UNIVERSITÀ
DEGLI STUDI
DI PADOVA

METODI STATISTICI PER LA BIOINGEGNERIA (B)

PARTE 13: REGRESSIONE LOGISTICA

A.A. 2024-2025

Prof. Martina Vettoretti

STUDIO DEI FATTORI ASSOCIATI AD UNA VARIABILE QUALITATIVA

- Spesso siamo interessati a capire se sussiste una relazione tra una serie di fattori o variabili indipendenti e una variabile dipendente di tipo **qualitativo**.
- Esempio: vogliamo investigare se sussiste una relazione tra la lunghezza di un impianto dentale e il fallimento dello stesso.
  - Variabile indipendente: lunghezza impianto dentale [mm]
  - Outcome: esito dell'impianto (fallimento o successo) → variabile qualitativa binaria
- Come possiamo affrontare questo problema?
- Abbiamo visto che nei modelli di regressione lineare multipla possiamo inserire variabili indipendenti di tipo qualitativo mediante opportuna codifica. Possiamo procedere allo stesso modo se la variabile qualitativa è la variabile di outcome?

REGRESSIONE LINEARE CON VARIABILE DI OUTCOME BINARIA

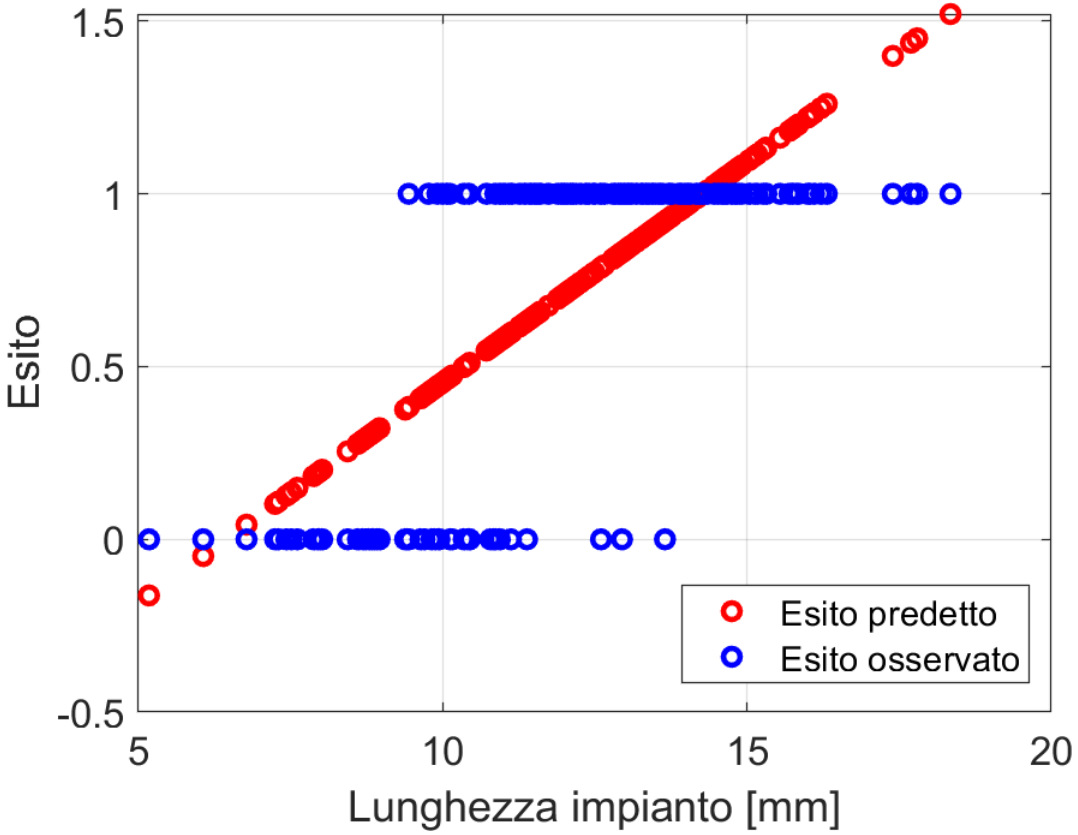
- Codifichiamo la variabile di outcome con una variabile 0/1.

$$Y = \begin{cases} 0 & \text{fallimento dell'impianto} \\ 1 & \text{successo dell'impianto} \end{cases}$$

- Proviamo a identificare un modello di regressione lineare semplice:

$$Y = \beta_0 + \beta \cdot X + \varepsilon$$

- Confrontiamo le predizioni del modello $\hat{Y}$ con i valori di Y. Che problema osserviamo?



- Vorremmo predire una quantità che vale 0 o 1.
- La regressione lineare predice una quantità che varia linearmente in un range illimitato che eccede l'intervallo [0 1].

MODELLI LINEARI GENERALIZZATI

- Il modello di regressione lineare non è appropriato per descrivere relazioni che coinvolgono una variabile dipendente di tipo qualitativo.
- Soluzione: **modelli lineari generalizzati** (*generalised linear models* o GLM)
- Viene applicata una funzione g detta **link function** per mappare il dominio (-∞, +∞) della regressione lineare su di un range limitato desiderato.

$$E(Y) = g^{-1}(\boldsymbol{\beta}^T \boldsymbol{X})$$

$$g(E(Y)) = \boldsymbol{\beta}^T \boldsymbol{X}$$

- Se siamo interessati a predire una quantità che è 0 o 1, potremmo scegliere una funzione g tale che $g^{-1}(\boldsymbol{\beta}^T \boldsymbol{X}) \in [0\ 1]$.

REGRESSIONE LOGISTICA

- La **regressione logistica** è un particolare tipo di GLM che impiega una funzione di link di tipo **logit** per mappare l'esito di una regressione lineare, compreso nell'intervallo (-∞, +∞), in una quantità compresa in [0 1].
- Funzione logit:

$$g(x) = \log\left(\frac{x}{1-x}\right), \qquad g^{-1}(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$

- Equazione della regressione logistica:

$$\log\left(\frac{p}{1-p}\right) = \boldsymbol{\beta}^T \boldsymbol{X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$$

$$p = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{X}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{X}}} = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m}}$$

IL SIGNIFICATO DEL VALORE PREDETTO DALLA REGRESSIONE LOGISTICA

- Problema iniziale: vogliamo stabilire se sussiste una relazione tra un set di variabili indipendenti, $X_1, X_2, \dots, X_m$, e una variabile di outcome Y, binaria, i cui valori sono rappresentati con 0 e 1.

- Applichiamo la regressione logistica:

$$\log\left(\frac{p}{1-p}\right) = \boldsymbol{\beta}^T \boldsymbol{X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$$

$$p = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{X}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{X}}} = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m}}$$

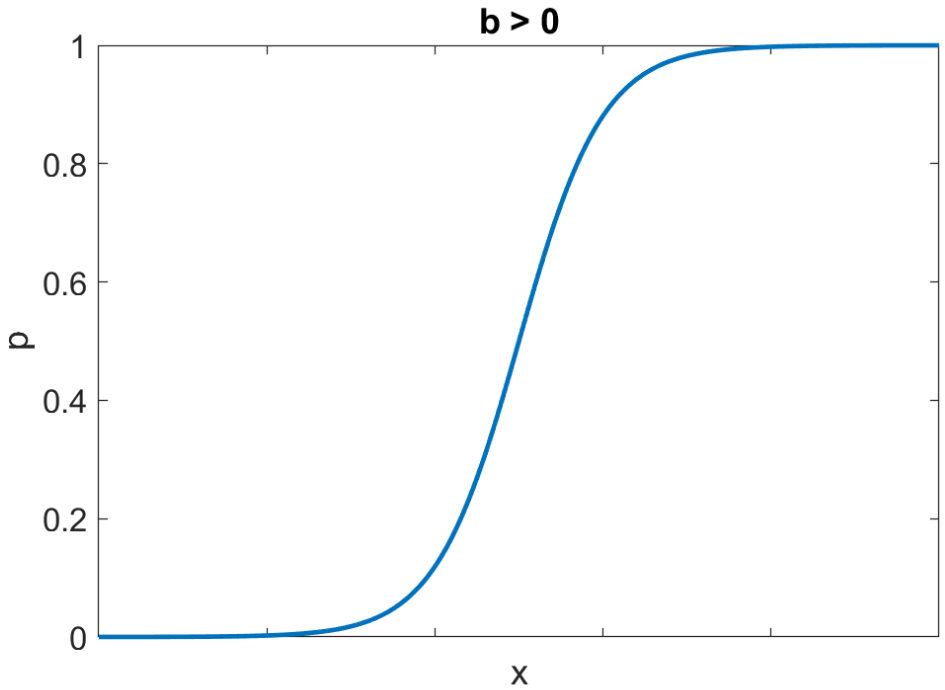
L'uscita p della regressione logistica è una quantità compresa in [0, 1] che rappresenta la probabilità che Y sia pari a 1:

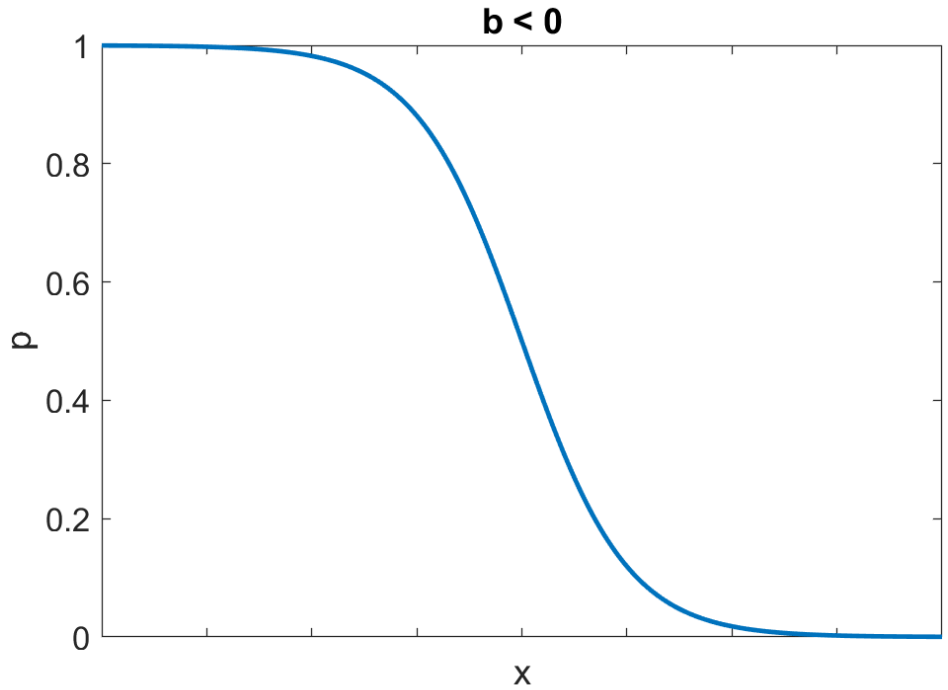
$$\boldsymbol{p = P(Y = 1 | X, \beta)}$$

REGRESSIONE LOGISTICA SEMPLICE

- Com'è fatta la quantità p predetta dalla regressione logistica?

$$p = \frac{e^{\beta_0+\beta\cdot X}}{1+e^{\beta_0+\beta\cdot X}}$$



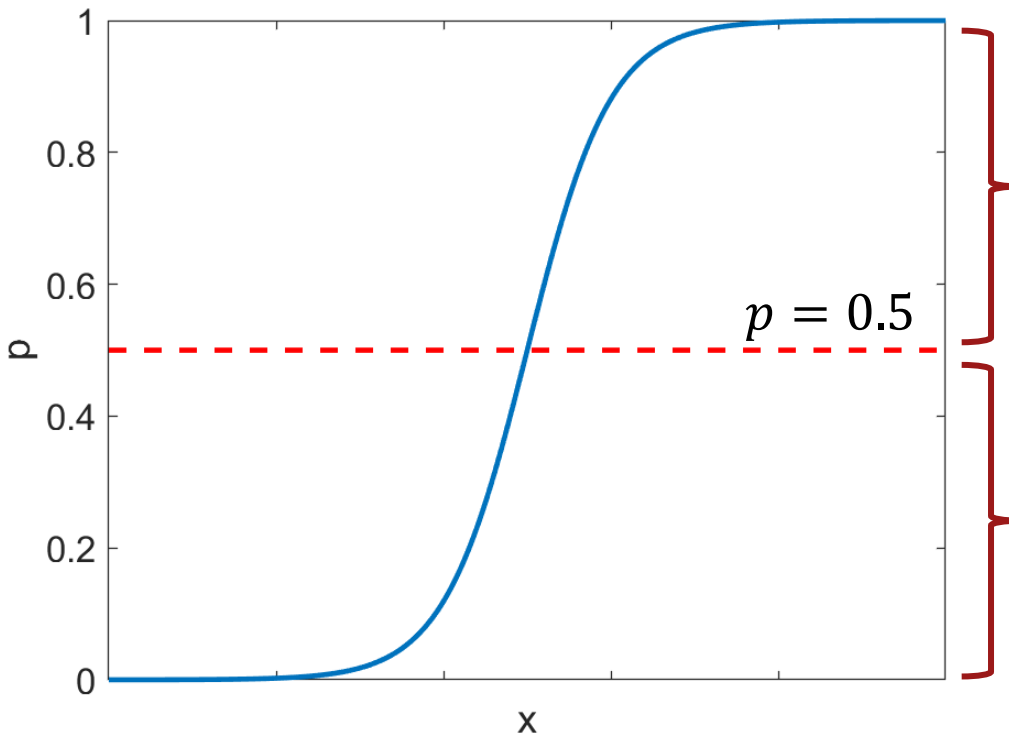


- Se vogliamo che il modello predica il valore di Y (0 o 1) come facciamo?

SOGLIA DI CLASSIFICAZIONE

- Applichiamo una soglia th sul valore di p:
  - Se $p < th \rightarrow \hat{Y} = 0$
  - Se $p \geq th \rightarrow \hat{Y} = 1$

$$p = P(Y = 1|X, \beta)$$



$p \geq 0.5 \rightarrow \hat{Y} = 1$
Prediciamo la classe 1

$p < 0.5 \rightarrow \hat{Y} = 0$
Prediciamo la classe 0

Di fatto il problema che approcciamo con la regressione logistica è un problema di **classificazione binaria**: cerchiamo di costruire un modello matematico per predire il valore di una variabile binaria Y usando delle variabili esplicative $X_1, X_2, \ldots, X_m$.

STIMA DEI COEFFICIENTI DELLA REGRESSIONE LOGISTICA

Dataset (training set)

	X_1	X_2	…	X_m	Y
Oss. 1					0
Oss. 2					1
…					…
Oss. n					0

Stima dei coefficienti mediante il **metodo di massima verosimiglianza***

Stime dei coefficienti del modello

$$\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_m$$

Equazione del modello

$$\hat{p} = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2+\cdots+\hat{\beta}_m X_m}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2+\cdots+\hat{\beta}_m X_m}}$$

*Il metodo della massima verosimiglianza stima i valori dei parametri β che massimizzano la probabilità a posteriori (noti i β) dei valori osservati per l'outcome.

LA FUNZIONE DI VEROSIMIGLIANZA

- $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_m]^{\mathrm{T}}$ → vettore dei parametri incogniti della regressione logistica
- $\mathbf{Y} = [\mathrm{Y}_1, \mathrm{Y}_2, \dots, \mathrm{Y}_\mathrm{n}]^{\mathrm{T}}$ → vettore aleatorio che rappresenta le n osservazioni di Y
- $\boldsymbol{y} = [y_1, y_2, \dots, y_n]^{\mathrm{T}}$ → vettore dei valori osservati per l'outcome Y (realizzazione di **Y**)
- $\boldsymbol{x_i} = [x_{i1}, \dots, x_{im}]^{\mathrm{T}}$ → vettore dei valori osservati per le variabili indipendenti in corrispondenza di y_i
- $\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}$ → Matrice avente sulle colonne i valori delle variabili indipendenti e una colonna di 1 in corrispondenza dell'intercetta

- **Funzione di verosimiglianza ("likelihood"):**

$$L(\boldsymbol{\beta}|\boldsymbol{Y} = \boldsymbol{y}) = \mathrm{P}(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}) = \mathrm{P}(Y_1 = y_1|\boldsymbol{x_1}, \boldsymbol{\beta}) \cdot \mathrm{P}(Y_2 = y_2|\boldsymbol{x_2}, \boldsymbol{\beta}) \cdot \cdots \cdot \mathrm{P}(Y_m = y_m|\boldsymbol{x_m}, \boldsymbol{\beta})$$

$$L(\boldsymbol{\beta}|\boldsymbol{y}) = \prod_{i=1}^{n} \mathrm{P}(Y_i = y_i|\boldsymbol{x_i}, \boldsymbol{\beta})$$

STIMATORE DI MASSIMA VEROSIMIGLIANZA

Noti i valori delle variabili esplicative $\boldsymbol{x_i}$, la funzione di verosimiglianza è una funzione dei parametri incogniti $\boldsymbol{\beta}$.

Stimatore di massima verosimiglianza:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, L(\boldsymbol{\beta}|\boldsymbol{y}) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \prod_{i=1}^{n} \mathrm{P}(Y_i = y_i|\boldsymbol{x_i}, \boldsymbol{\beta})$$

Cerchiamo il set di parametri $\boldsymbol{\beta}$ che massimizza la verosimiglianza dei valori osservati di Y, ovvero la probabilità dei valori osservati di Y dati i valori osservati per le variabili esplicative e $\boldsymbol{\beta}$.

STIMATORE DI MASSIMA VEROSIMIGLIANZA

BONUS

Equivalentemente possiamo massimizzare il logaritmo della funzione di verosimiglianza (i prodotti diventano somme e il calcolo è più semplice):

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\ \log(L(\boldsymbol{\beta}|\boldsymbol{y})) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\ \sum_{i=1}^{n} \log(\mathrm{P}(Y_i = y_i|\boldsymbol{x_i}, \boldsymbol{\beta}) =$$

$$= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{i=1}^{n} (y_i \cdot \log(\mathrm{P}(Y_i = 1|\boldsymbol{x_i}, \boldsymbol{\beta}))+ (1 - y_i) \cdot \log(\mathrm{P}(Y_i = 0|\boldsymbol{x_i}, \boldsymbol{\beta}))) =$$

$$= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{i=1}^{n} \left(y_i \cdot \log(\frac{\mathrm{P}(y_i = 1|\boldsymbol{x_i}, \boldsymbol{\beta})}{\mathrm{P}(y_i = 0|\boldsymbol{x_i}, \boldsymbol{\beta})}) + \log(\mathrm{P}(y_i = 0|\boldsymbol{x_i}, \boldsymbol{\beta})) \right)$$

STIMATORE DI MASSIMA VEROSIMIGLIANZA

BONUS

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \sum_{i=1}^{n} \left(y_i \cdot \left(\beta_0 + \sum_{j=1}^{m} \beta_j \, x_{ij} \right) - \log\left(1 + e^{\beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}}\right) \right)$$

Questo problema di ottimizzazione non presenta soluzione in forma chiusa (non si riesce a ricavare una formula per calcolare $\widehat{\boldsymbol{\beta}}$). La soluzione viene cercata mediante algoritmi di ottimizzazione iterativi.

In pratica, i vari software per l'analisi statistica mettono a disposizione funzioni che risolvono questo problema di stima per noi.

RISULTATI DEL PROCESSO DI STIMA

- **Stime dei coefficienti:** $\hat{\beta}_j, j = 0, \dots, m$
- **Standard error** sulle stime dei coefficienti: $SE_j, j = 0, \dots, m$

- **Intervallo di confidenza** al 95% sulle stime dei parametri:

$$\hat{\beta}_j \pm 1.96 \cdot SE_j, \qquad j = 0, \dots, m$$

- **Odds ratio** delle stime dei coefficienti:

$$OR_j = e^{\hat{\beta}_j}$$

- Intervallo di confidenza al 95% sugli odds ratio:

$$e^{\hat{\beta}_j \pm 1.96 \cdot SE_j}$$

ESEMPIO (1/2)

- Vogliamo investigare se sussiste una relazione tra il successo/fallimento di un impianto dentale (Y) e due variabili esplicative: la lunghezza dell'impianto dentale (X_1) e l'età del paziente al momento dell'impianto (X_2).
- Disponiamo di una dataset contenente 200 osservazioni per le 3 variabili in gioco.
  - Per 50 osservazioni l'esito è fallimentare (Y=0)
  - Per 150 osservazioni l'esito è successo (Y=1)
- Applichiamo il modello di regressione logistica
- Equazione del modello:

$$p = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}$$

ESEMPIO (2/2)

- Risultato della stima di massima verosimiglianza:

Variabile	Stima del coefficiente $\hat{\beta}_j$	Standard error SE_j	Intervallo di confidenza al 95% $\hat{\beta}_j \pm 1.96 \cdot SE_j$	Odds ratio $e^{\hat{\beta}_j}$	Intervallo di confidenza al 95% $e^{\hat{\beta}_j \pm 1.96 \cdot SE_j}$
Intercetta	-15.437	2.9884	[-21.29 -9.58]	1.9756×10^{-7}	[5.6×10^{-10} 6.9×10^{-5}]
Lunghezza impianto [mm]	1.5842	0.2559	[1.08 2.09]	4.8752	[2.95 8.05]
Età [anni]	-0.0231	0.0371	[-0.096 0.048]	0.9771	[0.91 1.05]

VALUTAZIONE DELLA BONTA' DEL MODELLO

- Valutazione dell'errore di classificazione
- Deviance e likelihood ratio test

ERRORE DI CLASSIFICAZIONE

Data una soglia th per la classificazione, dobbiamo confrontare:

- $\hat{y}_i$: valori dell'outcome predetti dal modello (0 o 1)
  - $\hat{y}_i = 1$ → valore predetto **positivo**
  - $\hat{y}_i = 0$ → valore predetto **negativo**
- y_i: valori reali dell'outcome (0 o 1)
- **Matrice di confusione:**

	# valori positivi $y_i = 1$	**# valori negativi** $y_i = 0$
# valori predetti positivi $\hat{y}_i = 1$	# VERI POSITIVI o TRUE POSITIVES (TP)	# FALSI POSITIVI o FALSE POSITIVES (FP)
# valori predetti negativi $\hat{y}_i = 0$	# FALSI NEGATIVI o FALSE NEGATIVES (FN)	# VERI NEGATIVI o TRUE NEGATIVES (TN)

METRICHE DI CLASSIFICAZIONE

- **Accuratezza:** frazione di predizioni corrette sul totale dei valori predetti

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

- **Sensibilità** (o Recall): frazione di valori positivi correttamente predetti

$$Sensitivity = \frac{TP}{TP + FN}$$

- **Specificità:** frazione di valori negativi correttamente predetti

$$Specificity = \frac{TN}{TN + FP}$$

- **Precisione** (o positive predictive value): frazione di predizioni positive corrette

$$Precision = \frac{TP}{TP + FP}$$

	$\mathbf{y_i = 1}$	$\mathbf{y_i = 0}$
$\mathbf{\hat{y}_i = 1}$	TP	FP
$\mathbf{\hat{y}_i = 0}$	FN	TN

Tutte queste metriche presentano valori compresi tra 0 e 1 e il loro valore è tanto migliore quanto più è vicino a 1.

METRICHE DI CLASSIFICAZIONE PER IL CLASSIFICATORE RANDOM

Un modello che assegna i valori predetti di Y a caso avrà le seguenti metriche di classificazione:

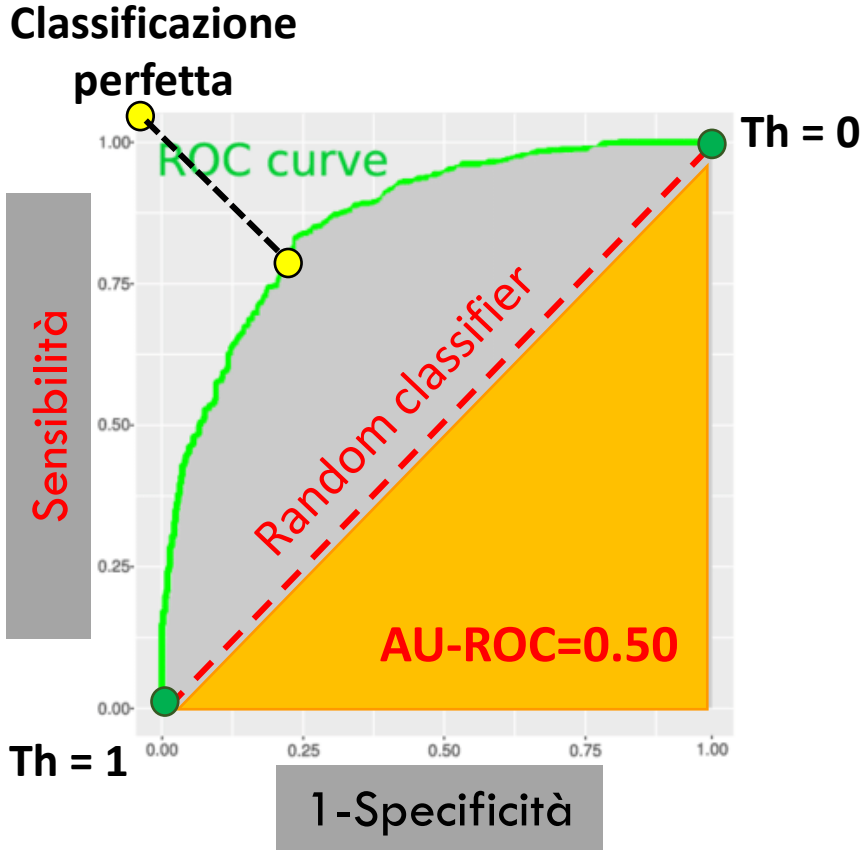
- Accuratezza = 0.5
- Sensibilità = 0.5
- Specificità = 0.5
- Precisione = (# osservazioni con Y =1)/# totale di osservazioni

SCELTA DELLA SOGLIA DI CLASSIFICAZIONE

- Come scegliamo la soglia di classificazione th?
- La scelta più intuitiva è 0.5, ma questa potrebbe non essere la scelta ottimale, specie se la prevalenza delle classi è sbilanciata (tante più osservazioni con Y=0 rispetto a Y=1 o viceversa).
- Per una scelta ottimale si possono testare diversi valori di th e scegliere il migliore sulla base delle metriche di classificazione.
- All'aumentare di th, tipicamente sensibilità diminuisce e specificità aumenta
  - Se $th = 0$ → tutti i valori predetti sono positivi → sensitività = 1, specificità = 0
  - Se $th = 1$ → tutti i valori predetti sono negativi → sensitività = 0, specificità = 1
- Tipicamente si cerca il valore di th per cui si ha un buon compromesso tra sensibilità e specificità.

LA CURVA ROC

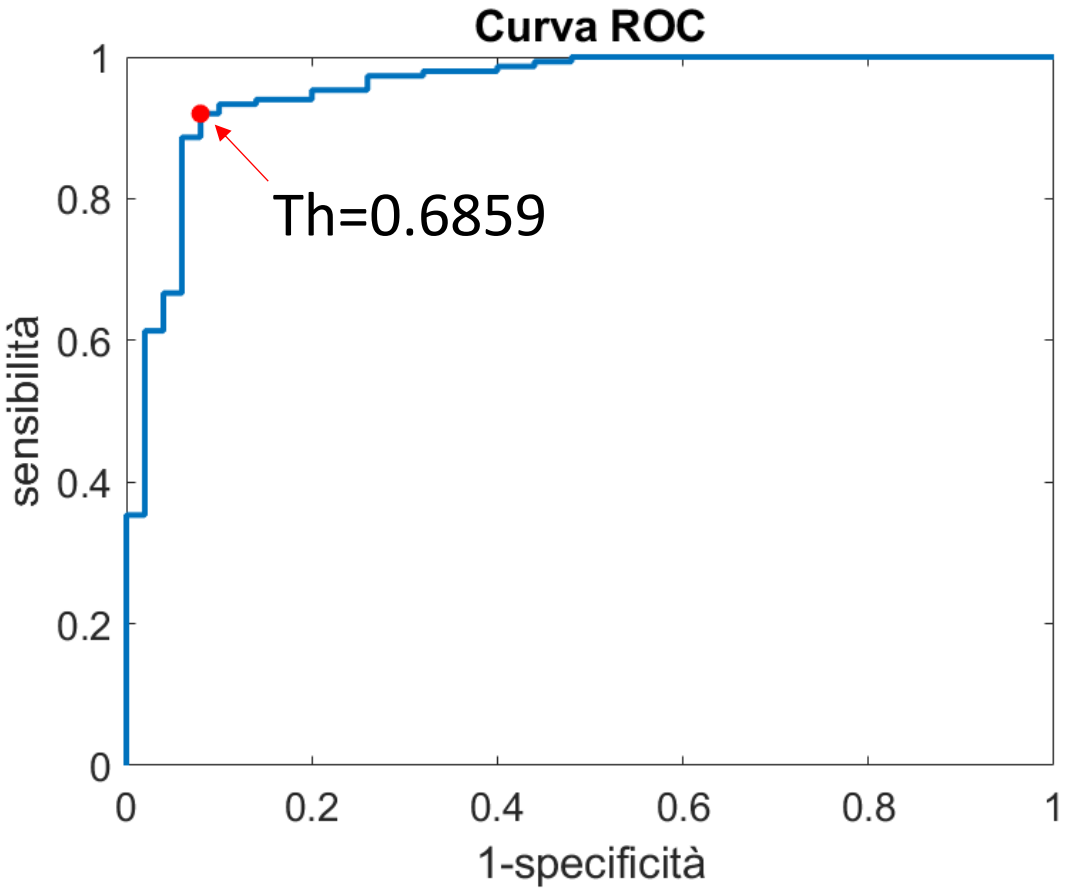
- **Curva ROC** (*Receiver Operating Characteristic* curve): grafico che mostra sensibilità vs. 1-specificità al variare di th.



- Possibile soglia ottima: il valore di th per cui la distanza della curva ROC dall'angolo in alto a sinistra è minima.
- L'**area sotto la curva ROC (AU-ROC)** è una metrica di classificazione indipendente da th
  - Valore compreso tra 0 e 1, tanto più è vicino a 1, tanto più il modello tende ad assegnare valori di probabilità elevati ad osservazioni per cui Y=1.
  - Il classificatore che assegna i valori predetti in modo random presenta AU-ROC=0.5.

ESEMPIO

- Valutiamo l'errore di classificazione per il modello dell'esempio precedente.



AU-ROC=0.9557

ESEMPIO

- Metriche di classificazione in corrispondenza della soglia ottima:

Matrice di confusione

	$y_i = 1$	$y_i = 0$
$\hat{y}_i = 1$	137	4
$\hat{y}_i = 0$	13	46

- Accuratezza: (137+46)/200 = 0.915
- Sensibilità: 137/(137+13) = 0.9133
- Specificità: 46/(46+4) = 0.92
- Precisione = 137/(137+4)=0.9716

DEVIANCE

- **Deviance** del modello:

$$D_m = -2 \cdot \log(L(\hat{\boldsymbol{\beta}}|\boldsymbol{y}))$$

- La deviance è una misura di quanto bene il modello descrive i dati (stesso ruolo di SSE per la regressione lineare).

LIKELIHOOD RATIO TEST (1/2)

- Il **likelihood ratio test** è un test statistico per valutare se il modello di regressione logistica considerato è significativamente diverso da quello nullo, ovvero quello che include solo l'intercetta.
- Sistema di ipotesi:
  - H_0: $\beta_1 = \beta_2 = \cdots = \beta_m = 0$ → il modello trovato è uguale a quello nullo
  - H_1: almeno uno dei coefficienti $\beta_j, j = 1, \dots, m$ è diverso da 0 → il modello trovato è diverso da quello nullo
- Statistica del test:

$$D = D_0 - D_m = -2 \cdot \log\left(\frac{likelihood\ del\ modello\ nullo}{L(\hat{\boldsymbol{\beta}}|\boldsymbol{y})}\right)$$

dove D_0 è la deviance del modello nullo.

LIKELIHOOD RATIO TEST (2/2)

- Se vale H_0, D è distribuita come una variabile aleatoria χ^2 avente m gradi.
- Livello di significatività α
- Regola decisionale:
  - $D > \chi^2_{\alpha,m}$ → rifiutiamo H_0 → il modello è significativamente diverso da quello nullo
  - $D \leq \chi^2_{\alpha,m}$ → non possiamo rifiutare H_0

ESEMPIO

Applichiamo il likelihood ratio test al modello dell'esempio precedente.

- Log-likelihood del modello identificato: $\log(L) = -44.42$
- Deviance del modello identificato: $D_m = -2 \cdot \log(L) = 88.84$
- Deviance del modello nullo: $D_0 = 224.93$
- Statistica del test:

$$D = D_0 - D_m = 224.93 - 88.84 = 136.09$$

- Livello di significatività $\alpha = 0.05$
- Valore critico:
  - $\chi^2_{\alpha,m} = \chi^2_{0.05,2} = 5.99$
- Conclusione: $D > 5.99$ → rifiutiamo H_0 → il modello è significativamente diverso da quello nullo → sussiste in effetti una relazione significativa tra le variabili esplicative considerate (lunghezza impianto ed età) e l'esito dell'impianto

INTERPRETAZIONE DEI COEFFICIENTI DELLA REGRESSIONE LOGISTICA

- β_0: intercetta → quando tutte le variabili indipendenti sono nulle la probabilità della classe 1 è $p = \frac{e^{\beta_0}}{1+e^{\beta_0}}$
- β_j rappresenta di quanto aumenta la funzione logit di p quando la variabile X_j aumenta di una unità e tutte le altre variabili esplicative sono costanti.
- Segno di β_j:
  - Se $\beta_j > 0$ o $e^{\beta_j} > 1$ → Se X_j aumenta, anche la probabilità che Y sia 1 aumenta.
  - Se $\beta_j < 0$ o $e^{\beta_j} < 1$ → Se X_j aumenta, la probabilità che Y sia 1 diminuisce.
- Valore assoluto di β_j:
  - Se $\beta_j = 0$ o $e^{\beta_j} = 1$ → X_j non ha effetto sulla probabilità che Y sia 1.
  - Se β_j è significativamente diverso da 0 → X_j ha un impatto importante sulla probabilità che Y sia 1.

SIGNIFICATIVITA' STATISTICA DEI COEFFICIENTI

Wald test

- Sistema di ipotesi:
  - H_0: $\beta_j = 0$
  - H_1: $\beta_j \neq 0$
- Statistica del test:

$$\text{Z-score del coefficiente } \beta_j \longrightarrow Z_j = \frac{\hat{\beta}_j}{SE_j}$$

- Quando vale H_0, z_j ha distribuzione normale standard.
- Regola decisionale:
  - $|Z_j| > z_{\frac{\alpha}{2}}$ → rifiutiamo H_0
  - $|Z_j| \leq z_{\frac{\alpha}{2}}$ → non possiamo rifiutare H_0

ESEMPIO

- Come interpreteresti le stime dei coefficienti del modello dell'esempio precedente alla luce del risultato del Wald test?

Variabile	Stima del coefficiente $\hat{\beta}_j$	Standard error SE_j	Z-score z_j*	P-value del Wald test
Intercetta	-15.437	2.9884	-5.1657	2.40×10^{-7}
Lunghezza impianto [mm]	1.5842	0.2559	6.1917	5.95×10^{-10}
Età [anni]	-0.0231	0.0371	-0.6221	0.53

*Valore critico per $\alpha = 5\% \rightarrow z_{0.025} = 1.96$

CONFRONTO TRA MODELLI IN COMPETIZIONE

- Log-likelihood ratio test:

$$D = D_{modello\ 1} - D_{modello\ 2} = -2 \cdot \log\left(\frac{likelihood\ modello\ 1}{likelihood\ modello\ 2}\right)$$

Quando vale H_0, D ha distribuzione χ^2 con gradi di libertà pari alla differenza tra il numero di parametri del modello 1 e il numero di parametri del modello 2.

- Indici di parsimonia:

$$AIC = -2 \cdot \log\left(L\left(\hat{\boldsymbol{\beta}}|\boldsymbol{y}\right)\right) + 2 \cdot p$$

$$BIC = -2 \cdot \log\left(L\left(\hat{\boldsymbol{\beta}}|\boldsymbol{y}\right)\right) + p \cdot \log(n)$$

QUIZ 1

Nel modello di regressione logistica avente equazione:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1$$

La variabile p è:

- una quantità che può assumere due soli valori, 0 o 1, che rappresentano i due livelli della variabile di outcome, Y.
- una quantità che può assumere valori in [0, 1] che rappresenta la probabilità che la variabile di outcome, Y, sia pari a 1 (classe 1)
- una quantità che può assumere valori in [0, 1] che rappresenta la probabilità che la variabile di outcome, Y, sia pari a 0 (classe 0)
- un iperparametro del modello che deve essere ottimizzato.

DOMANDA APERTA

- Quali metriche conosci per valutare l'errore di classificazione di un modello di regressione logistica? Che approccio potresti utilizzare per scegliere un valore ottimo per la soglia di classificazione?