Analisi Multinomiale su Diagnosi Alzheimer: Confronto tra Tecniche di Bilanciamento

# Obiettivo

L'obiettivo di questo studio è valutare l'effetto delle misure volumetriche cerebrali sulla probabilità di appartenenza alle tre categorie diagnostiche: CN (Controlli sani), MCI (Mild Cognitive Impairment) e AD (Alzheimer's Disease), utilizzando modelli di regressione multinomiale e confrontando diverse tecniche di bilanciamento del dataset.

# Variabili Utilizzate

Hippocampus\_Total, InfLatVentricle\_Total, Precuneus\_Total, Thalamus\_Total, Insula\_Total, SuperiorTemporal\_Total, Mammilare\_Total, CaudalMiddleFrontal\_Total

# Tecniche di Bilanciamento

1. Pesi di classe (Class Weights)  
2. SMOTE (Synthetic Minority Oversampling Technique)  
3. Undersampling casuale  
4. Undersampling con K-Medoids

# P-Values per Modelli Multinomiali Completi (8 variabili)

|  |  |  |  |
| --- | --- | --- | --- |
| Variabile | Class Weights (AD) | SMOTE (AD) | K-Medoids (AD) |
| Hippocampus\_Total | 0.6141 | 0.0000 | <0.001 |
| Thalamus\_Total | 0.8400 | 0.0005 | 0.0250 |
| SuperiorTemporal\_Total | 0.9616 | 0.5406 | 0.2100 |
| Insula\_Total | 0.8910 | 0.0113 | 0.0223 |
| Precuneus\_Total | 0.7660 | <1e-8 | <0.001 |
| Mammilare\_Total | 0.9132 | 0.0263 | 0.0913 |
| CaudalMiddleFrontal\_Total | 0.9888 | 0.6961 | 0.4467 |
| InfLatVentricle\_Total | 0.7041 | 1.02e-10 | 6.5e-6 |

# P-Values per Modelli Ridotti (3–5 variabili)

Modello Ridotto (SMOTE):

|  |  |  |
| --- | --- | --- |
| Variabile | SMOTE (AD) | SMOTE (MCI) |
| Hippocampus\_Total | 0.0000 | 0.1006 |
| InfLatVentricle\_Total | 7.10e-12 | 0.0051 |
| Precuneus\_Total | 1.06e-10 | 0.747 |
| Thalamus\_Total | 0.0005 | 0.515 |
| Insula\_Total | 0.0003 | 0.681 |

Modello Ridotto (K-Medoids):

|  |  |  |
| --- | --- | --- |
| Variabile | K-Medoids (AD) | K-Medoids (MCI) |
| Hippocampus\_Total | <0.001 | 0.127 |
| InfLatVentricle\_Total | <0.001 | 0.128 |
| Precuneus\_Total | 0.0002 | 0.912 |
| Thalamus\_Total | 0.0250 | 0.822 |

# Conclusioni

- Le tecniche di balancing SMOTE e K-medoids hanno migliorato drasticamente la significatività statistica rispetto ai class weights.  
- Le variabili più predittive di diagnosi AD risultano essere:  
 \* Hippocampus\_Total (negativa)  
 \* InfLatVentricle\_Total (positiva)  
 \* Precuneus\_Total (negativa)  
 \* Thalamus\_Total (negativa)  
- Il modello ridotto a 3–4 variabili offre una buona interpretabilità senza compromettere la significatività.

# Discussione Finale e Prospettive Future

I risultati ottenuti evidenziano come la scelta della tecnica di bilanciamento influenzi sensibilmente la significatività delle variabili nel modello. In particolare, SMOTE e K-Medoids hanno permesso di evidenziare relazioni più robuste tra le misure volumetriche e la diagnosi, riducendo il bias introdotto dallo sbilanciamento originale.  
  
Dal punto di vista clinico, l'importanza di strutture come l'ippocampo e il ventricolo laterale inferiore è coerente con quanto riportato in letteratura sulla neurodegenerazione. Queste variabili dovrebbero essere considerate come potenziali biomarcatori nel monitoraggio della progressione dell'Alzheimer.  
  
Per sviluppi futuri, sarebbe utile:  
- Integrare variabili genetiche (es. SNPs) nel modello per aumentare la capacità predittiva.  
- Validare i risultati su un dataset indipendente.  
- Utilizzare tecniche di interpretabilità dei modelli (es. SHAP) per analizzare in modo più granulare l'importanza delle variabili.

# Interpretazione dei Coefficienti Significativi

Di seguito si riportano i coefficienti delle variabili risultate significative nei modelli più informativi (SMOTE e K-Medoids). I coefficienti rappresentano il log-odds della probabilità di appartenenza a una categoria diagnostica rispetto alla categoria di riferimento (CN).

## Modello SMOTE (Diagnosi: AD)

|  |  |
| --- | --- |
| Variabile | Coefficiente (log-odds) |
| Hippocampus\_Total | -1.6488 |
| InfLatVentricle\_Total | 0.8823 |
| Precuneus\_Total | -0.9038 |
| Thalamus\_Total | 0.6013 |
| Insula\_Total | 0.4146 |

- Il coefficiente negativo per Hippocampus\_Total e Precuneus\_Total indica che una riduzione di volume in queste aree aumenta la probabilità di essere nella categoria AD rispetto a CN.  
- Il coefficiente positivo per InfLatVentricle\_Total suggerisce che un aumento del volume ventricolare è associato a maggior rischio di AD.  
- Anche le regioni Insula e Thalamus mostrano associazioni positive moderate.

## Modello K-Medoids (Diagnosi: AD)

|  |  |
| --- | --- |
| Variabile | Coefficiente (log-odds) |
| Hippocampus\_Total | -1.4999 |
| InfLatVentricle\_Total | 0.7560 |
| Precuneus\_Total | -0.7587 |
| Thalamus\_Total | 0.5016 |

- I coefficienti confermano le tendenze osservate nel modello SMOTE: ippocampo e precuneus con effetto negativo, mentre ventricolo, talamo (e insula) presentano effetto positivo.  
- Il pattern supporta l'ipotesi neurodegenerativa legata alla diagnosi di Alzheimer.