

Metodi Matematici e Statistici per le Decisioni Economiche

Corso di Laurea Magistrale in Management, Imprese e Mercati
2025-2026

Vincenzo Nardelli



vincenzo.nardelli@unicatt.it

Domanda 1

Utilizza il dataset College del pacchetto ISLR2 in R per eseguire una selezione delle variabili significative tra Room.Board (costo camera/mensa), Outstate (tasse scolastiche) e PhD (percentuale di docenti con PhD) al fine di costruire un modello logistico predittivo per Private (college privato o pubblico). Stima il modello con tutte le variabili e rimuovi quelle non significative per rispondere alle domande successive.

Descrizione delle Variabili:

- ▶ Private: Variabile binaria (Yes per privato, No per pubblico). Variabile dipendente.
- ▶ Room.Board: Costo per stanza e mensa.
- ▶ Outstate: Tasse scolastiche per studenti fuori stato.
- ▶ PhD: Percentuale di docenti con PhD.

Domanda 1 - Domande

1. Calcola gli odds e spiega il significato.
2. Come si valuta la bontà del modello? Giustifica numericamente utilizzando metriche come devianza residua, AIC e altre misure di adattamento.
3. Calcola le seguenti metriche del modello finale, considerando le università private come classe positiva (Yes): Accuratezza, Sensitività e Specificità.
4. Sono maggiori i falsi positivi (FP) o i falsi negativi (FN)? Quale è l'impatto pratico di un errore nella classificazione di un college come privato o pubblico?

Domanda 1 - Modello Finale

Dopo aver stimato il modello completo, la variabile `Room.Board` non è significativa (p-value = 0.804), quindi viene rimossa.

Modello finale:

$$\text{logit}(P(\text{Private} = \text{Yes})) = \beta_0 + \beta_1 \cdot \text{Outstate} + \beta_2 \cdot \text{PhD}$$

Variabile	Coeff.	Std. Error	p-value
Intercetta	2.3663	0.7522	0.00166
Outstate	0.0008603	0.0000662	< 2e-16
PhD	-0.1225	0.0127	< 2e-16

Risposta 1: Odds Ratio

Gli **odds ratio** si calcolano come e^{β_i} :

Variabile	Odds Ratio	Interpretazione
Intercetta	10.658	Odds di base
Outstate	1.0009	+0.09% per ogni \$ in più
PhD	0.885	-11.5% per ogni punto % in più

Interpretazione sostanziale:

- ▶ **Outstate** (positivo): college con tasse più alte hanno maggiori probabilità di essere privati
- ▶ **PhD** (negativo): più docenti con PhD è associato a minori probabilità di essere privato (università pubbliche di ricerca)

Risposta 2: Bontà del Modello

a) Devianza Residua:

- ▶ Null Deviance: 910.75
- ▶ Residual Deviance: 415.90
- ▶ Riduzione: 494.85 (54.3%)

b) AIC (Akaike Information Criterion):

- ▶ AIC modello completo: 423.84
- ▶ AIC modello finale: 421.9 (migliore)

c) Test del Rapporto di Verosimiglianza:

$$G^2 = 910.75 - 415.90 = 494.85 \quad (\text{p-value} \approx 0)$$

Conclusion: Modello robusto con ottimo potere predittivo.

Risposta 3: Metriche di Performance

Matrice di Confusione (threshold = 0.5):

		Predetto	
		No	Yes
Reale	No	166 (TN)	46 (FP)
	Yes	42 (FN)	523 (TP)

Calcolo delle Metriche:

$$\text{Accuratezza} = \frac{166 + 523}{777} = \mathbf{0.8867} \text{ (88.67\%)}$$

$$\text{Sensibilità} = \frac{523}{523 + 42} = \mathbf{0.9257} \text{ (92.57\%)}$$

$$\text{Specificità} = \frac{166}{166 + 46} = \mathbf{0.7830} \text{ (78.30\%)}$$

Risposta 4: Analisi degli Errori

Confronto: $FP = 46$, $FN = 42 \rightarrow FP > FN$ (differenza minima)

1. Falsi Negativi (FN = 42): College privato classificato come pubblico
Conseguenze:

- ▶ Studenti sottostimano significativamente i costi reali
- ▶ Problemi finanziari per le famiglie
- ▶ Possibile abbandono degli studi

Impatto: ALTO per studenti

2. Falsi Positivi (FP = 46): College pubblico classificato come privato
Conseguenze:

- ▶ Potenziali studenti scoraggiati dalla percezione di costi eccessivi
- ▶ Riduzione delle candidature
- ▶ Danno alla reputazione dell'istituzione

Impatto: MEDIO-ALTO per istituzioni

Domanda 2

Utilizza il dataset mtcars in R per eseguire una selezione forward delle variabili wt (peso), hp (cavalli) e disp (cilindrata) al fine di costruire un modello lineare predittivo per mpg (miglia per gallone).

Descrizione delle Variabili:

- ▶ mpg: Miglia per gallone (efficienza del carburante). Variabile dipendente.
- ▶ wt: Peso del veicolo (in migliaia di libbre). Un peso maggiore può ridurre l'efficienza del carburante.
- ▶ hp: Cavalli. Indica la potenza del motore. Più potenza può ridurre l'efficienza del carburante.
- ▶ disp: Cilindrata (capacità del motore in pollici cubici). Un motore più grande potrebbe influire sull'efficienza del carburante.

Domanda 2 - Risultati da fornire

1. Quali variabili significative sono state selezionate nel modello finale?
2. Qual è il valore minimo di RSE per il modello definitivo?
3. Qual è il valore di R-quadro del modello finale?
4. Interpreta i coefficienti del modello finale, specificando l'effetto di ogni variabile selezionata.

Selezione Forward - Step 1

Step 1: Selezione della prima variabile

Testiamo modelli con una singola variabile:

Modello	RSE	R ²	Selezionato
mpg ~ wt	3.046	0.7528	✓
mpg ~ hp	3.863	0.6024	
mpg ~ disp	3.251	0.7183	

Decisione: Il modello con **wt** ha il RSE più basso (3.046), quindi **wt** viene selezionata come prima variabile.

Selezione Forward - Step 2

Step 2: Selezione della seconda variabile

Aggiungiamo una seconda variabile al modello contenente **wt**:

Modello	RSE	R ²	Selezionato
mpg ~ wt + hp	2.593	0.8268	✓
mpg ~ wt + disp	2.917	0.7809	

Decisione: Il modello **wt + hp** ha RSE inferiore (2.593 vs 2.917), quindi **hp** viene aggiunta al modello.

Selezione Forward - Step 3

Step 3: Tentativo di aggiungere la terza variabile

Verifichiamo se l'aggiunta di `disp` migliora il modello:

Modello	RSE	R ²	p-value disp
<code>mpg ~ wt + hp</code>	2.593	0.8268	-
<code>mpg ~ wt + hp + disp</code>	2.639	0.8268	0.928

Decisione: L'aggiunta di `disp`:

- ▶ Peggiora l'RSE (aumenta da 2.593 a 2.639)
- ▶ Non aumenta l'R²
- ▶ Non è significativa (p-value = 0.928 \gg 0.05)

Modello finale: `mpg ~ wt + hp`

Risposta 1: Variabili Selezionate

Le variabili selezionate nel modello finale sono:

- ▶ **wt** (peso del veicolo)
- ▶ **hp** (potenza in cavalli)

La variabile *disp* (cilindrata) è stata **esclusa** perché non significativa e non migliora il modello.

Risposta 2: RSE Minimo

Il valore minimo dell'RSE (Residual Standard Error) per il modello definitivo è:

$$\mathbf{RSE = 2.593}$$

Interpretazione: Le previsioni del modello si discostano mediamente di ± 2.593 mpg dal valore reale.

Risposta 3: R-quadro

Il modello finale presenta:

$$R^2 = \mathbf{0.8268} \quad (\text{Multiple R-squared})$$

$$R^2_{\text{adj}} = \mathbf{0.8148} \quad (\text{Adjusted R-squared})$$

Interpretazione:

- ▶ L' $R^2 = \mathbf{0.8268}$ indica che il modello spiega l'**82.68%** della variabilità nell'efficienza del carburante (mpg)
- ▶ L' R^2 **aggiustato** = **0.8148** è più appropriato per confrontare modelli
- ▶ La differenza minima tra R^2 e R^2 aggiustato indica che il modello non è sovraparametrizzato

Risposta 4: Interpretazione Coefficienti

Il modello finale stimato è:

$$\widehat{\text{mpg}} = 37.23 - 3.88 \cdot \text{wt} - 0.032 \cdot \text{hp}$$

Variabile	Coeff.	Std. Error	t-value	p-value
Intercetta	37.23	1.60	23.29	$< 2 \times 10^{-16}$
wt	-3.88	0.63	-6.13	1.12×10^{-6}
hp	-0.032	0.009	-3.52	0.00145

Interpretazione Dettagliata dei Coefficienti

1. Intercetta ($\beta_0 = 37.23$):

- ▶ Valore atteso di mpg quando wt = 0 e hp = 0
- ▶ Non interpretabile nella pratica (baseline matematica)

2. Coefficiente di wt ($\beta_1 = -3.88$):

- ▶ Per ogni 1000 libbre di peso in più, l'efficienza *diminuisce* di 3.88 mpg (a parità di potenza)
- ▶ Altamente significativo (p-value $< 10^{-6}$)
- ▶ **Esempio:** Un'auto da 3000 lbs ha 3.88 mpg in meno di una da 2000 lbs (stessa potenza)

3. Coefficiente di hp ($\beta_2 = -0.032$):

- ▶ Per ogni cavallo in più, l'efficienza *diminuisce* di 0.032 mpg (a parità di peso)
- ▶ Significativo (p-value = 0.00145)
- ▶ **Esempio:** Un'auto con 200 hp ha 3.2 mpg in meno di una con 100 hp (stesso peso)

Confronto degli Effetti

Variabile	Effetto unitario	Effetto su 100 unità
wt (1000 lbs)	-3.88 mpg	-388 mpg
hp (1 cavallo)	-0.032 mpg	-3.2 mpg

Osservazione importante: L'effetto del **peso** è molto più marcato dell'effetto della **potenza**.

Implicazione pratica: Per migliorare l'efficienza del carburante, è più efficace ridurre il peso del veicolo che ridurre la potenza del motore.

Conclusione Domanda 2

La procedura di selezione forward ha identificato un modello:

- ▶ **Parsimonioso:** Solo 2 variabili su 3 candidate
- ▶ **Statisticamente robusto:** Entrambe le variabili altamente significative
- ▶ **Ad alto potere esplicativo:** $R^2 = 0.8268$ (spiega oltre l'82% della variabilità)
- ▶ **Fisicamente interpretabile:** I coefficienti negativi confermano che peso e potenza riducono l'efficienza del carburante

Test F: $F = 69.21$ su $(2, 29)$ gradi di libertà con $p\text{-value} = 9.11 \times 10^{-12}$

Il modello è significativamente migliore del modello nullo.