APPLICAZIONE DI UN METODO DI VALUTAZIONE PER DIVERSE TECNICHE DI IMPUTAZIONE MULTIPLA IN UN DATABASE CLINICO

Ciotta Alessandro[1], Pescini Dario[2,3], Soranna Davide[1], Lombardi Carolina[4,5], Zambon Antonella[1,2]

1 Biostatistics Unit, IRCCS Istituto Auxologico Italiano, Milan, Italy

2 Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

3 SYSBIO Centre of Systems Biology, Milan, Italy.

4 Sleep Disorders Center & Department of Cardiovascular, Neural and Metabolic Sciences, IRCCS Istituto Auxologico Italiano, Milan, Italy.

5 Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy

Introduzione

Gli studi multicentrici collaborativi dispongono generalmente di una grande mole di dati ma sono caratterizzati da un'alta frequenza di valori mancanti. L'approccio tradizionale prevede di escludere i soggetti con valori mancanti dall'analisi finale (complete case analysis) riducendo la sample size e quindi aumentando l'errore di secondo tipo e la probabilità di introdurre selection bias. Un approccio alternativo per la gestione dei valori mancanti consiste nell'imputazione ovvero nel sostituire un valore plausibile al dato mancante. Tra i vari metodi di imputazione quello che ha ricevuto maggiore attenzione negli ultimi anni è l'imputazione multipla ovvero l'imputazione per ogni valore mancante di una serie di valori plausibili. Tale approccio consente di integrare nelle stime finali l'incertezza legata al valore imputato. Ad oggi esistono numerosi approcci per effettuare l'imputazione multipla, ma spesso negli studi clinici non si pone sufficiente attenzione alla qualità dell'imputazione prodotta e non si comparano i diversi metodi nella capacità di produrre stime dell'effetto attendibili per il setting di dati disponibile. Per affrontare questa problematica Casiraghi et al. [1] hanno proposto la seguente strategia che considera un sottoinsieme di casi completi su cui si effettuano delle amputazioni (simulazioni di dati mancanti) ad hoc. Si confrontano quindi le stime dell'effetto calcolate sui casi completi (gold standard) con quelle ottenute dai diversi metodi di imputazione.

Obiettivi

L'obiettivo di questo studio consiste nel valutare diversi algoritmi di imputazione multipla, applicando l'approccio di proposto da Casiraghi et al. tramite la costruzione di una pipeline in Python per automatizzarne i passaggi e verificarne le performance nel contesto del progetto European Sleep Apnea Database (ESADA).

Metodi

Per la costruzione della pipeline e la verifica della sua performance si è fatto riferimento ai dati e ai risultati di una pubblicazione relativa al progetto ESADA che si proponeva di indagare la relazione tra gravità di apnee notturne e ipertensione [2]. In questa pubblicazione le analisi erano state condotte solo sui complete cases. Applicando la procedura di Casiraghi abbiamo introdotto in modo controllato valori

mancanti all'interno dei complete cases creandone molteplici versioni amputate simulando un meccanismo missing at random con il pacchetto missmethods di R. Specificatamente, per ogni variabile si è simulato un 10% di valori mancanti e questa operazione è stata replicata 20 volte, creando altrettanti dataset amputati. Successivamente su ognuno dei dataset amputati si sono applicati i seguenti metodi di imputazione multipla: Amelia che utilizza un approccio Joint Modeling (JM), Mice che usa il Fully Conditional Specification (FCS) e MissRanger che è una tecnica basata sul machine learning (ML). La pipeline, realizzata in Python, richiama questi metodi implementati tramite il software R permettendo l'ottimizzazione dei parametri da cui sono caratterizzati, ad esempio la tollerance per Amelia, il numero di iterazioni per Mice e il numero di neighbors e di alberi decisionali in missRanger. Infine, dopo aver effettuato il pooling delle stime ottenute dalle varie versioni amputate queste sono state confrontate con le stime ottenute sui complete cases (gold standard). Per il confronto fra le stime si è utilizzato il percent bias (PB) [3].

Risultati

La coorte di partenza era costituita da 12485 soggetti ma i casi completi erano solo 6018 (48%). Le stime gold standard e quelle ottenute dai diversi metodi di imputazione dell'associazione (β e standard error) tra le variabili di interesse e l'ipertensione sono mostrate in tabella.

effetto	Gold Standard β (SE)	Amelia β (SE)	Mice β (SE)	MissRanger β (SE)
OSA lieve	0.464 (0.08)	0.46 (0.084)	0.462 (0.08)	0.441 (0.116)
OSA moderata	0.622 (0.088)	0.617 (0.094)	0.622 (0.088)	0.604 (0.114)
Età (>60)	1.042 (0.092)	1.005 (0.078)	1.038 (0.073)	0.988 (0.084)
BMI (>30)	0.315 (0.073)	0.316 (0.084)	0.313 (0.08)	0.315 (0.083)
BPCO	0.092 (0.08)	0.108 (0.149)	0.086 (0.142)	0.14 (0.147)
Diabete tipo 2	0.839 (0.012)	0.79 (0.112)	0.825 (0.104)	0.79 (0.122)
ESS (>10)	-0.232 (0.459)	-0.229 (0.073)	-0.242 (0.069)	-0.208 (0.072)
Iperlipidemia	0.964 (0.078)	0.946 (0.08)	0.971 (0.075)	0.916 (0.087)
Circonferenza collo	0.065 (0.069)	0.06 (0.013)	0.066 (0.012)	0.058 (0.013)
Genere (M)	-0.654 (0.103)	-0.609 (0.099)	-0.662 (0.092)	-0.589 (0.098)
Fumo	-0.265 (0.075)	-0.247 (0.085)	-0.259 (0.078)	-0.253 (0.085)
Rapporto vita/fianchi	1.753 (0.142)	1.774 (0.49)	1.738 (0.461)	1.751 (0.484)
PB	-	4.36%	1.64%	8.98%

L'algoritmo di imputazione migliore risulta Mice (con PB pari a 1.64%). La versione ottimizzata di questo algoritmo è stata poi applicata al dataset dei 12485 soggetti. Le stime degli effetti calcolate dopo la multiple imputation non risultano troppo diverse da quelle dei complete cases nei valori puntuali ma sono caratterizzate da un maggiore precisione (OSA lieve: OR 1.53 (95% CI 1.38-1.70), OSA moderata: OR 1.58 (95% CI 1.40-1.78) paragonate alle stime Gold Standard: 1.59 (1.36-1.86) e 1.86 (1.57-2.21)).

Conclusione

Nello specifico contesto del progetto ESADA l'approccio FCS di Mice si è dimostrato il metodo di imputazione meno distorto. I diversi metodi di imputazione possono avere performance diverse all'interno

dello stesso setting di dati e quindi si ritiene opportuno far precedere l'applicazione del metodo di imputazione da una fase valutativa preliminare al fine di selezionare il metodo più performante.

Bibliografia

[1] Casiraghi E., Wong R., Hall M. et al. ≪A method for comparing multiple imputation techniques: A case study on the US national COVID cohort collaborative≫. In: Journal of Biomedical Informatics (2023).

[2] Bouloukaki I., Grote L., T.McNicholas W. et al. ≪Mild obstructive sleep apnea increases hypertension risk, challenging traditional severity classification≫. In: Journal of Clinical Sleep Medicine (2020).

[3] Smith, M. W., et al., ≪ Recommendations for catch-curve analysis. ≫. In: North American Journal of Fisheries Management (2012)