Cose fatte e da fare

**21 Marzo**

Stefano mette il database ricevuto da Mariachiara che era meno lavorato di quello precedete e aveva tutti gli zeri nel campo diagnosi

**23 Marzo**

Stefano e Mariagrazia rivedono tutta la pulizia dei dati fatti fino ad ora

* Ricodificare professione
  + Ho ricodificato la prima cifra di sdo1\_profes che indica occupati/inoccupati e ha creato le variabili profession e profession\_simple Adesso bisogna
    - Ricodificare la seconda cifra (per quelli occupati) che indica il tipo di occupazione: job\_type
    - Ricodifcare la terza cifra (che indica il settore) job\_sector
* Ricodificare dia\_pri:
  + Creare una variabile con i primi tre caratteri (diapri\_cod)
  + Estrarre l’elenco aggregato per verificare cosa hai
  + Quelli con la lettera codificarli per primi.
  + Per Quelli senza lettera Creare tabella excel da mettere in cartella “data” con i l’elenco preso da wikipedia in cui a seconda delle prime 3 cifre mette la macro-diagnosi.

27 Marzo

Mariagrazia

* Ricodificare sdo1\_terapia come Medical or Surgical
* Ho terminato la ricodifica della diagnosi primaria
* Ho creato la variabile numero\_di\_infezioni e almeno\_una\_infezione (TRUE = 10.48%)

28 Marzo

Mariagrazia

* Ho ricodificato anche le successive cinque diagnosi

31 marzo

Mariagrazia

* Pulizia file excel DRG
* Creata tabella con procedure invasive e messa nella cartella data
* Ho creato un nuovo campo proc\_inv che mi dice se almeno uno dei codici delle procedure invasive si trova nelle procedure primarie o secondarie
* Ho creato il campo cod\_proc che mi dice, solo per i TRUE, il codice del tipo di procedura hanno subito

3 aprile

Mariagrazia

* Ho creato il campo drg\_soglia associato al numero del drg
* Ho creato la variabile sup\_soglia, in cui TRUE se i giorni di degenza hanno superato la soglia prevista da DRG, altrimenti FALSE (NB: ci sono 817 missing nella soglia perché per alcuni codici di DRG non c’è una soglia assegnata. I codici sono 557, 558, 559 e 577. I missing sono stati controllati e corrispondono ai valori del drg di questi codici)

Cose da fare

* ~~Capire la questione procedure con Claudia o Luca. Quali campi della SDO indicano le procedure? Come categorizzarle? Probabilmente ICD 9 ma con quale criterio.~~
* ~~Creare una variabile del costo trasformata con log() che useremo come outcome principale della propensity~~
* RIcordificare le procedure in questo modo
  + ~~Fare elenco di quelle invasive (sono circa 18 e sono sui file nella cartella procedure\_mapping) e metterlo nella cartella “data”~~
  + ~~Creare un campo (True o False) che è True se uno dei quei codici si trova nelle colonne con le procedure primarie o secondarie.~~
  + ~~Eventualmente, solo per quelli TRUE, creare un campo con il tipo di procedura invasiva subita~~
* DRG
  + ~~Pulire file per codifica drg~~
  + ~~A seconda del DRG24 mettere per ciascun record la soglia~~
  + ~~Calcolare campo binario per quelli in cui la degenza supera la soglia (TRUE o FALSO) e campo numerico con la differenza di giorni tra degenza e soglia~~
  + Decidere come trattare i soggetti che hanno generato costo 0 (forse per il momento li togliamo, e poi alla fine vediamo se, e come re-inserirli)
* Dopo aver fatto tutto, rivedere le variabili escluse per capire se ripescarne qualcuna
* Capire bene come trattere quei missing in drg\_soglia. Si potrebbe fare una imputazione con i metodi per i missing, oppure calcolare noi una mediana e utilizzare quella come soglia. Per il momento li lasciamo un attimo come NA
* Verificare come mai alcuni pz risultano infetti di un batterio ma non è indicato il sito, mentre altri è indicato il sito ma non il batterio

# Considerazioni di metodo

* Istruzione e professione sono correlati con età, quindi forse si possono togliere
* I reparti con pochi infetti (< 5) forse si possono togliere perché sono reparti dove le ICA non sono un problema rilevante (vedi oculistica)
* L’età potrebbe essere fatta per classi, ma non so se dividere per classi di anni uguali o per numero di pazienti uguali

**DA FARE ANALISI – Riflessioni**

**I giorni di degenza li abbiamo usati come parametro del matching e del PS, ma forse andrebbero considerati come un outcome. Il costo può aumentare perché si registra il DRG dell’infezione, ma anche in caso di DRG non registrato l’infezione può aumentare i giorni extra soglia.**

**Ragionare su questa cosa**

**MG - 29/05/2023**

FATTO:

* Ho inserito nel db i 72 pz che avevano l’infezione sul sito ma non sul batterio.
* Ho aggiunto al batterio anche questi nuovi valori
* Ho calcolato num\_infezioni e batterio\_pos considerando questi nuovi dati
* Ho inserito in db i campi che contengono le date delle procedure primarie e secondario
* Ho convertito i campi formato numerico in formato data
* Ho creato la variabile “data\_proc\_inv” che mi dice in che data è stata effettuata solo la procedura invasiva legata ai codici
* Ho creato la variabile “proc\_inv\_real” che mi dice TRUE se “data\_proc\_inv” + 2days è < di date\_inv.