

Corso Big Data Public Health aa.2022/2023

Docenti: Prof.ssa Maria Grazia Valsecchi

Prof.ssa Paola Rebora

Tutor: Dott.ssa Anita Andreano

Indicazioni per la preparazione del progetto finale

Si hanno a disposizione quattro dataset, tre estratti di registri di popolazione e uno di un flusso sanitario amministrativo:

1. estratto del **German health registry** per l'anno **1984 relativo ad una piccola cittadina** (GermanH.csv) (dataset reale openData modificato). Le variabili che lo compongono sono le seguenti (e sono relative alla situazione del cittadino all'inizio del 1984):

- idnum*: codice identificativo del soggetto
- smoke* (yes or no): se il soggetto fuma
- sex* (Female or Male)
- married* (yes or no): se è coniugato
- kids* (yes or no): se il soggetto ha figli
- work* (yes or no): se lavora
- education* (no/low = da nessuna a diploma scuola media inferiore;
medium/high = diploma scuola media superiore o oltre)
- age* (numerica, in anni)

2. **Registro tumori tedesco** (Cancerregister.csv) (dataset simulato) relativo al mese di Gennaio 1984. Le variabili che lo compongono sono le seguenti:

- *idnum*: codice identificativo del paziente
- *stadio* (I, II, III, IV): stadio del tumore alla diagnosi
- *incidenza*: data di diagnosi del tumore
- *tipo di tumore*: seno, polmone, colon, altro
- *geneticm*: fattore genetico (1=positivo, 0=negativo)

3. **Schede di dimissione ospedaliera dei soggetti ricoverati in Germania tra gennaio 1984 e ottobre 1984 per trattamenti oncologici** (SDO.csv) (dataset simulato):

- idnum*: codice identificativo del soggetto
- *Prestazione*: tipo di trattamento ricevuto durante il ricovero (chirurgica, chemioterapia o radioterapica)
- *data prestazione*: data del trattamento
- *dimissione*: data di dimissione dall'ospedale
- *ospedale*: codice univoco dell'ospedale

4. estratto del **Registro di mortalità** della cittadina tedesca che riporta la mortalità dal 1984 al 1988 e lo stato in vita alla fine del 1988 (Deathregister.csv) (dataset simulato):

- idnum*: codice identificativo del cittadino
- dead*: stato in vita alla data *enddate*

enddate: data di ultima osservazione (se *dead*=1 data di morte)

- 1) **Esaminare i dati e riportare le statistiche descrittive in una tabella per ciascun dataset. Per le date riportare data minima e data massima. Fare attenzione alla possibilità di dati mancanti, incongruenze tra date, record ripetuti che potrebbero creare problemi in fase di linkage e analisi. I record con dati ripetuti o incongruenze tra date (eg. Data trattamento precedente alla data d'incidenza) devono essere segnalati nel report e poi eliminati per le analisi successive.**
- 2) **Effettuare il record-linkage con lo scopo di costruire l'indicatore 'Intervento chirurgico di asportazione del tumore al seno entro 60 giorni dalla data di diagnosi' su base mensile per i casi incidenti nel mese di gennaio 1984.**

Denominatore: tutte le pazienti di sesso femminile con tumore al seno insorto tra 01/01/1984 e 31/01/1984, in stadio I o II, che hanno subito un intervento chirurgico

Numeratore: tutte le pazienti al denominatore con intervallo tra la data d'incidenza e la data dell'intervento ≤ 60 giorni

- 3) **Calcolare l'indicatore 'Intervento chirurgico di asportazione del tumore al seno entro 60 giorni dalla data di diagnosi' per ospedale e darne rappresentazione grafica, includendo come valore di riferimento nel grafico l'indicatore calcolato sull'intero dataset. Per esempi relativi alla rappresentazione grafica fare riferimento al sito Piano Nazionale Esiti (PNE) o siti analoghi trattati a lezione.**
- 4) **Utilizzare il dataset ottenuto per valutare l'associazione a livello individuale tra il livello di educazione ed il valore dell'indicatore 'Intervento chirurgico di asportazione del tumore al seno entro 60 giorni dalla data di diagnosi'.**

Quale misura di effetto è possibile stimare?

Calcolate ed interpretate tale misura di effetto grezza. Riportare anche la relativa tabella di contingenza.

- 5) **Calcolate la stessa misura di effetto, questa volta aggiustata per la sola variabile 'working', mediante il metodo Mantel Haenszel. Interpretate il risultato.**
- 6) **Stimate l'associazione a livello individuale tra il livello di educazione ed il valore dell'indicatore, aggiustata per tutte le variabili disponibili che ritenete opportuno inserire come potenziali confondenti, mediante un modello di regressione logistica. Interpretate i risultati. Su quanti soggetti avete effettuato l'analisi?**

Quali variabili sono associate all'indicatore? In che modo?

Considerate ora tutti i tumori al colon insorti nel gennaio 1984. Unire i data-set utili per studiare la mortalità del tumore al colon nei soggetti inclusi nell'estrazione del German health Register (dataset 1).

- 7) Selezionate i record relativi ai tumori al colon e stimate la sopravvivenza a 5 anni.**

Quanti soggetti sono inclusi nell'analisi?

Quanti pazienti sono morti nel periodo di interesse?

Riportare graficamente la stima di sopravvivenza nei primi 5 anni dalla diagnosi stimata tramite lo stimatore di Kaplan-Meier.

Stimare approssimativamente la sopravvivenza mediana.

- 8) Stimare la di sopravvivenza nei primi 5 anni dalla diagnosi per Stadio e effettuare un test d'ipotesi per verificare se l'azzardo di morte sia diverso per stadio di malattia alla diagnosi.**
- 9) Applicare un modello per valutare l'associazione tra sesso e mortalità e interpretare la misura di effetto stimata.**
- 10) Quali variabili sono associate alla mortalità? Riportare le relative stime di effetto con gli intervalli di confidenza.**
- 11) Valutare la presenza di confondenti e/o modificatori di effetto tra le variabili disponibili nel German health register e nel registro tumori nella valutazione dell'associazione tra sesso e mortalità.**
Se identificate un'interazione tra sesso e un'altra variabile riportare le stime di effetto per maschi e femmine separatamente e commentare il tipo di interazione trovato.
- 12) A seguito delle considerazioni effettuate nei punti precedenti scegliete un modello finale per valutare i fattori di rischio della mortalità dopo diagnosi di tumore al colon e commentate i risultati.**