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. 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.

Cose da fare

* Dopo aver fatto il codice per dia\_pri applicarlo alle altre 5 possibili diagnosi, e infine cancellare il codice originale
* Capire la questione procedure con Claudia o Luca. Quali campi della SDO indicano le procedure? Come categorizzarle? Probabilmente ICD 9 ma con quale criterio.
* Ricodificare sdo1\_terapia come Medical or Surgical
* Creare una variabile del costo trasformata con log() che useremo come outcome principale della propensity
* Per le infezioni creare un campo con “almeno una infezione = (Si o no)” e un campo con numero di infezioni. Il primo campo dovrebbe dare un Si a circa il 13% del campione. Se non è così c’è un errore nella codifica.
* …
* Dopo aver fatto queste cose fissare riunione con Claudia e Luca per vedere:
  + Le variabili calcolate cosa significano e quanto sono utili a noi
  + Le procedure sono state già lavorate da loro? Va bene quello che hanno fatto oppure ha senso lavorarci ancora?
* Dopo aver fatto tutto, rivedere le variabili escluse per capire se ripescarne qualcuna