q3 + 1.5*IQR)
−4
−3
−2
−1
0
1
2
3
4



Variance

['ver-ē-ən(t)s]

A measurement of how far each number in a data set is from the mean (average), and thus from every other number in the set.

Investopedia

Influential observation and leverage

- The influence of an observation $(y_i, \widetilde{x}_i)$ is measured formally by its deletion from the computation of the estimator:

$$\widehat{\beta}_{[i]} = (B'_{[i]}B_{[i]})^{-1}B'_{[i]})\mathbf{Y}_{[i]}, \quad i = 1, \ldots, n.$$

I.5.41 Definition

Consider a multiple linear regression model as in Definition I.5.3, that is, $\mathbf{Y} = B\beta + \varepsilon$ as well as the respective model by deleting the i-th observation, that is, $\mathbf{Y}_{[i]} = B_{[i]}\beta_{[i]} + \varepsilon_{[i]}$. Let $\widehat{\beta}$ and $\widehat{\beta}_{[i]}$ be the correspondings LSEs.

For $i \in \{1, \ldots, n\}$, **Cook's distance** is defined by

$$D_i = \frac{(\widehat{\beta}_{[i]} - \widehat{\beta})'B'B(\widehat{\beta}_{[i]} - \widehat{\beta})}{(m+1)\text{SSE}/(n-m-1)} = \frac{\|B(\widehat{\beta}_{[i]} - \widehat{\beta})\|^2}{(m+1)\text{SSE}/(n-m-1)} = \frac{\|\widehat{y}_{[i]} - \widehat{y}\|^2}{(m+1)\text{SSE}/(n-m-1)}$$

with $\text{SSE} = \mathbf{Y}'(I_n - B(B'B)^{-1}B')\mathbf{Y}$

I.5.42 Remark

D_i can be written as $D_i = \dfrac{\widehat{r}_i^2}{m+1} \cdot \dfrac{q_{ii}}{1-q_{ii}}$ with studentized residual $\widehat{r}_i$ as in (I.12).

R^2 and adjusted R^2

I.5.43 Definition

For a given data set and the associated TSS value, the larger the value of SSR, the more effective the explanatory variables are in 'explaining' the response variable. A summary of this predictive power of the underlying model is

$$R^2 = \frac{SSR}{TSS} = \frac{TSS - SSE}{TSS},$$

known as *coefficient of determination*.

I.5.44 Properties

- R^2 measures the proportional reduction in error due to the regressors ($0 \leqslant R^2 \leqslant 1$).
- The sample correlation $r_{Y,\hat{Y}}$ is equal to $r_{Y,\hat{Y}} = +\sqrt{R^2} = R$, which is called *multiple correlation* ($0 \leqslant R \leqslant 1$). When $m = 1$, then $R = r_{Y,x_1}$.

I.5.45 Remark

When n is small and a model has several explanatory variables, R^2 is biased (overestimates the corresponding population value). For this, an *adjusted* R^2 has been proposed:

$$R_a^2 = \frac{(n-1)R^2 - m}{n - m - 1}.$$