Descrizione della ricerca

In una comunità di pazienti con doppia diagnosi composta da circa 200 soggetti è stata condotta una ricerca per cercare di capire alcune caratteristiche legate alla permanenza nella struttura.
Nel file di dati sono presenti alcune variabili socio-demo come età, se si hanno figli o meno, se si hanno denuncie e quali tipi di diagnosi hanno ricevuto.

Descrizione dei dati

Età: età del paziente in anni
figli_minori: 0=no 1=si
precedenti_trattamenti: 0=no, 1=si
permanenza: numero di giorni di permanenza nella comunità
fine_programma: 1=concluso, 2=non concluso
abbandono= 0: no; 1:si.

Domande

1. Controlla se avere figli minori ed aver effettuato un qualsiasi precedente trattamento influenza il tempo di permanenza nella struttura.
    - VD = tempo di permanenza
    - VI = figli 2 livelli
    - VI = trattamenti precedenti, 2 livelli
    - Abbiamo calcolato una analisi della varianza univariata con le variabili dicotomiche tra i soggetti, con un disegno sperimentale 2 x 2. Il modello spiega il 2,6% della varianza del tempo di permanenza, tuttavia non risulta significativo ($F_{(3, 187)}$=1.682, p=.172, eta quadrato parziale = .026), non posso fare inferenze. Gli effetti principali di Figli minori ($F_{(1, 187)}$= 1.721, p=.191, eta quadrato parziale = .009) e precedenti trattamenti ($F_{(1,187)}$=2.765, p=.098, eta quadrato parziale = .015) non sono statisticamente significativi. L'interazione tra le variabili indipendenti avere figli e avere avuto un trattamento precedente è significativa ($F_{(1, 187)}$ = 4.351, p=.038, eta quadrato parziale = .023). Osservando il grafico dell'interazione di tipo non ordinale, vediamo che chi non ha subito precedenti trattamenti avere figli corrisponde ad un tempo di permanenze maggiore rispetto a chi non ha subito un trattamento precedente. Per chi non ha figli, l'aver avuto un precedente trattamento vede una permanenza maggiore rispetto a chi non ha avuto un trattamento precedente.

2. Calcola i giorni di permanenza per un soggetto che ha una deviazione standard in più della media di età.
    - PRIMO MODO: trovo ds-età e media-età usando descrittive, descrittive: Valore predittore = media(età) + dev.st(età) = 33.59 + 7.171 = 40,761 anni perché l'età era espressa in anni. Quindi la domanda la possiamo trasformare in "quanti sono i giorni di permanenza per un soggetto che ha 40,761 anni. Un soggetto di età di una deviazione standard maggiore della media, ovvero di 40,761 anni, resterà nella comunità per 396 giorni.
    - SECONDO MODO: Posso risolvere anche guardando ai punteggi standardizzati del mio modello che ho già pronti. Y= Beta * X. All'aumentare di una deviazione standard di X, Y aumenta di beta quindi di .064 punteggio standard. Y = Beta * x -> Y=.064 dev.st ma quanti giorni sono?

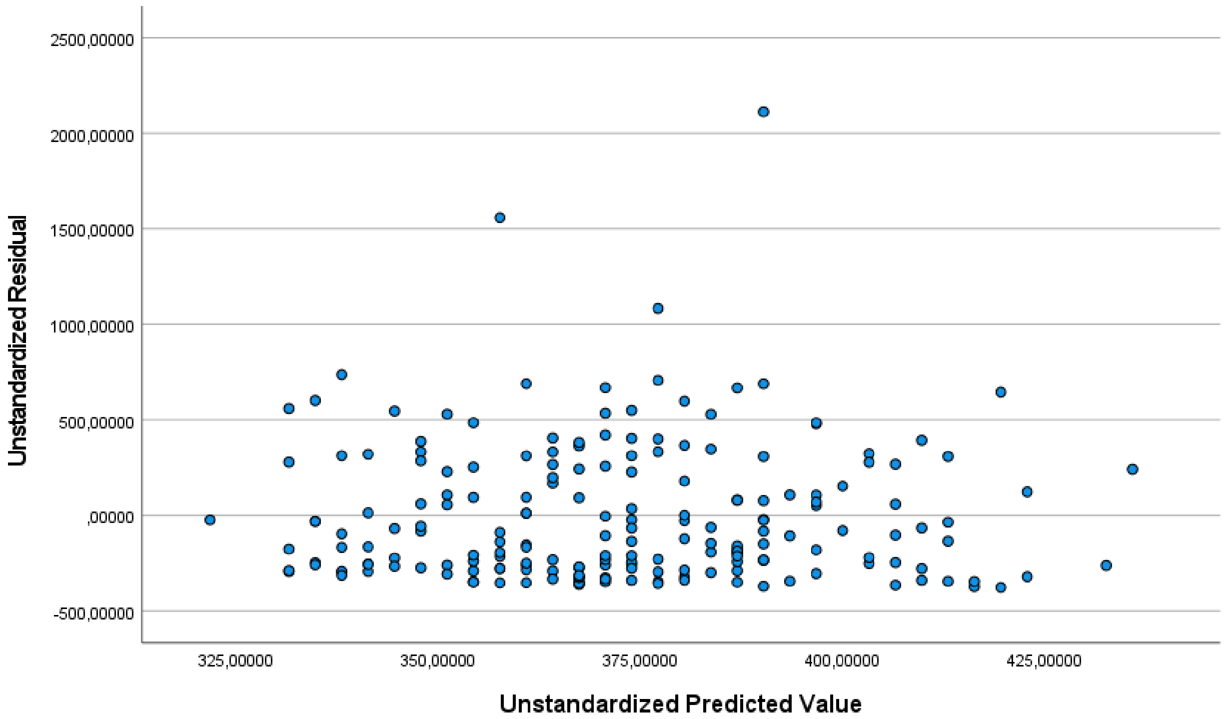
- Se voglio sapere un punteggio nell'unità di misura partendo da un punteggio standard cosa faccio? Ricordate come si calcola un punteggio standard e andate indietro.
    1. Punteggio standard = (punteggio – media) / dev standard.
    2. 0.064 = (x – 372.56) / 363.713 -> 23.277632 = x – 372.56 -> x=395.838.

- TERZO MODO. Oppure posso standardizzare l'età e mettere posto di x il valore 1, e poi calcolare a+b. x e poi rifare la regressione semplice con Zetà su permanenza. Ovviamente gli inferenziali del modello sono gli stessi. Andiamo a vedere a e b e li sostituiamo: Y=a+b*x -> a+b -> 372.564 + 23.375 = 395.939

3. Controlla se la permanenza nella struttura incide sulla probabilità di abbandonare o meno il percorso.
    - VI: permanenza nella struttura
    - VD: abbandono -> variabile dipendente dicotomica -> regressione logistica
    - Il modello spiega il 18,7% della varianza della variabile dipendente (R^2 di Nagelkerke = .187, Chi quadrato (1) = 29.338, p>.001). La probabilità di abbandonare il percorso quando i giorni di permanenza sono zero è di 3.101 volte quella di non abbandonarlo (W(1)=21.763, p <.001). All'aumentare di un giorno di permanenza, la probabilità di abbandonare il percorso diminuisce di .997 volte (W(1)=22.952, p<.001).

4. Inciderebbe qualora parzializzassimo l'effetto di età?
    - Il modello spiega il 20,5% della varianza della VD (Rquadrato di Nagelkerke = .205, Chi quadrato (2) = 32.500, p>.001). La probabilità di abbandonare il percorso quando i giorni di permeanza sono zero e abbiamo parzializzato l'età, è di 11.432 volte quella di non abbandonarlo (W(1)=9.511, p =.002). All'aumentare di un giorno di permanenza, e parzializzando l'età, la probabilità di abbandonare il percorso diminuisce di .997 volte (W(1)=22.459, p<.001).
    - L'effetto di età cambia la probabilità di uscire dal percorso appena arrivati (è più probabile abbandonare il percorso a zero giorni), ma la probabilità di lasciare il percorso all'aumentare dei giorni di permeanza rimane uguale.

5. Testa le assunzioni della domanda 2.
    - Osservando il grafico a dispersione dei valori predetti e dei residui, possiamo dire che linearità, omoschedasticità sono rispettate. Potrebbero esserci 2 o 3 outlier. Osservando il grafico istogramma dei residui, ipotizziamo che l'assunto di normalità non venga rispettato; inoltre vediamo 3 outlier. Il test di K-S risulta significativo, quindi l'assunto di normalità non viene rispettato.


2500,00000
2000,00000
1500,00000
1000,00000
500,00000
,00000
-500,00000
Unstandardized Residual
325,00000
350,00000
375,00000
400,00000
425,00000
Unstandardized Predicted Value