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

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
* Dopo aver fatto tutto, rivedere le variabili escluse per capire se ripescarne qualcuna