Statistica II (750AA)

Lezione 7

Dario Trevisan – *https://web.dm.unipi.it/trevisan*

3/11/2025

Regressione lineare multipla

Richiami della lezione precedente

- Regressione come previsione di un fattore di uscita y a partire da uno di ingresso x

Richiami della lezione precedente

- Regressione come previsione di un fattore di uscita y a partire da uno di ingresso x
- **Indicatori di performance:** MSE, RMSE, MAP, R^2

Richiami della lezione precedente

- Regressione come previsione di un fattore di uscita y a partire da uno di ingresso x
- **Indicatori di performance:** MSE, RMSE, MAP, R^2
- **k-Nearest Neighbors (k-NN)** per la regressione $\rightarrow$ non parametrico

Richiami della lezione precedente

- Regressione come previsione di un fattore di uscita y a partire da uno di ingresso x
- **Indicatori di performance:** MSE, RMSE, MAP, R^2
- **k-Nearest Neighbors (k-NN)** per la regressione $\rightarrow$ non parametrico
- **Modelli parametrici:** $h(x) = h(x; \theta)$

Richiami della lezione precedente

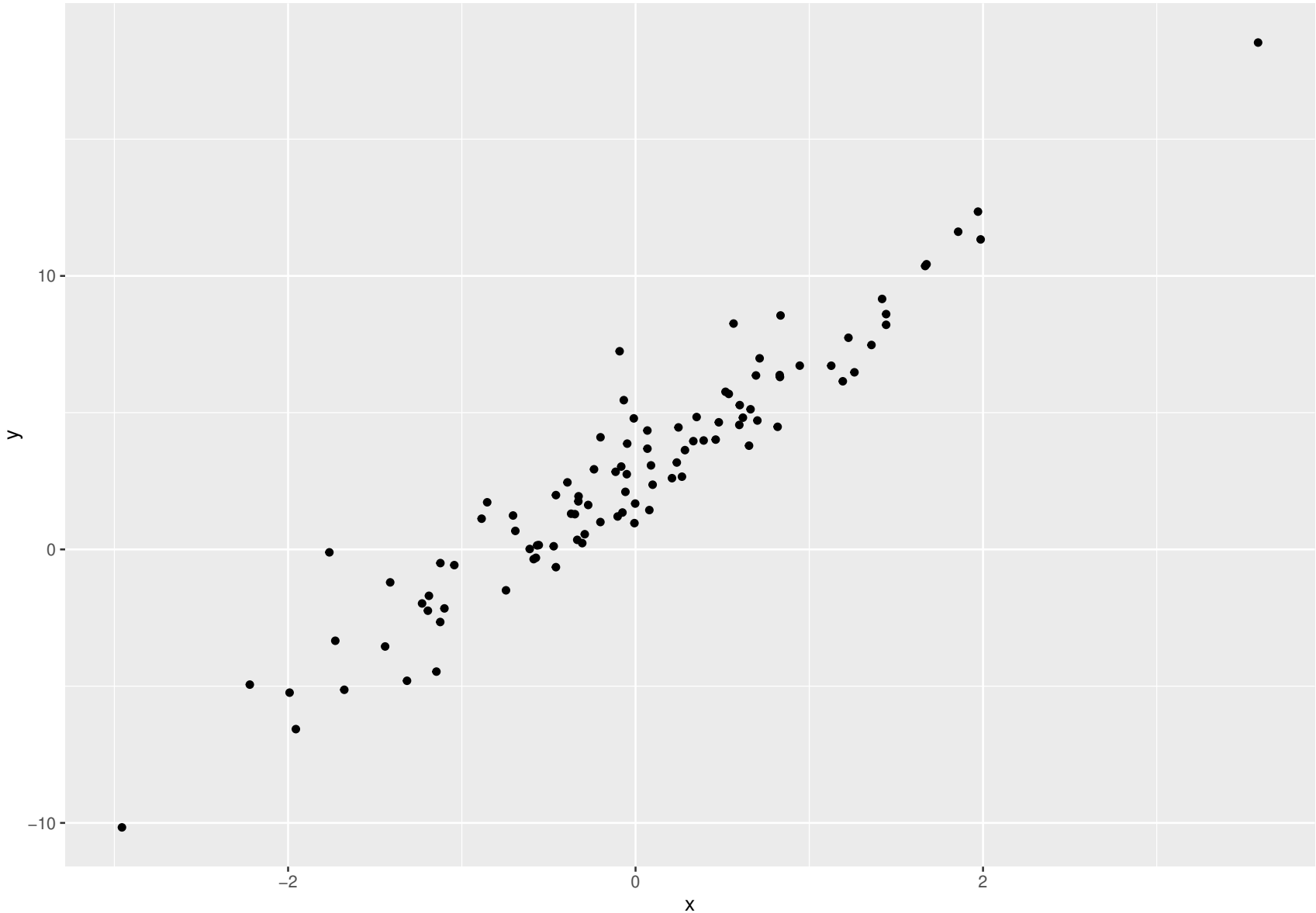
- Regressione come previsione di un fattore di uscita y a partire da uno di ingresso x
- **Indicatori di performance:** MSE, RMSE, MAP, R^2
- **k-Nearest Neighbors (k-NN)** per la regressione $\rightarrow$ non parametrico
- **Modelli parametrici:** $h(x) = h(x; \theta)$
- **Minimi quadrati (OLS)**: $\theta_{OLS} \in \arg\min_\theta \sum_i (y_i - h(x_i; \theta))^2$

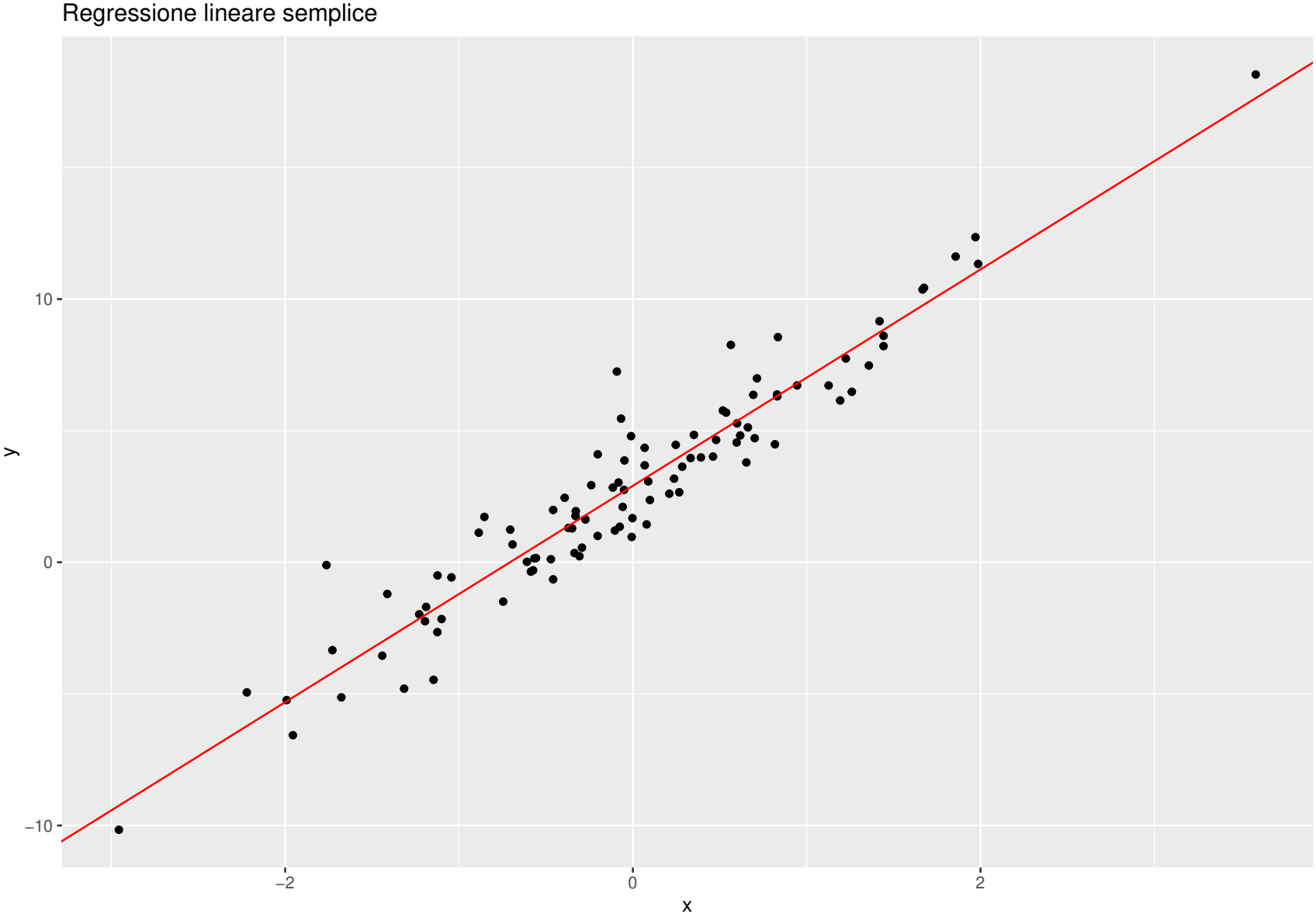
Richiami della lezione precedente

- Regressione come previsione di un fattore di uscita y a partire da uno di ingresso x
- **Indicatori di performance:** MSE, RMSE, MAP, R^2
- **k-Nearest Neighbors (k-NN)** per la regressione $\rightarrow$ non parametrico
- **Modelli parametrici:** $h(x) = h(x; \theta)$
- **Minimi quadrati (OLS)**: $\theta_{OLS} \in \arg\min_\theta \sum_i (y_i - h(x_i; \theta))^2$
- **Stima OLS per la regressione semplice:** $h(x) = \beta_0 + \beta_1 x$

$$\beta_{OLS,1} = \mathrm{cor}(x, y)\frac{\mathrm{sd}(y)}{\mathrm{sd}(x)}, \quad \beta_{OLS,0} = \bar{y} - \beta_{OLS,1}\bar{x}$$

Un esempio (generato)




Regressione lineare semplice
10
0
−10
y
−2
0
2
x

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

- Modello $h(x;\beta) = \Phi(x)\beta$

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

- Modello $h(x;\beta) = \Phi(x)\beta$
- Residuo $\varepsilon := y - h(x;\beta)$

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

- Modello $h(x;\beta) = \Phi(x)\beta$
- Residuo $\varepsilon := y - h(x;\beta)$
- **Esempi:**

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

- Modello $h(x;\beta) = \Phi(x)\beta$
- Residuo $\varepsilon := y - h(x;\beta)$
- **Esempi:**
  - lineare semplice, $\Phi(x) = (1, x)$, $x \in \mathbb{R}$, $\rightarrow p = 2$

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

- Modello $h(x;\beta) = \Phi(x)\beta$
- Residuo $\varepsilon := y - h(x;\beta)$
- **Esempi:**
  - lineare semplice, $\Phi(x) = (1, x)$, $x \in \mathbb{R}$, $\rightarrow p = 2$
  - polinomiale: $\Phi(x) = (1, x, x^2, \ldots, x^q)$, $x \in \mathbb{R}$, $p = q + 1$,

Modello lineare

Forma più generale di modello lineare

$$h(x;\beta) = \phi_1(x)\beta_1 + \phi_2(x)\beta_2 + \cdots + \phi_p(x)\beta_p$$

$$= \phi(x) \cdot \beta$$

- Scrittura **compatta**:

$$\Phi(x) = (\phi_i(x))_{i=1\ldots,p} \in \mathbb{R}^p \quad \text{(vettore riga)}$$
$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p \quad \text{(vettore colonna)}$$

- Modello $h(x;\beta) = \Phi(x)\beta$
- Residuo $\varepsilon := y - h(x;\beta)$
- **Esempi:**
  - lineare semplice, $\Phi(x) = (1, x)$, $x \in \mathbb{R}$, $\rightarrow p = 2$
  - polinomiale: $\Phi(x) = (1, x, x^2, \ldots, x^q)$, $x \in \mathbb{R}$, $p = q + 1$,
  - multipla lineare $\Phi(x) = (1, x_1, \ldots, x_d)$, $p = d + 1$.

Notazione matriciale

- Training set $\{(x_i, y_i)\}_{i=1,\dots n}$: notazione *matriciale*

$$\boldsymbol{y} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \underbrace{\mathbb{R}^n} \quad \Phi(\boldsymbol{x}) := \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_p(x_1) \\ \vdots & \vdots & & \vdots \\ \phi_1(x_n) & \phi_2(x_n) & \dots & \phi_p(x_n) \end{bmatrix} \in \underbrace{\mathbb{R}^{n \times p}}$$

Notazione matriciale

- Training set $\{(x_i, y_i)\}_{i=1,\dots n}$: notazione *matriciale*

$$\boldsymbol{y} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n \quad \Phi(\boldsymbol{x}) := \begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_p(x_1) \\ \vdots & \vdots & & \vdots \\ \phi_1(x_n) & \phi_2(x_n) & \dots & \phi_p(x_n) \end{bmatrix} \in \mathbb{R}^{n \times p}$$

- Vettore dei **residui** (sul training set):

$$\varepsilon := \boldsymbol{y} - \Phi(\boldsymbol{x})\beta$$

Metodo dei minimi quadrati

- **Obiettivo**: minimizzare la somma dei quadrati dei residui

$$\beta_{OLS} \in \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \Phi(x_i)\beta)^2 \cdot \frac{1}{n}$$

Metodo dei minimi quadrati

- **Obiettivo**: minimizzare la somma dei quadrati dei residui

$$\beta_{OLS} \in \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \Phi(x_i)\beta)^2$$

- **Formula esplicita**: matrice di **Gram** $G := \Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x}) \in \mathbb{R}^{p \times p}$

$$\beta_{OLS} = G^{-1} \Phi(\boldsymbol{x})^T \boldsymbol{y}$$

$G^{-1} \in \mathbb{R}^{p \times p}$, $\Phi(\boldsymbol{x})^T \in \mathbb{R}^{p \times n}$, $\boldsymbol{y} \in \mathbb{R}^n$

Metodo dei minimi quadrati

- **Obiettivo**: minimizzare la somma dei quadrati dei residui

$$\beta_{OLS} \in \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \Phi(x_i)\beta)^2$$

- **Formula esplicita**: matrice di **Gram** $G := \Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x})$

$$\beta_{OLS} = G^{-1} \Phi(\boldsymbol{x})^T \boldsymbol{y}$$

- **Osservazione:** G *non* è invertibile $\leftrightarrow$ $p > n$ oppure le colonne di $\Phi(\boldsymbol{x})$ sono linearmente dipendenti (collinearità)

Metodo dei minimi quadrati

- **Obiettivo**: minimizzare la somma dei quadrati dei residui

$$\beta_{OLS} \in \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \Phi(x_i)\beta)^2$$

- **Formula esplicita**: matrice di **Gram** $G := \Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x})$

$$\beta_{OLS} = G^{-1} \Phi(\boldsymbol{x})^T \boldsymbol{y}$$

- **Osservazione:** G *non* è invertibile $\leftrightarrow$ $p > n$ oppure le colonne di $\Phi(\boldsymbol{x})$ sono linearmente dipendenti (collinearità)
- Se G non è invertibile $\rightarrow$ metodi di regolarizzazione (prossima lezione).

Dimostrazione

Notazione semplificata: $\Phi(\boldsymbol{x}) = \boldsymbol{x}$, $G = \boldsymbol{x}^T\boldsymbol{x}$.

1. Sviluppiamo (in notazione matriciale)

$$n\,\mathrm{MSE}(h(;\beta)) = \sum_{i=1}^{n}(y_i - x_i\beta)^2 = \|\boldsymbol{y} - \boldsymbol{x}\beta\|^2 = (\boldsymbol{y} - \boldsymbol{x}\beta)^T(\boldsymbol{y} - \boldsymbol{x}\beta)$$

$$= \|\boldsymbol{y}\|^2 + \beta^T\boldsymbol{x}^T\boldsymbol{x}\beta - \beta^T\boldsymbol{x}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{x}\beta$$

Dimostrazione

Notazione semplificata: $\Phi(\boldsymbol{x}) = \boldsymbol{x}$, $G = \boldsymbol{x}^T\boldsymbol{x}$.

1. Sviluppiamo (in notazione matriciale)

$$\sum_{i=1}^{n}(y_i - x_i\beta)^2 = \|\boldsymbol{y} - \boldsymbol{x}\beta\|^2 = (\boldsymbol{y} - \boldsymbol{x}\beta)^T(\boldsymbol{y} - \boldsymbol{x}\beta)$$

$$= \|\boldsymbol{y}\|^2 + \beta^T \underbrace{\boldsymbol{x}^T\boldsymbol{x}}_{G}\beta - \beta^T\boldsymbol{x}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{x}\beta$$

2. **Completiamo il quadrato** con $\boldsymbol{y}^T\boldsymbol{x}G^{-1}\boldsymbol{x}^T\boldsymbol{y}$:

$$\sum_{i=1}^{n}(y_i - x_i\beta)^2$$

$$= \|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\boldsymbol{x}G^{-1}\boldsymbol{x}^T\boldsymbol{y} + (\beta - G^{-1}\boldsymbol{x}^T\boldsymbol{y})^T G(\beta - G^{-1}\boldsymbol{x}^T\boldsymbol{y})$$

quadrato ≥ 0

Dimostrazione

Notazione semplificata: $\Phi(\boldsymbol{x}) = \boldsymbol{x}$, $G = \boldsymbol{x}^T\boldsymbol{x}$.

1. Sviluppiamo (in notazione matriciale)

$$\sum_{i=1}^{n}(y_i - x_i\beta)^2 = \|\boldsymbol{y} - \boldsymbol{x}\beta\|^2 = (\boldsymbol{y} - \boldsymbol{x}\beta)^T(\boldsymbol{y} - \boldsymbol{x}\beta)$$

$$= \|\boldsymbol{y}\|^2 + \beta^T\boldsymbol{x}^T\boldsymbol{x}\beta - \beta^T\boldsymbol{x}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{x}\beta$$

2. **Completiamo il quadrato** con $\boldsymbol{y}^T\boldsymbol{x}G^{-1}\boldsymbol{x}^T\boldsymbol{y}$:

$$\sum_{i=1}^{n}(y_i - x_i\beta)^2$$

$$= \|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\boldsymbol{x}G^{-1}\boldsymbol{x}^T\boldsymbol{y} + (\beta - G^{-1}\boldsymbol{x}^T\boldsymbol{y})^T G(\beta - G^{-1}\boldsymbol{x}^T\boldsymbol{y})$$

3. **Minimizziamo** la quantità rispetto a β:

$$\beta_{OLS} = G^{-1}\boldsymbol{x}^T\boldsymbol{y}$$

MSE nella regressione multipla

$$MSE(h(\cdot;\beta)) =$$
$$\frac{1}{n}\left[\|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\Phi(\boldsymbol{x})G^{-1}\Phi(\boldsymbol{x})^T\boldsymbol{y} + (\beta-\beta_{OLS})^T G(\beta-\beta_{OLS})\right]$$
$$= MSE(h(\cdot;\beta_{OLS})) + (\beta-\beta_{OLS})^T\frac{G}{n}(\beta-\beta_{OLS}) \geq 0$$

- Per $\beta = \beta_{OLS}$ otteniamo

$$MSE(h(;\beta_{OLS})) = \frac{1}{n}\left(\|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\Phi(\boldsymbol{x})G^{-1}\Phi(\boldsymbol{x})^T\boldsymbol{y}\right)$$
$$= \frac{1}{n}\|(\mathrm{Id} - \Pi_{\Phi(\boldsymbol{x})})\boldsymbol{y}\|^2$$

MSE nella regressione multipla

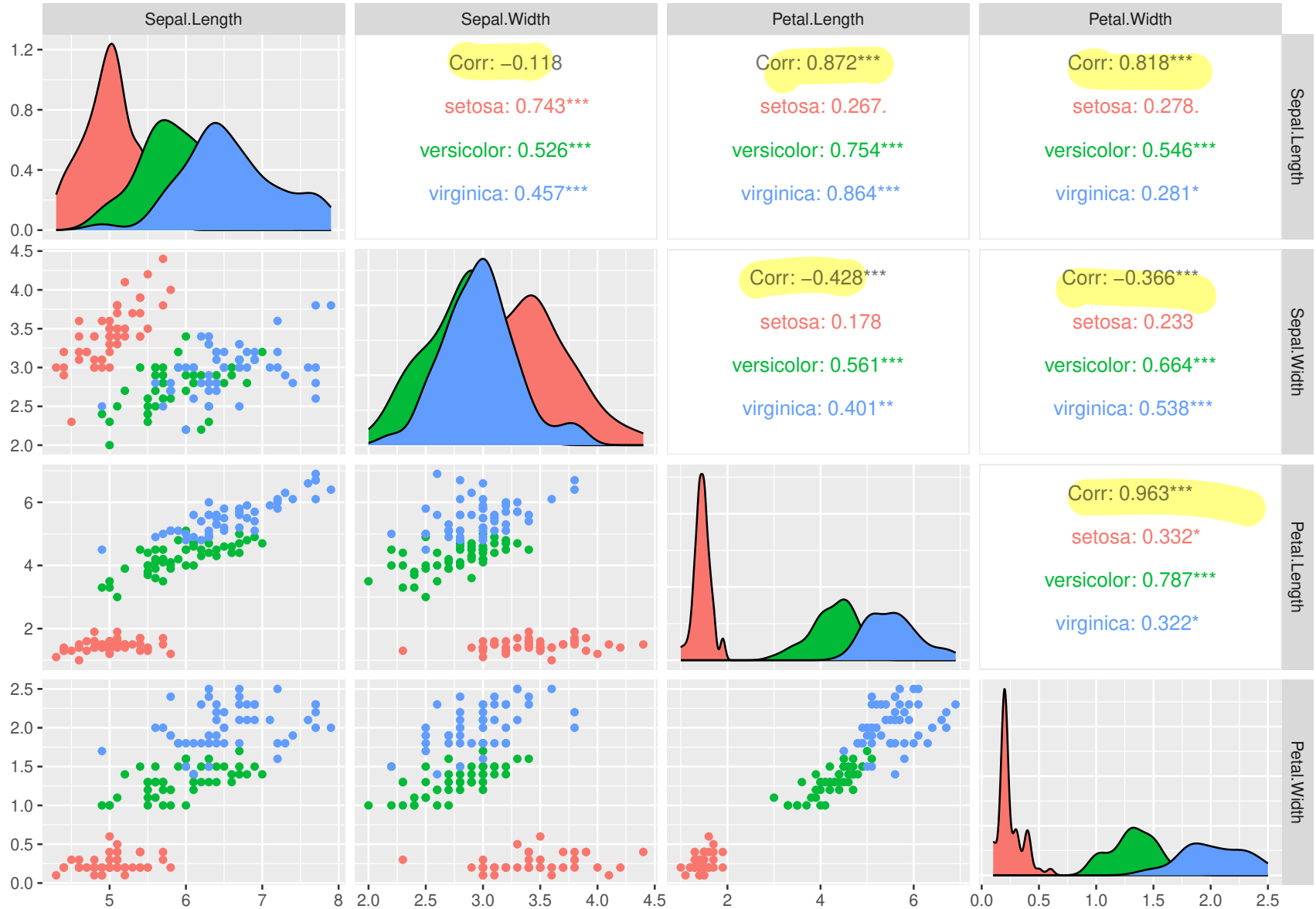
$$MSE(h(\cdot;\beta)) =$$
$$\frac{1}{n}\left[\|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\Phi(\boldsymbol{x})G^{-1}\Phi(\boldsymbol{x})^T\boldsymbol{y} + (\beta-\beta_{OLS})^T G(\beta-\beta_{OLS})\right]$$
$$= MSE(h(\cdot;\beta_{OLS})) + (\beta-\beta_{OLS})^T\frac{G}{n}(\beta-\beta_{OLS})$$

- Per $\beta = \beta_{OLS}$ otteniamo

$$MSE(h(;\beta_{OLS})) = \frac{1}{n}\left(\|\boldsymbol{y}\|^2 - \boldsymbol{y}^T\Phi(\boldsymbol{x})G^{-1}\Phi(\boldsymbol{x})^T\boldsymbol{y}\right)$$
$$= \frac{1}{n}\|(\mathrm{Id} - \Pi_{\Phi(\boldsymbol{x})})\boldsymbol{y}\|^2$$

- **Interpetazione geometrica**: $MSE \propto$ norma al quadrato della proiezione di $\boldsymbol{y}$ sul **complemento ortogonale** dello spazio generato dalle colonne di $\Phi(\boldsymbol{x})$.

Esempio (iris)

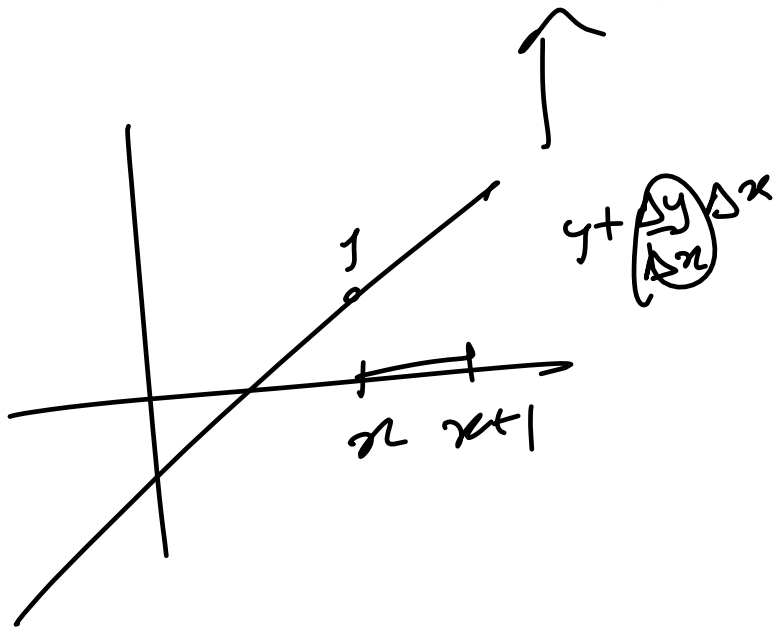


Confronto dei coefficienti della regressione multipla di Sepal.Length in funzione di Sepal.Width, Petal.Length, Petal.Width tramite comando `lm()` e tramite formula matriciale:

1. Comando `lm()` di R:

```
##  (Intercept)  Sepal.Width Petal.Length  Petal.Width
##    1.8559975    0.6508372    0.7091320   -0.5564827
```

```
##                      [,1]
## 1               1.8559975
## Sepal.Width     0.6508372
## Petal.Length    0.7091320
## Petal.Width    -0.5564827
```



Confronto dei coefficienti della regressione multipla di Sepal.Length in funzione di Sepal.Width, Petal.Length, Petal.Width tramite comando `lm()` e tramite formula matriciale:

1. Comando `lm()` di R:

```
##  (Intercept)  Sepal.Width Petal.Length  Petal.Width
##    1.8559975    0.6508372    0.7091320   -0.5564827
```

2. Formula matriciale:

```
##                     [,1]
## 1              1.8559975
## Sepal.Width    0.6508372
## Petal.Length   0.7091320
## Petal.Width   -0.5564827
```

Stima dell'incertezza nella regressione

Problema generale

Precisare la stima puntuale dei parametri

$$\theta_{OLS} \in \arg\min_{\theta} \sum_i (y_i - h(x_i; \theta))^2$$

indicandone l'*incertezza* dovuta ai dati:

$$\{(X_i, Y_i)\}_i \to \theta_{OLS}$$

- **Osservazione:** l'incertezza su θ_{OLS} si *propaga*

Problema generale

Precisare la stima puntuale dei parametri

$$\theta_{OLS} \in \arg\min_{\theta} \sum_{i} (y_i - h(x;\theta))^2$$

indicandone l'*incertezza* dovuta ai dati:

$$\{(X_i, Y_i)\}_i \to \theta_{OLS}$$

- **Osservazione:** l'incertezza su θ_{OLS} si *propaga*
  - nella **confidenza** sul modello $h(x;\theta)$

Problema generale

Precisare la stima puntuale dei parametri

$$\theta_{OLS} \in \arg\min_{\theta} \sum_{i} (y_i - h(x;\theta))^2$$

indicandone l'*incertezza* dovuta ai dati:

$$\{(X_i, Y_i)\}_i \to \theta_{OLS}$$

- **Osservazione:** l'incertezza su θ_{OLS} si *propaga*
  - nella **confidenza** sul modello $h(x;\theta)$
  - nella **previsione** $h(x;\theta) + \varepsilon$

Problema generale

Precisare la stima puntuale dei parametri

$$\theta_{OLS} \in \arg\min_{\theta} \sum_{i} (y_i - h(x;\theta))^2$$

indicandone l'*incertezza* dovuta ai dati:

$$\{(X_i, Y_i)\}_i \to \theta_{OLS}$$

- **Osservazione:** l'incertezza su θ_{OLS} si *propaga*
  - nella **confidenza** sul modello $h(x;\theta)$
  - nella **previsione** $h(x;\theta) + \varepsilon$
- **Due approcci**:

Problema generale

Precisare la stima puntuale dei parametri

$$\theta_{OLS} \in \arg\min_{\theta} \sum_{i} (y_i - h(x;\theta))^2$$

indicandone l'*incertezza* dovuta ai dati:

$$\{(X_i, Y_i)\}_i \to \theta_{OLS}$$

- **Osservazione:** l'incertezza su θ_{OLS} si *propaga*
  - nella **confidenza** sul modello $h(x;\theta)$
  - nella **previsione** $h(x;\theta) + \varepsilon$
- **Due approcci**:
  1. Non parametrico (bootstrap)

Problema generale

Precisare la stima puntuale dei parametri

$$\theta_{OLS} \in \arg\min_{\theta} \sum_{i} (y_i - h(x;\theta))^2$$

indicandone l'*incertezza* dovuta ai dati:

$$\{(X_i, Y_i)\}_i \to \theta_{OLS}$$

- **Osservazione:** l'incertezza su θ_{OLS} si *propaga*
  - nella **confidenza** sul modello $h(x;\theta)$
  - nella **previsione** $h(x;\theta) + \varepsilon$
- **Due approcci**:
  1. Non parametrico (bootstrap)
  2. Parametrico (bayesiano/frequentista)

Approccio non parametrico: bootstrap

- **Idea:** simulare più volte il processo di apprendimento su dati diversi per stimare la variabilità di θ_{OLS}.

Approccio non parametrico: bootstrap

- **Idea:** simulare più volte il processo di apprendimento su dati diversi per stimare la variabilità di θ_{OLS}.
- **Problema:** non conosciamo la distribuzione dei dati $P(X, Y)$.

Approccio non parametrico: bootstrap

- **Idea:** simulare più volte il processo di apprendimento su dati diversi per stimare la variabilità di θ_{OLS}.
- **Problema:** non conosciamo la distribuzione dei dati $P(X, Y)$.
- **Bootstrap**:

Approccio non parametrico: bootstrap

- **Idea:** simulare più volte il processo di apprendimento su dati diversi per stimare la variabilità di θ_{OLS}.
- **Problema:** non conosciamo la distribuzione dei dati $P(X, Y)$.
- **Bootstrap**:
  1. generazione di nuovi campioni tramite **estrazione con rimpiazzo** dal training set $D := \{(x_i, y_i)\}_{i=1,\ldots,n}$:

$$D_1, D_2, \ldots, D_B$$

Approccio non parametrico: bootstrap

- **Idea:** simulare più volte il processo di apprendimento su dati diversi per stimare la variabilità di θ_{OLS}.
- **Problema:** non conosciamo la distribuzione dei dati $P(X, Y)$.
- **Bootstrap**:
  1. generazione di nuovi campioni tramite **estrazione con rimpiazzo** dal training set $D := \{(x_i, y_i)\}_{i=1,\dots,n}$:

$$D_1, D_2, \dots, D_B$$

  2. training del modello sui dati generati

$$\theta_{OLS,i}, \theta_{OLS,2}, \dots, \theta_{OLS,B} \in \mathbb{R}^p$$

Approccio non parametrico: bootstrap

- **Idea:** simulare più volte il processo di apprendimento su dati diversi per stimare la variabilità di θ_{OLS}.
- **Problema:** non conosciamo la distribuzione dei dati $P(X, Y)$.
- **Bootstrap**:

1. generazione di nuovi campioni tramite **estrazione con rimpiazzo** dal training set $D := \{(x_i, y_i)\}_{i=1,\dots,n}$:

$$D_1, D_2, \dots, D_B$$

2. training del modello sui dati generati

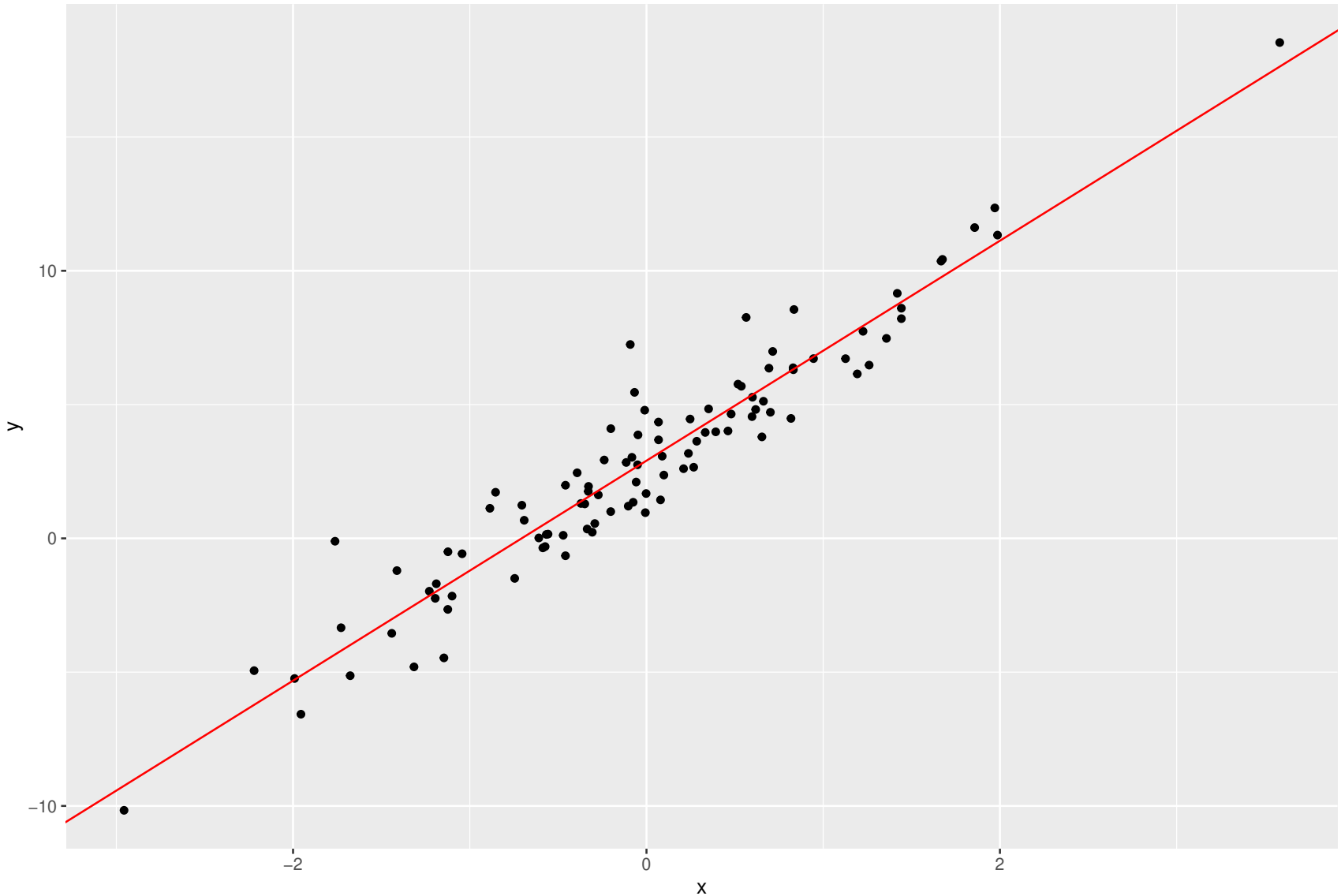
$$\theta_{OLS,i}, \theta_{OLS,2}, \dots, \theta_{OLS,B} \in \mathbb{R}^p$$

3. stima della variabilità di θ_{OLS} tramite distribuzione empirica:

$$m_\theta = \frac{1}{B}\sum_{i=1}^{B}\theta_{OLS,i}, \quad \Sigma_\theta^2 := \frac{1}{B-1}\sum_{i=1}^{B}(\theta_{OLS,i}-m_\theta)(\theta_{OLS,i}-m_\theta)^T$$

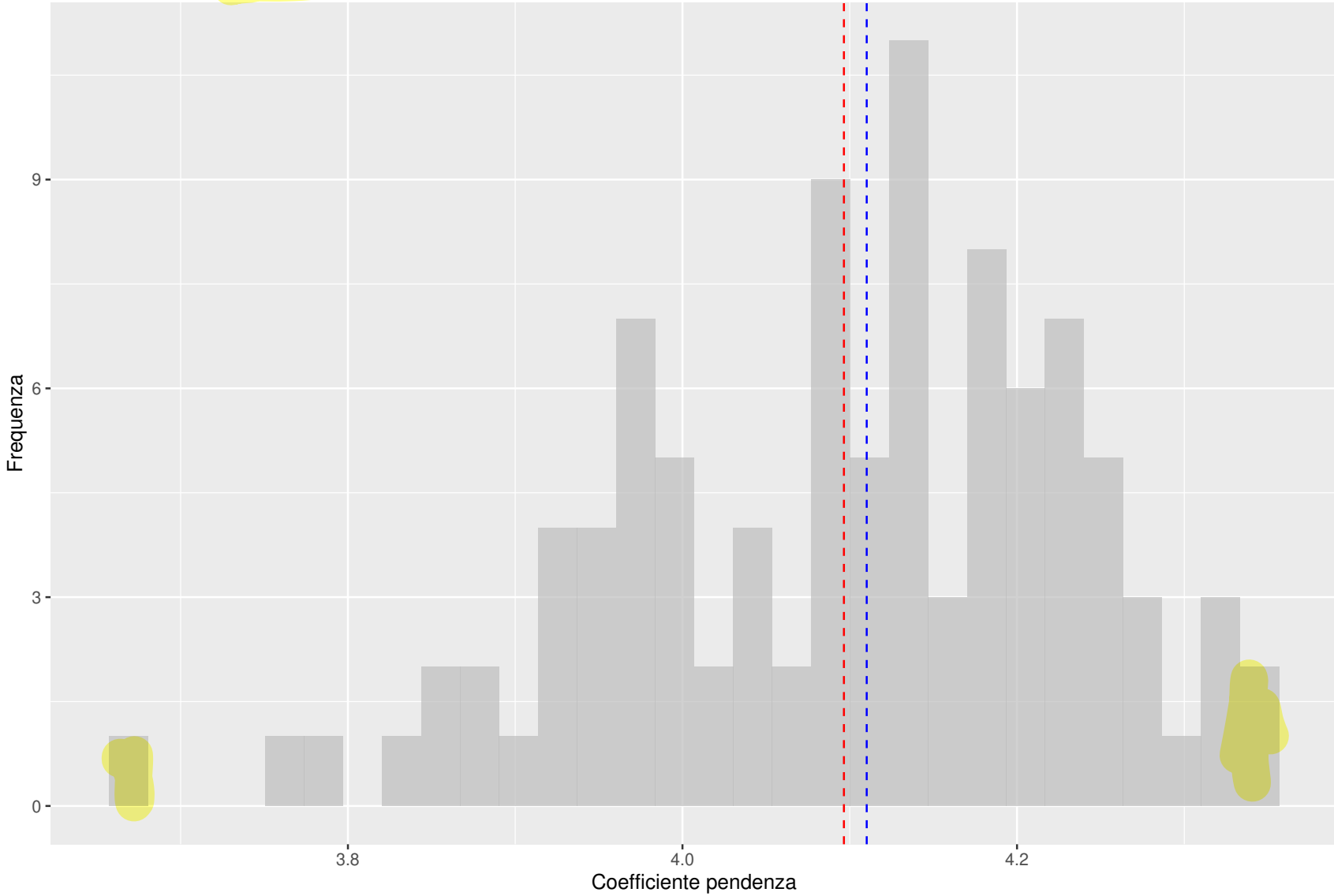
Esempio

Dati generati e modello OLS




```
## [1] "Coefficienti OLS: intercetta 2.9 pendenza 4.11"
```

Distribuzione del coefficiente pendenza (rosso: media bootstrap), (blu: stima OLS)



Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati
- **Cross-validation**: $\rightarrow$ stimare la performance del modello

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati
- **Cross-validation**: $\rightarrow$ stimare la performance del modello
  - si divide il training set in sottoinsiemi disgiunti

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati
- **Cross-validation**: $\rightarrow$ stimare la performance del modello
  - si divide il training set in sottoinsiemi disgiunti
  - i dati non sono ripetuti

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati
- **Cross-validation**: $\rightarrow$ stimare la performance del modello
  - si divide il training set in sottoinsiemi disgiunti
  - i dati non sono ripetuti
  - interessano (primariamente) i dati non usati per il training

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati
- **Cross-validation**: $\rightarrow$ stimare la performance del modello
  - si divide il training set in sottoinsiemi disgiunti
  - i dati non sono ripetuti
  - interessano (primariamente) i dati non usati per il training
- **Osservazione:** si può comunque usare la generazione di dati mediante bootstrap ai fini della cross-validation (e viceversa).

Differenza tra bootstrap e cross-validation

- **Bootstrap:** $\rightarrow$ stimare la variabilità del modello
  - si campiona su tutto il training set
  - i dati possono essere ripetuti (estrazioni con rimpiazzo)
  - non interessano (primariamente) i dati non campionati
- **Cross-validation**: $\rightarrow$ stimare la performance del modello
  - si divide il training set in sottoinsiemi disgiunti
  - i dati non sono ripetuti
  - interessano (primariamente) i dati non usati per il training
- **Osservazione:** si può comunque usare la generazione di dati mediante bootstrap ai fini della cross-validation (e viceversa).
  - I dati non campionati sono detti *Out-Of-Bag* (OoB).

Intervalli di confidenza bootstrap

- Intervallo di confidenza al livello $1 - \alpha$ per ciascun β_j si ottengono tramite i quantili della distribuzione empirica di bootstrap:

$$\left[q_{\alpha/2}, q_{1-\alpha/2}\right]$$

dove q_α è il quantile di ordine α della distribuzione empirica di β_j.

Intervalli di confidenza bootstrap

- Intervallo di confidenza al livello $1-\alpha$ per ciascun β_j si ottengono tramite i quantili della distribuzione empirica di bootstrap:

$$\left[q_{\alpha/2}, q_{1-\alpha/2}\right]$$

  dove q_α è il quantile di ordine α della distribuzione empirica di β_j.

- **Interpretazione precisa:** con frequenza $1-\alpha$ sui dati generati dal bootstrap l'intervallo contiene il parametro $\beta_{OLS,j}$.

Intervalli di confidenza bootstrap

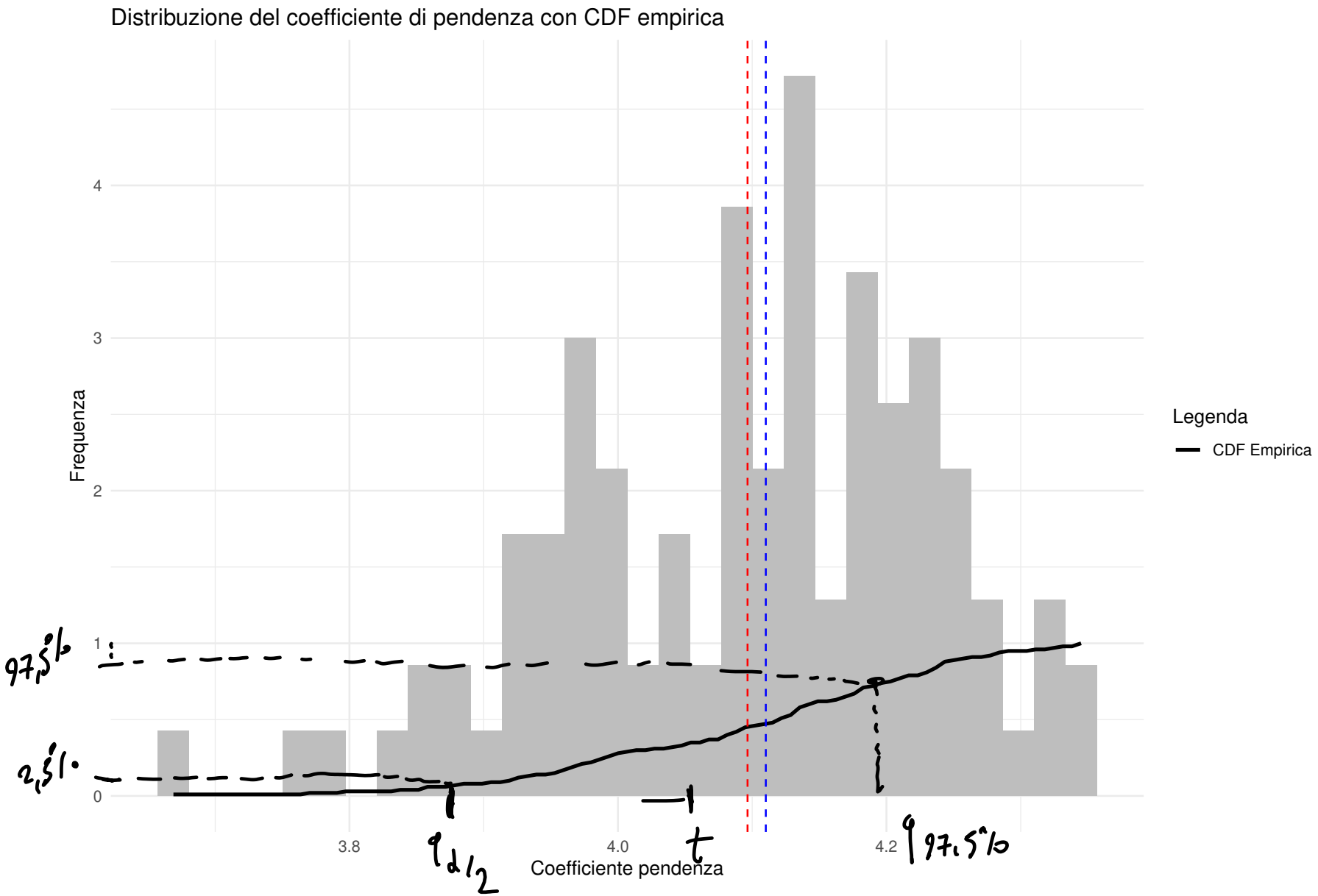
- Intervallo di confidenza al livello $1-\alpha$ per ciascun β_j si ottengono tramite i quantili della distribuzione empirica di bootstrap:

$$\left[q_{\alpha/2}, q_{1-\alpha/2}\right]$$

  dove q_α è il quantile di ordine α della distribuzione empirica di β_j.

- **Interpretazione precisa:** con frequenza $1-\alpha$ sui dati generati dal bootstrap l'intervallo contiene il parametro $\beta_{OLS,j}$.
- **Intepretazione pratica:** l'intervallo fornisce una stima della variabilità del parametro $\beta_{OLS,j}$.

Esempio di intervalli di confidenza



Esempio intervalli di confidenza bootstrap (dati generati)

Intervalli di confidenza bootstrap al 95% i per i coefficienti:

- intercetta: [2.66, 3.17]

Esempio intervalli di confidenza bootstrap (dati generati)

Intervalli di confidenza bootstrap al 95% i per i coefficienti:

- intercetta: [2.66, 3.17]
- pendenza: [3.81, 4.32]

Intervalli di confidenza bootstrap per il modello

- Dato x fisso, si può stimare un intervallo di confidenza per la previsione del modello $\underbrace{h(x; \beta_{OLS})}$

Intervalli di confidenza bootstrap per il modello

- Dato x fisso, si può stimare un intervallo di confidenza per la previsione del modello $h(x; \beta_{OLS})$
- Procedura: sui dati di bootstrap D_i si calcolano i parametri $\beta_{OLS,i}$ e si ottengono i valori

$$h(x; \beta_{OLS,i}) \quad i = 1, \ldots, B$$

Intervalli di confidenza bootstrap per il modello

- Dato x fisso, si può stimare un intervallo di confidenza per la previsione del modello $h(x; \beta_{OLS})$
- Procedura: sui dati di bootstrap D_i si calcolano i parametri $\beta_{OLS,i}$ e si ottengono i valori

$$h(x; \beta_{OLS,i}) \quad i = 1, \ldots, B$$

- Si calcolano indicatori: media, deviazione standard, quantili ...

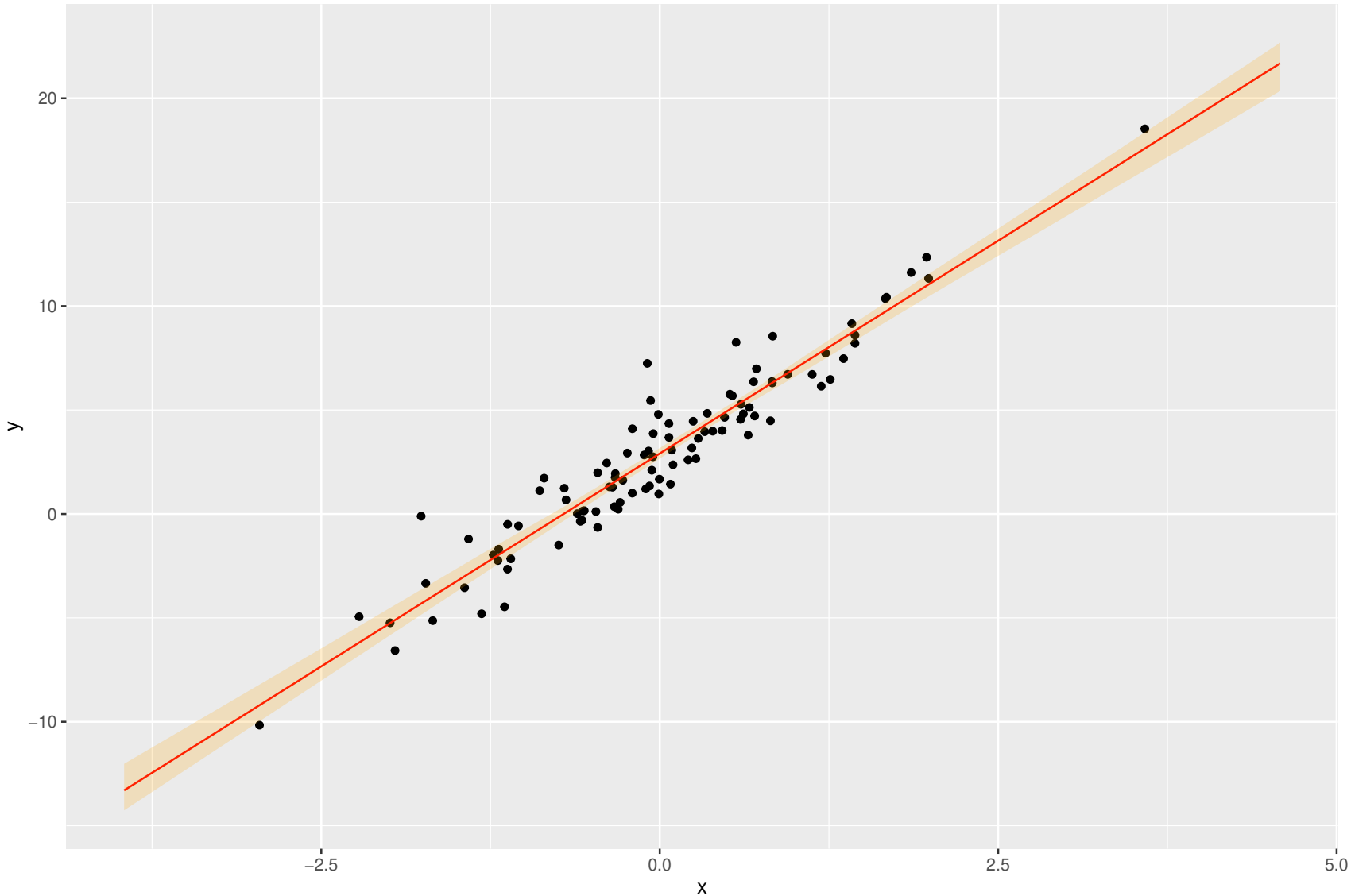
Esempio intervalli di confidenza bootstrap per il modello

- Media bootstrap di h in $x =$`1.5`: `9.05`

Esempio intervalli di confidenza bootstrap per il modello

- Media bootstrap di h in $x =$`1.5`: `9.05`
- Intervallo di confidenza bootstrap al 95% per h in $x =$`1.5`: [`8.57`, `9.45`]

Intervalli di confidenza bootstrap al 95% per il modello lineare semplice



Intervalli di previsione

- Si può stimare un intervallo di previsione per una nuova osservazione

$$y_{new} := h(x_{new}; \beta_{OLS}) + \varepsilon$$

Intervalli di previsione

- Si può stimare un intervallo di previsione per una nuova osservazione

$$y_{new} := h(x_{new}; \beta_{OLS}) + \varepsilon$$

- **Procedura**: sui dati di bootstrap D_i si calcolano i parametri $\beta_{OLS,i}$ e si ottengono i valori

$$y_{new,i} = h(x_{new}; \beta_{OLS,i}) + \varepsilon_i \quad i = 1, \ldots, B$$

dove ε_i sono campioni del rumore stimato sui residui del modello OLS.

Intervalli di previsione

- Si può stimare un intervallo di previsione per una nuova osservazione

$$y_{new} := h(x_{new}; \beta_{OLS}) + \varepsilon$$

- **Procedura**: sui dati di bootstrap D_i si calcolano i parametri $\beta_{OLS,i}$ e si ottengono i valori

$$y_{new,i} = h(x_{new}; \beta_{OLS,i}) + \varepsilon_i \quad i = 1, \ldots, B$$

dove ε_i sono campioni del rumore stimato sui residui del modello OLS.

- la distribuzione dei residui ε_i è stimata usando una cross-validation sui punti out-of-bag (OoB) per ogni campione bootstrap.

Intervalli di previsione

- Si può stimare un intervallo di previsione per una nuova osservazione

$$y_{new} := h(x_{new}; \beta_{OLS}) + \varepsilon$$

- **Procedura**: sui dati di bootstrap D_i si calcolano i parametri $\beta_{OLS,i}$ e si ottengono i valori

$$y_{new,i} = h(x_{new}; \beta_{OLS,i}) + \varepsilon_i \quad i = 1, \ldots, B$$

  dove ε_i sono campioni del rumore stimato sui residui del modello OLS.

- la distribuzione dei residui ε_i è stimata usando una cross-validation sui punti out-of-bag (OoB) per ogni campione bootstrap.

- Si calcolano indicatori: media, deviazione standard, quantili ...

Intervalli di previsione

- Si può stimare un intervallo di previsione per una nuova osservazione

$$y_{new} := h(x_{new}; \beta_{OLS}) + \varepsilon$$

- **Procedura**: sui dati di bootstrap D_i si calcolano i parametri $\beta_{OLS,i}$ e si ottengono i valori

$$y_{new,i} = h(x_{new}; \beta_{OLS,i}) + \varepsilon_i \quad i = 1, \ldots, B$$

  dove ε_i sono campioni del rumore stimato sui residui del modello OLS.

- la distribuzione dei residui ε_i è stimata usando una cross-validation sui punti out-of-bag (OoB) per ogni campione bootstrap.

- Si calcolano indicatori: media, deviazione standard, quantili ...

- **Attenzione:** gli intervalli di previsione sono più ampi degli intervalli di confidenza sul modello, in quanto tengono conto della variabilità del rumore.

Esempio intervalli di previsione bootstrap

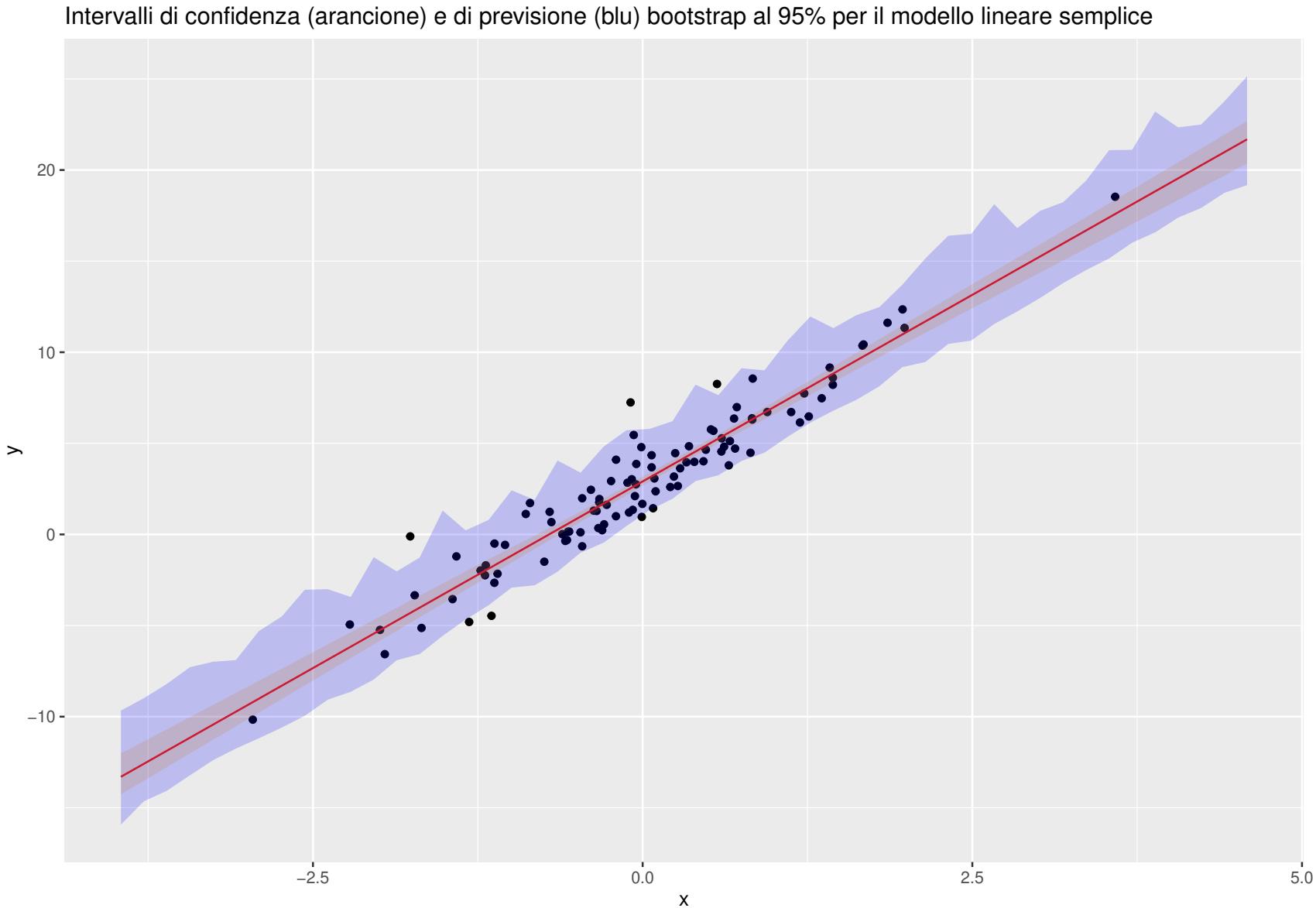
- Media bootstrap di y_{new} in $x =$`1.5`: `9.09`

Esempio intervalli di previsione bootstrap

- Media bootstrap di y_{new} in x =1.5: 9.09
- Intervallo di previsione bootstrap al 95% per y_{new} in x =1.5: [7.26, 11.56]

Esempio intervalli di previsione bootstrap

- Media bootstrap di y_{new} in x =1.5: 9.09
- Intervallo di previsione bootstrap al 95% per y_{new} in x =1.5: [7.26, 11.56]
- confronto con l'intervallo di confidenza del modello: [8.57, 9.45]


Intervalli di confidenza (arancione) e di previsione (blu) bootstrap al 95% per il modello lineare semplice
20
10
0
−10
y
−2.5
0.0
2.5
5.0
x

Approccio parametrico

- Modello probabilistico (rumore gaussiano):

$$Y = h(X; \beta) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Approccio parametrico

- Modello probabilistico (rumore gaussiano):

$$Y = h(X; \beta) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

- Minimi quadrati $\leftrightarrow$ massima verosimiglianza:

$$(\beta_{OLS}, \cancel{\sigma^2_{OLS}}) = \arg\max_{\beta, \cancel{\sigma^2}} P(D|\beta, \sigma^2)$$

Approccio parametrico

- Modello probabilistico (rumore gaussiano):

$$Y = h(X; \beta) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

- Minimi quadrati $\leftrightarrow$ massima verosimiglianza:

$$(\beta_{OLS}, \sigma^2_{OLS}) = \arg\max_{\beta, \sigma^2} P(D|\beta, \sigma^2)$$

- $\sigma^2_{OLS} = \frac{1}{n}\sum_{i=1}^{n}(y_i - h(x_i; \beta_{OLS}))^2 = \underbrace{MSE(h(\cdot; \beta_{OLS}))}$.

Approccio parametrico

- Modello probabilistico (rumore gaussiano):

$$Y = h(X; \beta) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

- Minimi quadrati $\leftrightarrow$ massima verosimiglianza:

$$(\beta_{OLS}, \sigma^2_{OLS}) = \arg\max_{\beta, \sigma^2} P(D|\beta, \sigma^2)$$

- $\sigma^2_{OLS} = \frac{1}{n}\sum_{i=1}^{n}(y_i - h(x_i; \beta_{OLS}))^2 = MSE(h(\cdot; \beta_{OLS}))$.
- **Obiettivo**: fornire una stima dell'incertezza sui parametri (β, σ^2) e sulla previsione.

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$
  - si stima di incertezza con intervalli di credibilità (via quantili)

$$P(\theta \in I(D)|D) \geq 1 - \alpha$$

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$
  - si stima di incertezza con intervalli di credibilità (via quantili)

$$P(\theta \in I(D)|D) \geq 1 - \alpha$$

  - decisioni bayesiane (test)

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$
  - si stima di incertezza con intervalli di credibilità (via quantili)

$$P(\theta \in I(D)|D) \geq 1 - \alpha$$

  - decisioni bayesiane (test)
- **Frequentista:** i parametri sono *fissi* ma sconosciuti, i dati sono *aleatori* (da osservare).

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$
  - si stima di incertezza con intervalli di credibilità (via quantili)

$$P(\theta \in I(D)|D) \geq 1 - \alpha$$

  - decisioni bayesiane (test)
- **Frequentista:** i parametri sono *fissi* ma sconosciuti, i dati sono *aleatori* (da osservare).
  - si lavora con la verosimiglianza $\mathcal{L}(\theta; D) = P(D|\theta)$

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$
  - si stima di incertezza con intervalli di credibilità (via quantili)

$$P(\theta \in I(D)|D) \geq 1 - \alpha$$

  - decisioni bayesiane (test)
- **Frequentista:** i parametri sono *fissi* ma sconosciuti, i dati sono *aleatori* (da osservare).
  - si lavora con la verosimiglianza $\mathcal{L}(\theta; D) = P(D|\theta)$
  - stima di incertezza con intervalli di fiducia

$$P(\theta \in I(D)|\theta) \geq 1 - \alpha \quad \text{per ogni } \theta$$

Confronto approccio bayesiano e frequentista

- **Bayesiano:** i parametri sono *aleatori* (perché non noti), i dati sono *fissi* (osservati).
  - fissa una distribuzione a priori $P(\theta)$
  - calcola la densità a posteriori $P(\theta|D) \propto P(\theta)\mathcal{L}(\theta; D)$
  - si stima di incertezza con intervalli di credibilità (via quantili)

$$P(\theta \in I(D)|D) \geq 1 - \alpha$$

  - decisioni bayesiane (test)
- **Frequentista:** i parametri sono *fissi* ma sconosciuti, i dati sono *aleatori* (da osservare).
  - si lavora con la verosimiglianza $\mathcal{L}(\theta; D) = P(D|\theta)$
  - stima di incertezza con intervalli di fiducia

$$P(\theta \in I(D)|\theta) \geq 1 - \alpha \quad \text{per ogni } \theta$$

  - test frequentisti

Esempio: modello costante gaussiano

Consideriamo il modello *costante* (in X): $\phi_1(x) = 1 \quad \forall x$

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(y_i)_{i=1,\dots,n}$

- **Parametri**: $\theta = (\beta_0, \sigma)$

Esempio: modello costante gaussiano

Consideriamo il modello *costante* (in X):

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(y_i)_{i=1,\dots,n}$

- **Parametri**: $\theta = (\beta_0, \sigma)$
- **Dati**: $D = \{Y_i = y_i\}_{i=1}^n$

Esempio: modello costante gaussiano

Consideriamo il modello *costante* (in X):

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(y_i)_{i=1,\dots,n}$

- **Parametri**: $\theta = (\beta_0, \sigma)$
- **Dati**: $D = \{Y_i = y_i\}_{i=1}^n$
- **Verosimiglianza**:

$$\mathcal{L}(\theta; D) = P(D|\theta) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0}{\sigma}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}}.$$

$\mathcal{N}(\beta_0, \sigma^2)$

Esempio: modello costante gaussiano

Consideriamo il modello *costante* (in X):

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(y_i)_{i=1,\dots,n}$

- **Parametri**: $\theta = (\beta_0, \sigma)$
- **Dati**: $D = \{Y_i = y_i\}_{i=1}^n$
- **Verosimiglianza**:

$$\mathcal{L}(\theta; D) = P(D|\theta) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0}{\sigma}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}}.$$

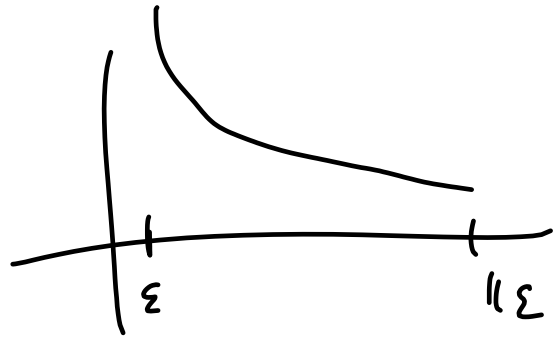
- Stima MLE (OLS): $\beta_{0,MLE} = \bar{y}_n$,

$$\sigma^2_{MLE} = \mathrm{Var}\,(y)_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = MSE(h(\cdot; \beta_{0,MLE})$$

Approccio bayesiano

- Densità a priori *uniforme* (completa ignoranza)

$$P(\beta_0, \sigma) \propto \frac{1}{\sigma} \quad \text{per } \beta_0 \in \mathbb{R}, \sigma > 0$$



Approccio bayesiano

- Densità a priori *uniforme* (completa ignoranza)

$$P(\beta_0, \sigma) \propto \frac{1}{\sigma} \quad \text{per } \beta_0 \in \mathbb{R}, \sigma > 0$$

- Verosimiglianza:

$$P(D|\beta_0, \sigma) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0}{\sigma}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}}$$
$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0)^2\right)\sigma^{-n}$$

Approccio bayesiano

- Densità a priori *uniforme* (completa ignoranza)

$$P(\beta_0, \sigma) \propto \frac{1}{\sigma} \quad \text{per } \beta_0 \in \mathbb{R}, \sigma > 0$$

- Verosimiglianza:

$$P(D|\beta_0, \sigma) = \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \beta_0}{\sigma}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0)^2\right) \sigma^{-n}$$

- **Densità a posteriori**:

$$P(\beta_0, \sigma|D) \propto \exp\left(-\frac{n}{2\sigma^2}\left[\text{Var}\,(y)_n + (\beta_0 - \bar{y}_n)^2\right]\right) \sigma^{-(n+1)}$$

Densità Student

- Integrando su σ si ottiene la densità a posteriori di β_0:

$$
\begin{aligned}
p(\beta_0|D) &\propto \int_0^\infty P(\beta_0, \sigma|D) d\sigma \\
&\propto \int_0^\infty \exp\left(-\frac{n}{2\sigma^2}\left[\mathrm{Var}\,(y)_n + (\beta_0 - \bar{y}_n)^2\right]\right) \sigma^{-(n+1)} d\sigma \\
&\propto \left(\mathrm{Var}\,(y) + (\beta_0 - \bar{y})^2\right)^{-n/2} \\
&\propto \left(1 + \frac{1}{n-1}\frac{(\beta_0 - \bar{y})^2}{\mathrm{var}(y)_n/n}\right)^{-n/2}
\end{aligned}
$$

$z = \frac{A}{\sigma^2} \quad \left|dz\right| = \left|\frac{A}{\sigma^3} d\sigma^2\right|$

Densità Student

- Integrando su σ si ottiene la densità a posteriori di β_0:

$$
\begin{aligned}
p(\beta_0|D) &\propto \int_0^\infty P(\beta_0, \sigma|D) d\sigma \\
&\propto \int_0^\infty \exp\left(-\frac{n}{2\sigma^2}\left[\mathrm{Var}\,(y)_n + (\beta_0 - \bar{y}_n)^2\right]\right) \sigma^{-(n+1)} d\sigma \\
&\propto \left(\mathrm{Var}\,(y) + (\beta_0 - \bar{y})^2\right)^{-n/2} \\
&\propto \left(1 + \frac{1}{n-1}\frac{(\beta_0 - \bar{y})^2}{\mathrm{var}(y)_n/n}\right)^{-n/2}
\end{aligned}
$$

- Si riconosce una distribuzione di Student-t con $n-1$ gradi di libertà, media $\bar{y}_n$ e scala al quadrato $\frac{\mathrm{var}(y)_n}{n}$.

$$
p(\beta_0|D) = t\left(\bar{y}_n, \frac{\mathrm{var}(y)_n}{n}, n-1\right)
$$

Densità di Student-t

La formula generale della densità di Student-t con df gradi di libertà parametri di media μ e scala al quadrato σ^2 è:

$$p(Z = z) \propto \left(1 + \frac{1}{df} \frac{(z-\mu)^2}{\sigma^2}\right)^{-\frac{df+1}{2}}$$

- **Notazione:** $p(Z) = t(\mu, \sigma^2, df)$

Densità di Student-t

La formula generale della densità di Student-t con df gradi di libertà parametri di media μ e scala al quadrato σ^2 è:

$$p(Z = z) \propto \left(1 + \frac{1}{df}\frac{(z-\mu)^2}{\sigma^2}\right)^{-\frac{df+1}{2}}$$

- **Notazione:** $p(Z) = t(\mu, \sigma^2, df)$
- **Osservazioni:**

Densità di Student-t

La formula generale della densità di Student-t con df gradi di libertà parametri di media μ e scala al quadrato σ^2 è:

$$p(Z = z) \propto \left(1 + \frac{1}{df}\frac{(z-\mu)^2}{\sigma^2}\right)^{-\frac{df+1}{2}}$$

- **Notazione:** $p(Z) = t(\mu, \sigma^2, df)$
- **Osservazioni:**
  1. code più pesanti della distribuzione normale per df piccoli (maggiore incertezza)

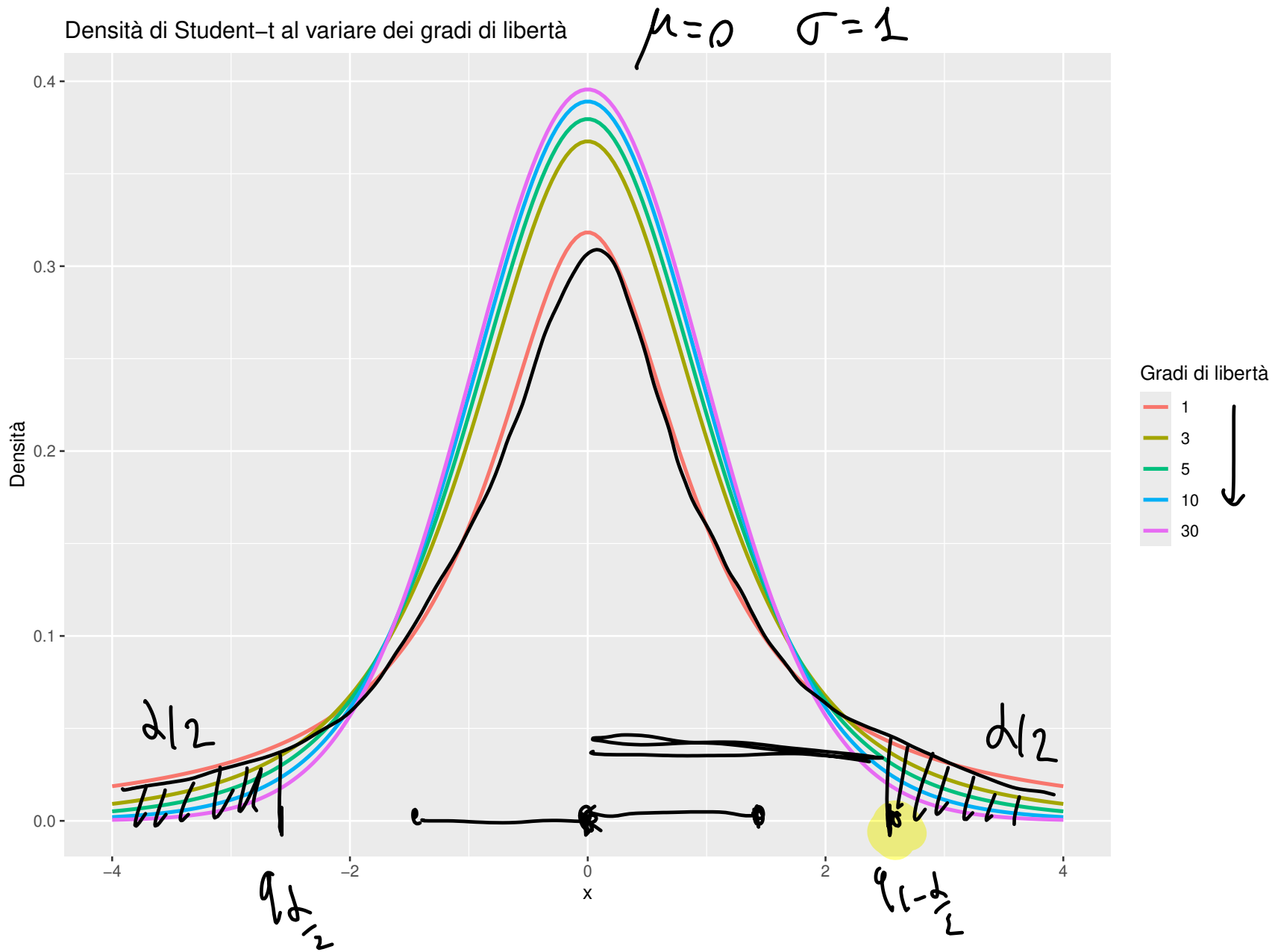
Densità di Student-t

La formula generale della densità di Student-t con df gradi di libertà parametri di media μ e scala al quadrato σ^2 è:

$$p(Z = z) \propto \left(1 + \frac{1}{df}\frac{(z-\mu)^2}{\sigma^2}\right)^{-\frac{df+1}{2}}$$

- **Notazione:** $p(Z) = t(\mu, \sigma^2, df)$
- **Osservazioni:**
  1. code più pesanti della distribuzione normale per df piccoli (maggiore incertezza)
  2. per $df \to \infty$ la distribuzione di Student-t converge alla distribuzione gaussiana $\mathcal{N}(\mu, \sigma^2)$ (in pratica, per $df > 30$ la differenza è trascurabile).

Grafici della densità al variare dei gradi di libertà



Stime e incertezza

- Incertezza su β_0:

Stime e incertezza

- Incertezza su β_0:
  1. **Approssimata** (usando stima *MLE* per σ, valida per $n \to \infty$):

$$p(\beta_0|D) = \mathcal{N}\left(\bar{y}, \frac{\text{var}(y)}{n}\right)$$

Stime e incertezza

- Incertezza su β_0:
  1. **Approssimata** (usando stima *MLE* per σ, valida per $n \to \infty$):

$$p(\beta_0|D) = \mathcal{N}\left(\bar{y}, \frac{\text{var}(y)}{n}\right)$$

  2. **Bayesiana esatta**: Student-t con $n-1$ gradi di libertà:

$$p(\beta_0|D) = t\left(\bar{y}, \frac{\text{var}(y)}{n}, n-1\right)$$

Stime e incertezza

- Incertezza su β_0:
  1. **Approssimata** (usando stima *MLE* per σ, valida per $n \to \infty$):

$$p(\beta_0|D) = \mathcal{N}\left(\bar{y}, \frac{\text{var}(y)}{n}\right)$$

  2. **Bayesiana esatta**: Student-t con $n-1$ gradi di libertà:

$$p(\beta_0|D) = t\left(\bar{y}, \frac{\text{var}(y)}{n}, n-1\right)$$

- Intervalli di credibilità per β_0 tramite quantili della distribuzione a posteriori.

$$P\left(|\beta_0 - \bar{y}_n| \leq q_{1-\alpha/2}\sqrt{\frac{\text{var}(y)}{n}}\middle| D\right) = 1-\alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1-\alpha/2$ della distribuzione di Student-t con $n-1$ gradi di libertà (oppure della distribuzione gaussiana se si usa l'approssimazione).

Previsione bayesiana

- Incertezza su previsione di nuova osservazione Y_{n+1}:

Previsione bayesiana

- Incertezza su previsione di nuova osservazione Y_{n+1}:
  1. **Approssimata** (usando stima *MLE*):

$$p(Y_{n+1}|D) = \mathcal{N}\left(\bar{y}_n, \left(1 + \frac{1}{n}\right) \text{var}(y)_n\right)$$

Previsione bayesiana

- Incertezza su previsione di nuova osservazione Y_{n+1}:
  1. **Approssimata** (usando stima *MLE*):

$$p(Y_{n+1}|D) = \mathcal{N}\left(\bar{y}_n, \left(1 + \frac{1}{n}\right) \text{var}(y)_n\right)$$

  2. **Bayesiana esatta**: Student-t con $n-1$ gradi di libertà:

$$p(Y_{n+1}|D) = t\left(\bar{y}_n, \left(1 + \frac{1}{n}\right) \text{var}(y)_n, n-1\right)$$

Previsione bayesiana

- Incertezza su previsione di nuova osservazione Y_{n+1}:
  1. **Approssimata** (usando stima *MLE*):

$$p(Y_{n+1}|D) = \mathcal{N}\left(\bar{y}_n, \left(1 + \frac{1}{n}\right) \text{var}(y)_n\right)$$

  2. **Bayesiana esatta**: Student-t con $n-1$ gradi di libertà:

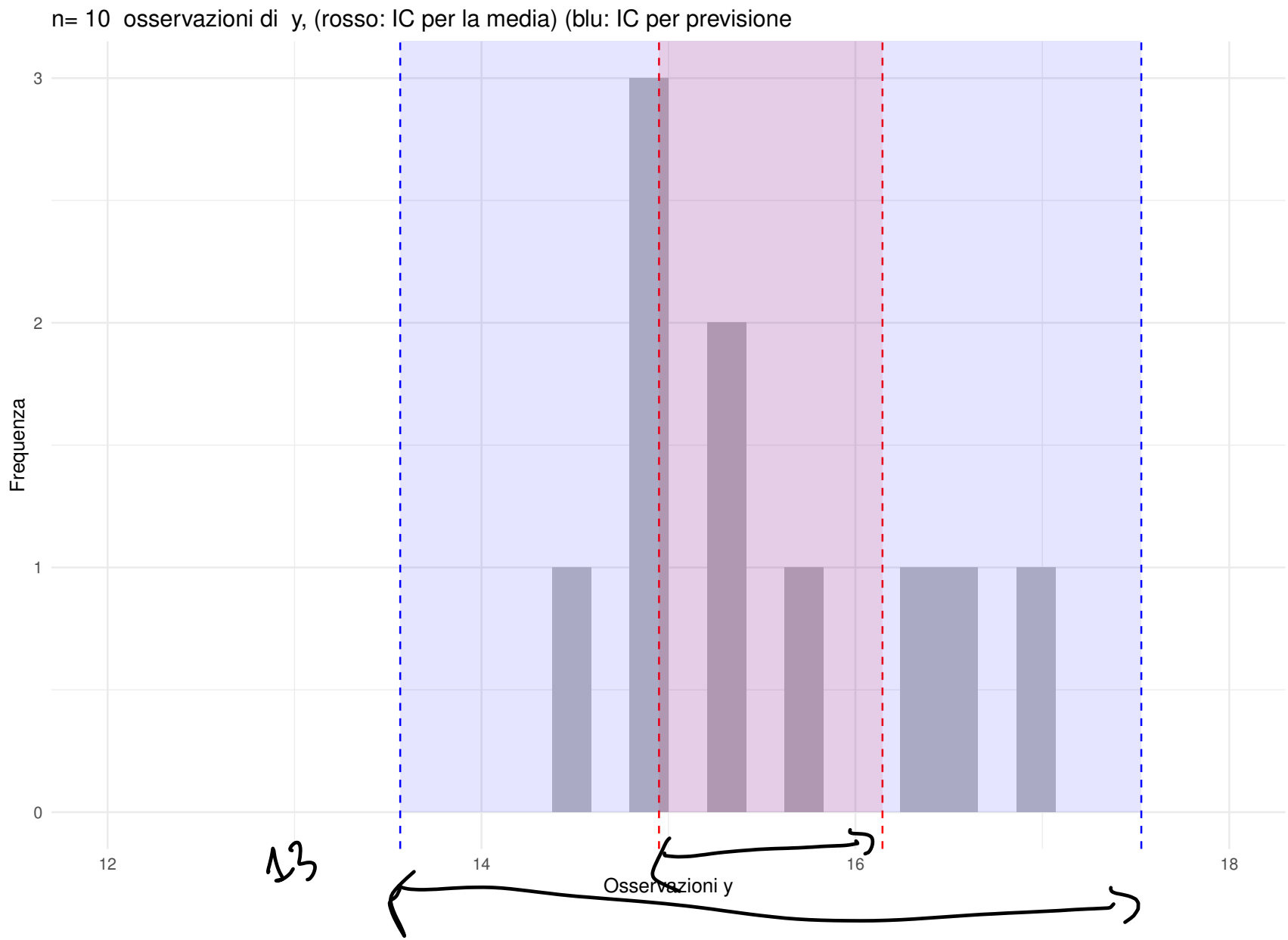
$$p(Y_{n+1}) = t\left(\bar{y}_n, \left(1 + \frac{1}{n}\right) \text{var}(y)_n, n-1\right)$$

- Intervalli di credibilità per la previsione:

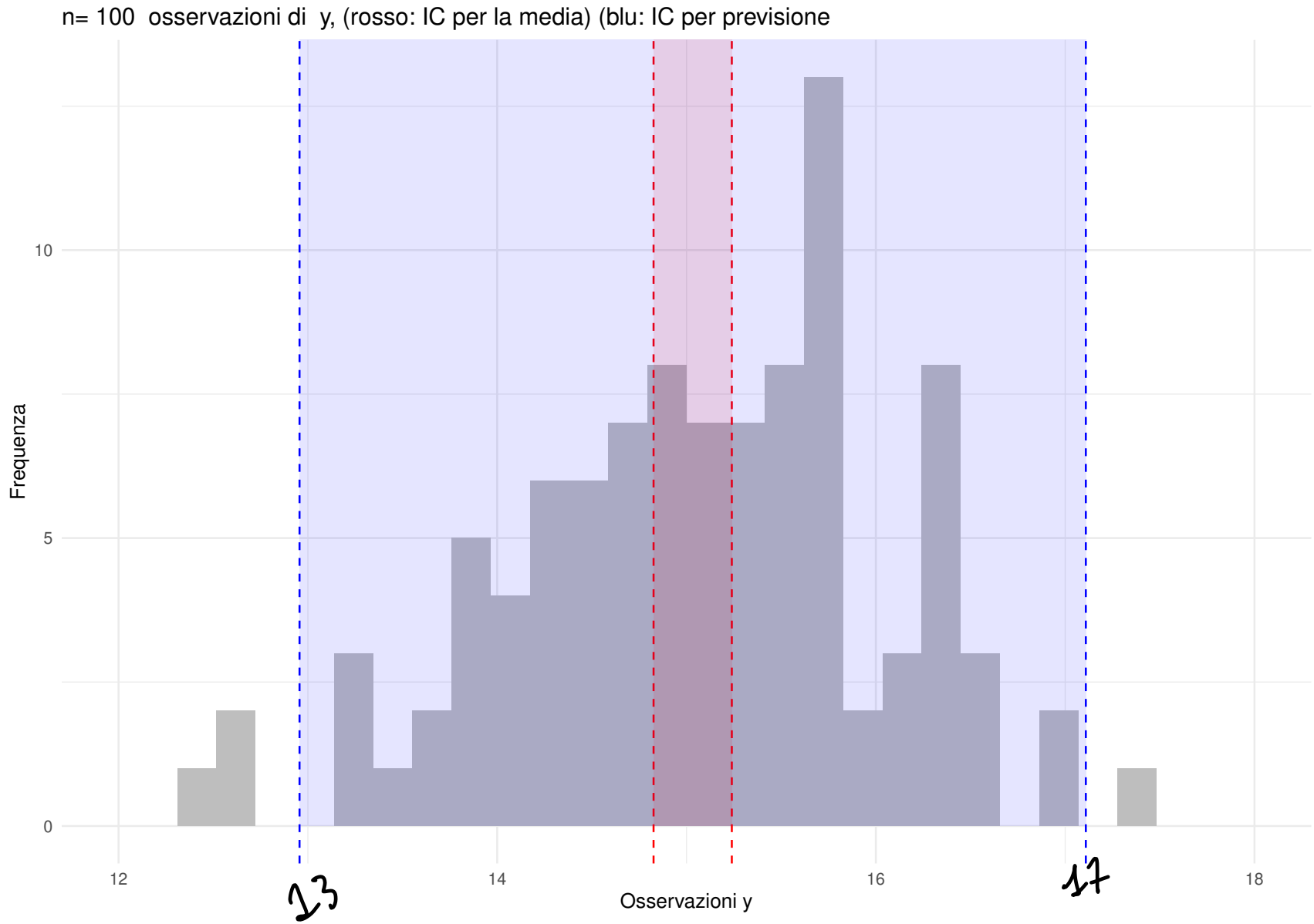
$$P\left(|Y_{n+1} - \bar{y}_n| \leq q_{1-\alpha/2}\sqrt{\left(1 + \frac{1}{n}\right) \text{var}(y)_n} \middle| D\right) = 1 - \alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1 - \alpha/2$ della distribuzione di Student-t con $n-1$ gradi di libertà (oppure della distribuzione gaussiana se si usa l'approssimazione).

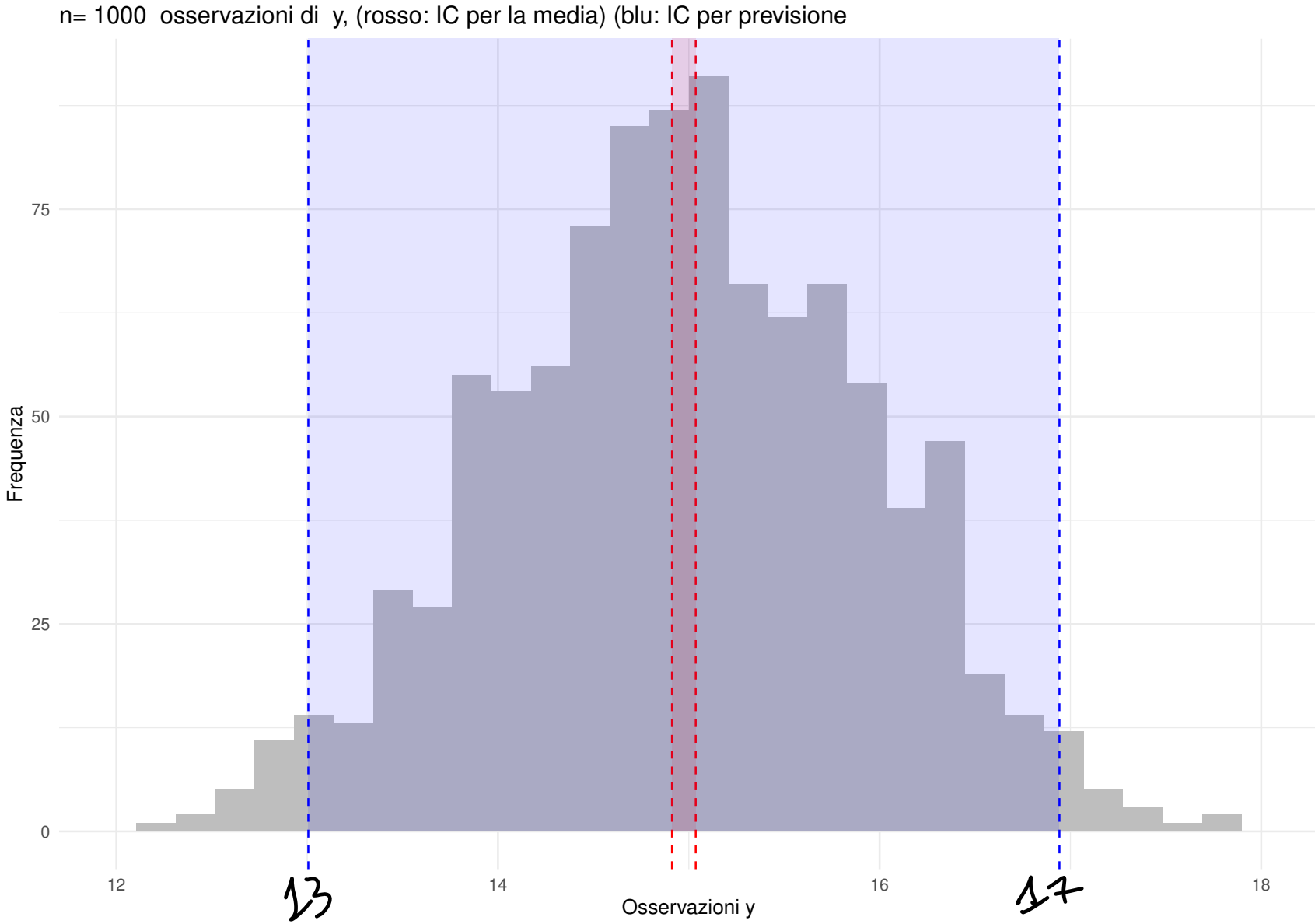
Esempio


n= 10 osservazioni di y, (rosso: IC per la media) (blu: IC per previsione
Frequenza
0
1
2
3
12
14
16
18
Osservazioni y

Esempio


n= 100 osservazioni di y, (rosso: IC per la media) (blu: IC per previsione
Frequenza
10
5
0
12
14
16
18
Osservazioni y
13
17

Esempio


n= 1000 osservazioni di y, (rosso: IC per la media) (blu: IC per previsione
Frequenza
0
25
50
75
12
13
14
16
17
18
Osservazioni y

Approccio frequentista

- Parametri fissi ma sconosciuti: $\theta = (\beta_0, \sigma^2)$

Approccio frequentista

- Parametri fissi ma sconosciuti: $\theta = (\beta_0, \sigma^2)$
- Training set $D = \{(X_i, Y_i)\}_{i=1}^n$ aleatorio.

Approccio frequentista

- Parametri fissi ma sconosciuti: $\theta = (\beta_0, \sigma^2)$
- Training set $D = \{(X_i, Y_i)\}_{i=1}^n$ aleatorio.
- **Intervallo di fiducia**: con frequenza $1 - \alpha$ sui possibili dati di training D, l'intervallo $I(D)$ contiene il parametro θ *"vero"*:

$$P\left(\theta \in I(D) | \theta\right) \geq 1 - \alpha \quad \text{per ogni } \theta.$$

Approccio frequentista

- Parametri fissi ma sconosciuti: $\theta = (\beta_0, \sigma^2)$
- Training set $D = \{(X_i, Y_i)\}_{i=1}^n$ aleatorio.
- **Intervallo di fiducia**: con frequenza $1 - \alpha$ sui possibili dati di training D, l'intervallo $I(D)$ contiene il parametro θ *"vero"*:

$$P\left(\theta \in I(D)|\theta\right) \geq 1 - \alpha \quad \text{per ogni } \theta.$$

- **Intepretazione:** l'intervallo di fiducia è una *procedura* che, ripetuta su più campioni, contiene il vero parametro con frequenza (almeno) $1 - \alpha$.

Caso costante (modello gaussiano)

Torniamo al modello costante (in X):

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(Y_i)_{i=1,\dots,n}$

- **Intervallo di fiducia per** β_0:

$$P\left(|\beta_0 - \bar{Y}_n| \leq q_{1-\alpha/2} \sqrt{\frac{\mathrm{var}(Y)_n}{n}} \middle| \beta_0 \right) = 1 - \alpha$$

Caso costante (modello gaussiano)

Torniamo al modello costante (in X):

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(Y_i)_{i=1,\dots,n}$

- **Intervallo di fiducia per** β_0:

$$P\left(|\beta_0 - \bar{Y}_n| \leq q_{1-\alpha/2} \sqrt{\frac{\text{var}(Y)_n}{n}} \middle| \beta_0 \right) = 1 - \alpha$$

- **Intervallo di fiducia per la previsione**:

$$P\left(|Y_{n+1} - \bar{Y}_n| \leq q_{1-\alpha/2} \sqrt{\left(1 + \frac{1}{n}\right) \text{var}(Y)_n} \middle| \beta_0 \right) = 1 - \alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1 - \alpha/2$ della distribuzione di Student-t con $n - 1$ gradi di libertà.

Caso costante (modello gaussiano)

Torniamo al modello costante (in X):

$$Y = \beta_0 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Problema: stima della media e varianza in un campione gaussiano $(Y_i)_{i=1,\dots,n}$

- **Intervallo di fiducia per** β_0:

$$P\left(|\beta_0 - \bar{Y}_n| \leq q_{1-\alpha/2} \sqrt{\frac{\text{var}(Y)_n}{n}} \middle| \beta_0 \right) = 1 - \alpha$$

- **Intervallo di fiducia per la previsione**:

$$P\left(|Y_{n+1} - \bar{Y}_n| \leq q_{1-\alpha/2} \sqrt{\left(1 + \frac{1}{n}\right) \text{var}(Y)_n} \middle| \beta_0 \right) = 1 - \alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1 - \alpha/2$ della distribuzione di Student-t con $n - 1$ gradi di libertà.

- **Come le formule bayesiane ma interpretazione diversa!**

Modello lineare generale: stime bayesiane

- Modello probabilistico sui dati:

$$Y = \Phi(X)\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

Modello lineare generale: stime bayesiane

- Modello probabilistico sui dati:

$$Y = \Phi(X)\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- Distribuzione a priori *uniforme*: $p(\beta, \sigma) = \frac{1}{\sigma}$ per $\beta \in \mathbb{R}^p, \sigma > 0$

Modello lineare generale: stime bayesiane

- Modello probabilistico sui dati:

$$Y = \Phi(X)\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- Distribuzione a priori *uniforme*: $p(\beta, \sigma) = \frac{1}{\sigma}$ per $\beta \in \mathbb{R}^p, \sigma > 0$
- Verosimiglianza (notazione matriciale):

$$P(D|\beta, \sigma) \propto \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \Phi(\boldsymbol{x})\beta\|^2\right)\sigma^{-n}$$

$$\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(y_i - \phi(x_i)\beta)^2\right)\frac{1}{\sigma}$$

Modello lineare generale: stime bayesiane

- Modello probabilistico sui dati:

$$Y = \Phi(X)\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- Distribuzione a priori *uniforme*: $p(\beta, \sigma) = \frac{1}{\sigma}$ per $\beta \in \mathbb{R}^p, \sigma > 0$
- Verosimiglianza (notazione matriciale):

$$P(D|\beta, \sigma) \propto \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \Phi(\boldsymbol{x})\beta\|^2\right) \sigma^{-n}$$

$\sim n\,\mathrm{MSE}(h(;\beta))$

- Densità a posteriori (usando l'identità per i minimi quadrati):

$$p(\beta, \sigma|D)$$

$$\propto \exp\left(-\frac{n\mathrm{MSE}(h(;\beta_{OLS})) + (\beta - \beta_{OLS})^T G(\beta - \beta_{OLS})}{2\sigma^2}\right) \sigma^{-(n+1)}$$

$G = \Phi(x)^T \Phi(x)$

Modello lineare generale: stime bayesiane

- Modello probabilistico sui dati:

$$Y = \Phi(X)\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$$

- Distribuzione a priori *uniforme*: $p(\beta, \sigma) = \frac{1}{\sigma}$ per $\beta \in \mathbb{R}^p, \sigma > 0$
- Verosimiglianza (notazione matriciale):

$$P(D|\beta, \sigma) \propto \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \Phi(\boldsymbol{x})\beta\|^2\right)\sigma^{-n}$$

- Densità a posteriori (usando l'identità per i minimi quadrati):

$$p(\beta, \sigma|D)$$
$$\propto \exp\left(-\frac{nMSE(h(;\beta_{OLS})) + (\beta - \beta_{OLS})^T G(\beta - \beta_{OLS})}{2\sigma^2}\right)\sigma^{-(n+1)}$$

- con $G = \Phi(\boldsymbol{x})^T\Phi(\boldsymbol{x})$ matrice di Gram.

Stime dell'incertezza

- Densità marginale per $\beta \in \mathbb{R}^i$

$$p(\beta|D) \propto \left(1 + \frac{1}{n-p} \frac{(\beta - \beta_{OLS})^T G (\beta - \beta_{OLS})}{RSE^2}\right)^{-n/2}$$

Stime dell'incertezza

- Densità marginale per β:

$$p(\beta|D) \propto \left(1 + \frac{1}{n-p} \frac{(\beta - \beta_{OLS})^T G (\beta - \beta_{OLS})}{RSE^2}\right)^{-n/2}$$

- **Residual Standard Error**
  $RSE = \sqrt{\frac{1}{n-p} \sum_{i=1}^{n} (y_i - \Phi(x_i)\beta_{OLS})}.$

$p \ll n \quad \frac{1}{n-p} \approx \frac{1}{n}$

$RSE \approx RMSE$

Stime dell'incertezza

- Densità marginale per β:

$$p(\beta|D) \propto \left(1 + \frac{1}{n-p} \frac{(\beta - \beta_{OLS})^T G (\beta - \beta_{OLS})}{RSE^2}\right)^{-n/2}$$

- **Residual Standard Error**
  $RSE = \sqrt{\frac{1}{n-p} \sum_{i=1}^{n} (y_i - \Phi(x_i)\beta_{OLS})}$.
- **Densità di Student multivariata**.

Densità di Student multivariata

Densità Student-t multivariata a valori in $\mathbb{R}^p$, con df gradi di libertà parametri di media μ e scala al quadrato Σ:

$$P(Z = z) \propto \left(1 + \frac{1}{df}(\boldsymbol{z} - \mu)^T \Sigma^{-1} (\boldsymbol{z} - \mu)\right)^{-\frac{df+p}{2}}$$

- Generalizza la densità di Student univariata.

GAUSSIANA multivariata

$$p(z=z) \propto \exp\left(-\frac{1}{2}(z-\mu)\Sigma^{-1}(z-\mu)\right)$$

Densità di Student multivariata

Densità Student-t multivariata a valori in $\mathbb{R}^p$, con df gradi di libertà parametri di media μ e scala al quadrato Σ:

$$P(Z = z) \propto \left(1 + \frac{1}{df}(\boldsymbol{z} - \mu)^T \Sigma^{-1} (\boldsymbol{z} - \mu)\right)^{-\frac{df+p}{2}}$$

- Generalizza la densità di Student univariata.
- **Notazione:** $p(Z) = t_p(\mu, \Sigma, df)$.

Densità di Student multivariata

Densità Student-t multivariata a valori in $\mathbb{R}^p$, con df gradi di libertà parametri di media μ e scala al quadrato Σ:

$$P(Z = z) \propto \left(1 + \frac{1}{df}(\boldsymbol{z} - \mu)^T \Sigma^{-1} (\boldsymbol{z} - \mu)\right)^{-\frac{df+p}{2}}$$

- Generalizza la densità di Student univariata.
- **Notazione:** $p(Z) = t_p(\mu, \Sigma, df)$.
- **Trasformazioni lineari:** se $P(Z) = t_p(\mu, \Sigma, df)$ e A è una matrice $m \times p$, allora

$$P(AZ) = t_m\left(A\mu, A\Sigma A^T, df\right)$$

Densità di Student multivariata

Densità Student-t multivariata a valori in $\mathbb{R}^p$, con df gradi di libertà parametri di media μ e scala al quadrato Σ:

$$P(Z = z) \propto \left(1 + \frac{1}{df}(\boldsymbol{z} - \mu)^T \Sigma^{-1} (\boldsymbol{z} - \mu)\right)^{-\frac{df+p}{2}}$$

- Generalizza la densità di Student univariata.
- **Notazione:** $p(Z) = t_p(\mu, \Sigma, df)$.
- **Trasformazioni lineari:** se $P(Z) = t_p(\mu, \Sigma, df)$ e A è una matrice $m \times p$, allora

$$P(AZ) = t_m\left(A\mu, A\Sigma A^T, df\right)$$

- Le marginali di una distribuzione di Student multivariata sono distribuzioni di Student univariate:

$$Z_j \sim t\left(\mu_j, \Sigma_{jj}, df\right)$$

Incertezza sui parametri e sul modello

- Con la notazione della Student multivariata:

$$P(\beta|D) = t_p\left(\beta_{OLS}, RSE^2 \cdot G^{-1}, n-p\right)$$

Incertezza sui parametri e sul modello

- Con la notazione della Student multivariata:

$$P(\beta|D) = t_p\left(\beta_{OLS}, RSE^2 \cdot G^{-1}, n-p\right)$$

- Ciascun β_j ha distribuzione di Student univariata:

$$P(\beta_j|D) = t_1\left(\beta_{OLS,j}, RSE^2 \cdot (G^{-1})_{jj}, n-p\right)$$

Incertezza sui parametri e sul modello

- Con la notazione della Student multivariata:

$$P(\beta|D) = t_p\left(\beta_{OLS}, RSE^2 \cdot G^{-1}, n-p\right)$$

- Ciascun β_j ha distribuzione di Student univariata:

$$P(\beta_j|D) = t_1\left(\beta_{OLS,j}, RSE^2 \cdot (G^{-1})_{jj}, n-p\right)$$

- Intervalli di credibilità per ciascun β_j:

$$P\left(|\beta_j - \beta_{OLS,j}| \leq q_{1-\alpha/2} RSE \sqrt{(G^{-1})_{jj}} \middle| D\right) = 1-\alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1-\alpha/2$ della distribuzione di Student-t con $n-p$ gradi di libertà.

Esempio (iris)

- Calcoliamo gli intervalli di credibilità bayesiani al 95% per i coefficienti di un modello di regressione lineare multipla sul dataset iris in cui si predice la lunghezza del sepalo in funzione della larghezza del sepalo e delle dimensioni del petalo.

```
## [1] "Intervallo di credibilità al 95% per beta_ 0 :"
## [1] "[1.08,2.64]"
## [1] "Intervallo di credibilità al 95% per beta_ 1 :"
## [1] "[-2.28,3.58]"
## [1] "Intervallo di credibilità al 95% per beta_ 2 :"
## [1] "[-2.74,4.16]"
## [1] "Intervallo di credibilità al 95% per beta_ 3 :"
## [1] "[-2.09,0.98]"
```

Esempio (iris)

- Calcoliamo gli intervalli di credibilità bayesiani al 95% per i coefficienti di un modello di regressione lineare multipla sul dataset iris in cui si predice la lunghezza del sepalo in funzione della larghezza del sepalo e delle dimensioni del petalo.
- $p = 4$ (incluso l'intercetta), $n = 150$.

```
## [1] "Intervallo di credibilità al 95% per beta_ 0 :"
## [1] "[1.08,2.64]"
## [1] "Intervallo di credibilità al 95% per beta_ 1 :"
## [1] "[-2.28,3.58]"
## [1] "Intervallo di credibilità al 95% per beta_ 2 :"
## [1] "[-2.74,4.16]"
## [1] "Intervallo di credibilità al 95% per beta_ 3 :"
## [1] "[-2.09,0.98]"
```

Intervalli di credibilità per il modello

- Dato x, il modello $\Phi(x)\beta$ è una trasformazione lineare di β:

$$P(h(x;\beta)|D) = t_1\left(h(x;\beta_{OLS}), RSE^2 \cdot s^2(x), n-p\right)$$

dove $s^2(x) = \phi(x) G^{-1} \phi(x)^T$.

$\phi(x)$ vettore riga

Intervalli di credibilità per il modello

- Dato x, il modello $\Phi(x)\beta$ è una trasformazione lineare di β:

$$P(h(x;\beta)|D) = t_1\left(h(x;\beta_{OLS}), RSE^2 \cdot s^2(x), n-p\right)$$

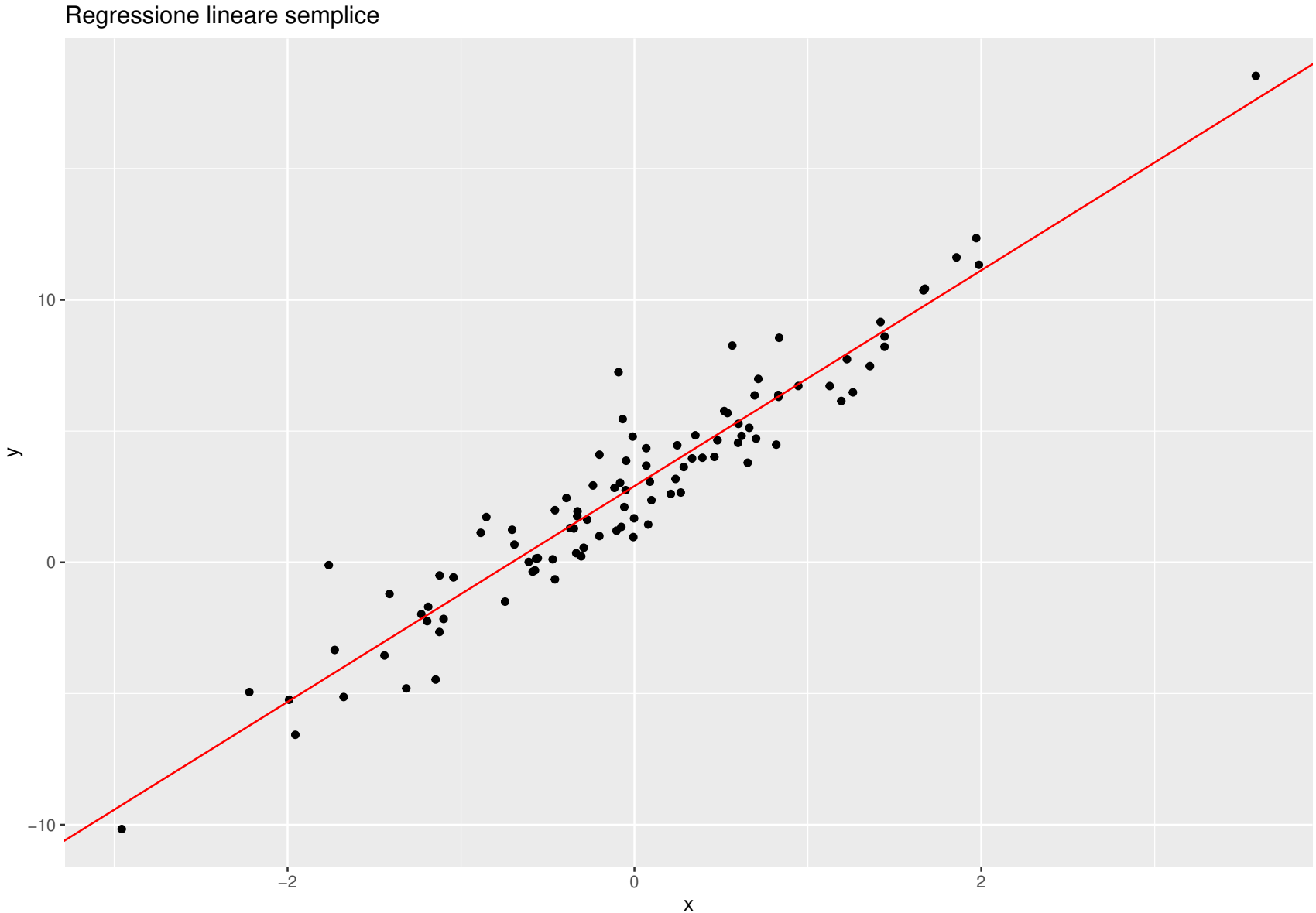
dove $s^2(x) = \phi(x)^T G^{-1}\phi(x)$.

- Intervalli di credibilità

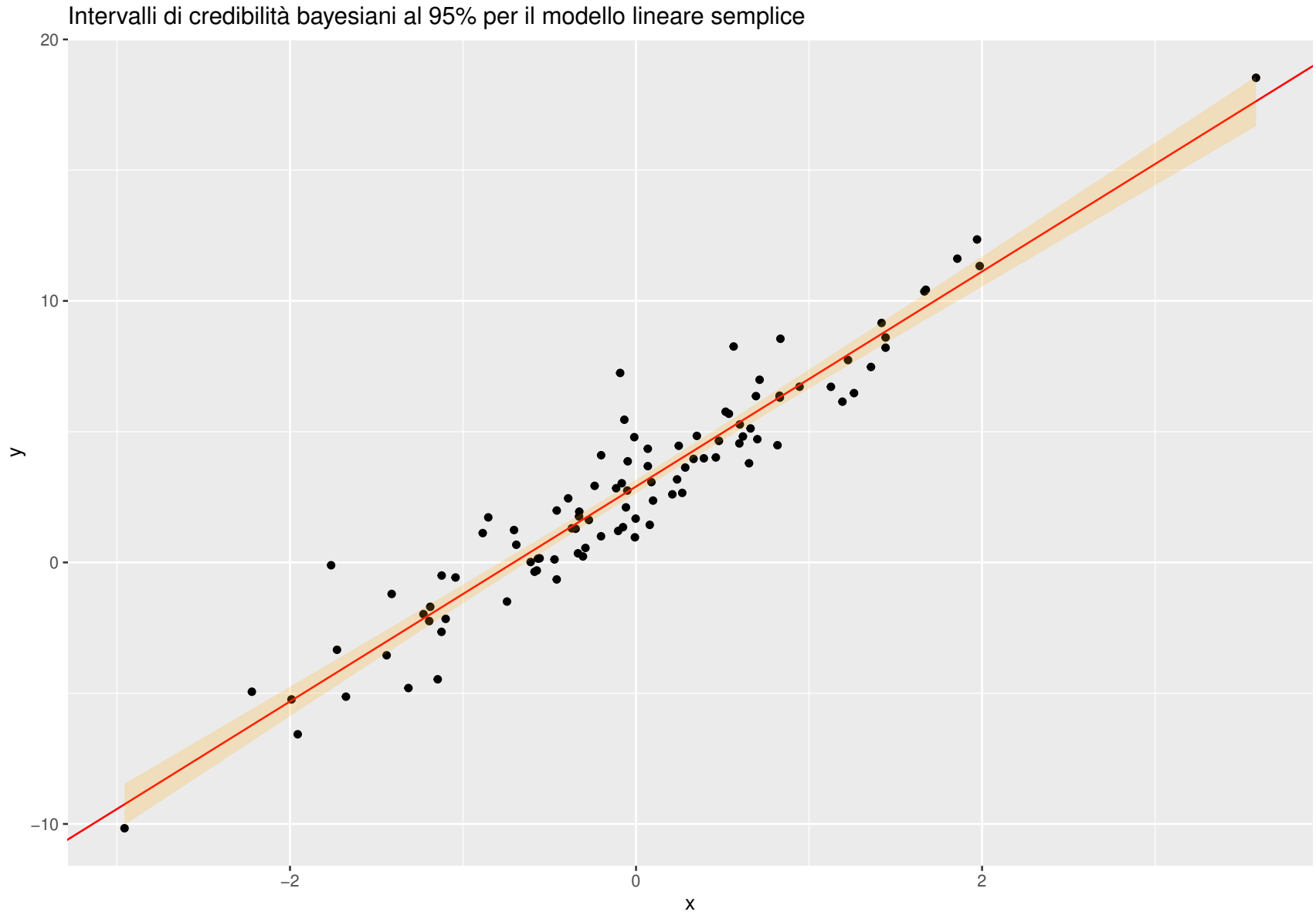
$$P\left(|h(x;\beta) - h(x;\beta_{OLS})| \leq q_{1-\alpha/2} RSE\sqrt{s^2(x)}\middle| D\right) = 1-\alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1-\alpha/2$ della distribuzione di Student-t con $n-p$ gradi di libertà.

Esempio (regressione lineare semplice)



Esempio (regressione lineare semplice)



Previsione bayesiana

- Densità a posteriori della previsione per un valore $X = x$ fissato: $Y = h(x; \beta) + \varepsilon$

$$p(Y = y|X = x, D) \propto \left(1 + \frac{1}{n-p} \frac{(y - h(x; \beta_{OLS}))^2}{RSE^2(1 + s^2(x))}\right)^{-\frac{n-p+1}{2}}$$

dove $s^2(x) = \phi(x)^T G^{-1} \phi(x)$.

Previsione bayesiana

- Densità a posteriori della previsione per un valore $X = x$ fissato: $Y = h(x; \beta) + \varepsilon$

$$p(Y = y|X = x, D) \propto \left(1 + \frac{1}{n-p} \frac{(y - h(x; \beta_{OLS}))^2}{RSE^2(1 + s^2(x))}\right)^{-\frac{n-p+1}{2}}$$

dove $s^2(x) = \phi(x)^T G^{-1} \phi(x)$.

- Densità Student-t univariata con $n - p$ gradi di libertà:

$$P(Y|X = x, D) = t_1\left(h(x; \beta_{OLS}), RSE^2 \cdot (1 + s^2(x)), n - p\right)$$

Previsione bayesiana

- Densità a posteriori della previsione per un valore $X = x$ fissato: $Y = h(x; \beta) + \varepsilon$

$$p(Y = y|X = x, D) \propto \left(1 + \frac{1}{n-p} \frac{(y - h(x; \beta_{OLS}))^2}{RSE^2(1 + s^2(x))}\right)^{-\frac{n-p+1}{2}}$$

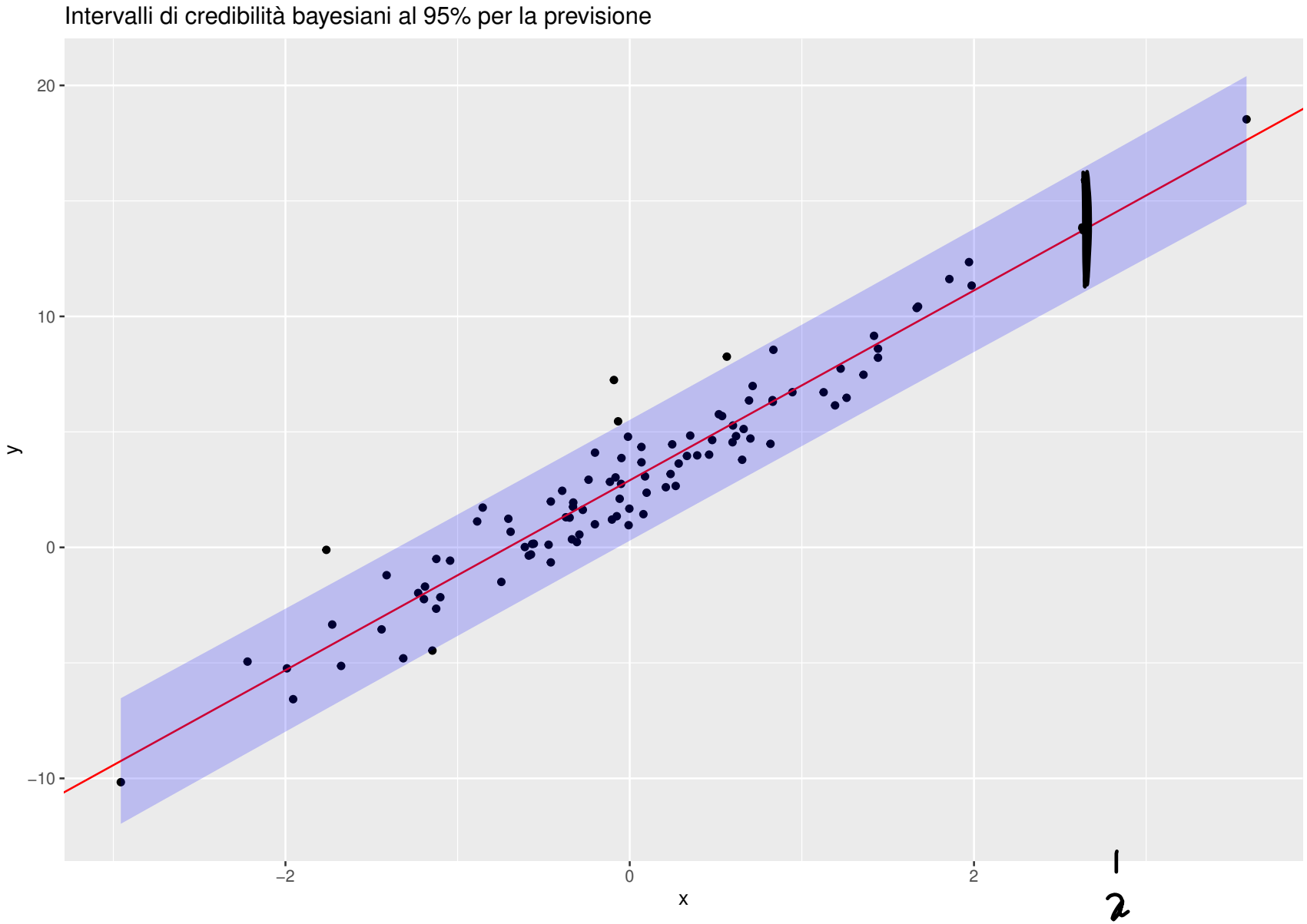
dove $s^2(x) = \phi(x)^T G^{-1} \phi(x)$.

- Densità Student-t univariata con $n - p$ gradi di libertà:

$$P(Y|X = x, D) = t_1\left(h(x; \beta_{OLS}), RSE^2 \cdot (1 + s^2(x)), n - p\right)$$

- Intervalli di credibilità per la previsione tramite quantili della distribuzione a posteriori.

Esempio (regressione lineare semplice)



Modello lineare generale: stime frequentiste

- Intervallo di fiducia per i parametri:

$$P\left(|\beta_j - \beta_{OLS,j}| \leq q_{1-\alpha/2} RSE \sqrt{(G^{-1})_{jj}} \,\middle|\, \beta\right) = 1 - \alpha$$

Fisso aleatorio

Modello lineare generale: stime frequentiste

- Intervallo di fiducia per i parametri:

$$P\left(|\beta_j - \beta_{OLS,j}| \leq q_{1-\alpha/2} RSE \sqrt{(G^{-1})_{jj}} \middle| \beta\right) = 1 - \alpha$$

- Intervallo di fiducia per il modello:

$$P\left(|h(x;\beta) - h(x;\beta_{OLS})| \leq q_{1-\alpha/2} RSE \sqrt{s^2(x)} \middle| \beta\right) = 1 - \alpha$$

Modello lineare generale: stime frequentiste

- Intervallo di fiducia per i parametri:

$$P\left(|\beta_j - \beta_{OLS,j}| \leq q_{1-\alpha/2} RSE \sqrt{(G^{-1})_{jj}} \middle| \beta\right) = 1 - \alpha$$

- Intervallo di fiducia per il modello:

$$P\left(|h(x;\beta) - h(x;\beta_{OLS})| \leq q_{1-\alpha/2} RSE \sqrt{s^2(x)} \middle| \beta\right) = 1 - \alpha$$

- Intervallo di fiducia per la previsione:

$$P\left(|Y - h(x;\beta_{OLS})| \leq q_{1-\alpha/2} RSE \sqrt{1 + s^2(x)} \middle| \beta\right) = 1 - \alpha$$

dove $q_{1-\alpha/2}$ è il quantile di ordine $1 - \alpha/2$ della distribuzione di Student-t con $n - p$ gradi di libertà.

Test di significatività dei parametri (frequentista)

- Ipotesi nulla: $H_0 : \beta_j = 0$ contro l'alternativa $H_1 : \beta_j \neq 0$.

Test di significatività dei parametri (frequentista)

- Ipotesi nulla: $H_0 : \beta_j = 0$ contro l'alternativa $H_1 : \beta_j \neq 0$.
- **Decisione:** rifiutare H_0 se 0 non appartiene all'intervallo di fiducia per β_j per un determinato livello $1 - \alpha$.

Test di significatività dei parametri (frequentista)

- Ipotesi nulla: $H_0 : \beta_j = 0$ contro l'alternativa $H_1 : \beta_j \neq 0$.
- **Decisione:** rifiutare H_0 se 0 non appartiene all'intervallo di fiducia per β_j per un determinato livello $1 - \alpha$.
- **Errore di I specie** (falsi positivi): probabilità di rifiutare H_0 quando è vera:

$$\begin{aligned}\alpha &= P(\text{rifiuto } H_0 | H_0 \text{ vera}) \\ &= P(0 \text{ non appartiene all'intervallo di fiducia} | \beta_j = 0) \\ &= P(|\beta_{OLS,j}| > q_{1-\alpha/2} RSE \sqrt{(G^{-1})_{jj}} | \beta_j = 0)\end{aligned}$$

Test di significatività dei parametri (frequentista)

- Ipotesi nulla: $H_0 : \beta_j = 0$ contro l'alternativa $H_1 : \beta_j \neq 0$.
- **Decisione:** rifiutare H_0 se 0 non appartiene all'intervallo di fiducia per β_j per un determinato livello $1 - \alpha$.
- **Errore di I specie** (falsi positivi): probabilità di rifiutare H_0 quando è vera:

$$\begin{aligned}\alpha &= P(\text{rifiuto } H_0 | H_0 \text{ vera}) \\ &= P(0 \text{ non appartiene all'intervallo di fiducia} | \beta_j = 0) \\ &= P(|\beta_{OLS,j}| > q_{1-\alpha/2} RSE \sqrt{(G^{-1})_{jj}} | \beta_j = 0)\end{aligned}$$

- Il **p-value** è il più piccolo livello di significatività α per cui si rifiuta H_0, una volta osservati i dati.

Test di significatività del modello (F-test)

- Ipotesi nulla: H_0: il modello non spiega nulla, cioè tutti i coefficienti sono nulli: $\beta = 0$.

Test di significatività del modello (F-test)

- Ipotesi nulla: H_0: il modello non spiega nulla, cioè tutti i coefficienti sono nulli: $\beta = 0$.
- Alternativa: almeno un coefficiente è diverso da zero.

Test di significatività del modello (F-test)

- Ipotesi nulla: H_0: il modello non spiega nulla, cioè tutti i coefficienti sono nulli: $\beta = 0$.
- Alternativa: almeno un coefficiente è diverso da zero.
- Statistica del test:

$$F = \frac{\frac{1}{p-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - h(x_i, \beta_{OLS}))^2} = \frac{\text{Explained Variance}/(p-1)}{RSE^2}$$

Test di significatività del modello (F-test)

- Ipotesi nulla: H_0: il modello non spiega nulla, cioè tutti i coefficienti sono nulli: $\beta = 0$.
- Alternativa: almeno un coefficiente è diverso da zero.
- Statistica del test:

$$F = \frac{\frac{1}{p-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - h(x_i, \beta_{OLS}))^2} = \frac{\text{Explained Variance}/(p-1)}{RSE^2}$$

- Sotto H_0, F ha distribuzione F di Fisher con $(p-1, n-p)$ gradi di libertà.

Test di significatività del modello (F-test)

- Ipotesi nulla: H_0: il modello non spiega nulla, cioè tutti i coefficienti sono nulli: $\beta = 0$.
- Alternativa: almeno un coefficiente è diverso da zero.
- Statistica del test:

$$F = \frac{\frac{1}{p-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - h(x_i, \beta_{OLS}))^2} = \frac{\text{Explained Variance}/(p-1)}{RSE^2}$$

- Sotto H_0, F ha distribuzione F di Fisher con $(p-1, n-p)$ gradi di libertà.
- **Decisione:** rifiutare H_0 se $F > F_{1-\alpha}(p-1, n-p)$, dove $F_{1-\alpha}(p-1, n-p)$ è il quantile di ordine $1-\alpha$ della distribuzione di Fisher con $(p-1, n-p)$ gradi di libertà.

Test di significatività del modello (F-test)

- Ipotesi nulla: H_0: il modello non spiega nulla, cioè tutti i coefficienti sono nulli: $\beta = 0$.
- Alternativa: almeno un coefficiente è diverso da zero.
- Statistica del test:

$$F = \frac{\frac{1}{p-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - h(x_i, \beta_{OLS}))^2} = \frac{\text{Explained Variance}/(p-1)}{RSE^2}$$

- Sotto H_0, F ha distribuzione F di Fisher con $(p-1, n-p)$ gradi di libertà.
- **Decisione:** rifiutare H_0 se $F > F_{1-\alpha}(p-1, n-p)$, dove $F_{1-\alpha}(p-1, n-p)$ è il quantile di ordine $1-\alpha$ della distribuzione di Fisher con $(p-1, n-p)$ gradi di libertà.
- Il **p-value** è il più piccolo livello di significatività α per cui si rifiuta H_0, una volta osservati i dati.