

Creazione Dataset virtuale

A mio parere (ma vi prego di contribuire in tal senso con critiche e suggerimenti), il dataset virtuale andrà creato in questo modo:

- Sappiamo che la prevalenza sulla popolazione generale è di 1:100. Possiamo quindi generare un certo numero di batch di pazienti, ognuno composto da 1 paziente positivo ed un numero random di pazienti negativi, estratto casualmente da una distribuzione gaussiana con media 100 e varianza 2 sigma.
 - o **Pazienti Negativi**
 - I dati relativi al **questionario** si limiteranno ai campi evidenziati (Anemia, Osteopenia, Diarrea, Mancata crescita, Disturbi Genetici, Madre celiaca). Si genereranno questionari con una (o più) positive con prevalenza 1:2.
 - Il **POCT** avrà esito negativo o inconclusivo, mantenendo la distribuzione degli inconclusivi (1:600), e considerando un numero di test difettosi pari a 1:100.
 - Nel caso di POCT negativo e Questionario negativo, **l'esame del sangue** avrà valori mancanti. Altrimenti si seguirà la stessa logica del POCT:
 - Prima si genererà un valore per le IGA totali.
 - o Nel caso di deficit (1:600), si genererà un valore per le TTG Igg con media 2 e varianza 2 sigma ed il valore delle TTG Iga rimarrà mancante
 - o Negli altri casi (599:600) si genererà un valore per la TTG Iga da una distribuzione random con media 4.5 e coda molto lunga a destra, ed il valore per le TTG Igg rimarrà mancante.
 - Nel caso l'analisi del sangue risulti positiva, La **biopsia** avrà ovviamente esito negativo (classe 1 o 2), altrimenti avrà valore mancante.
 - o **Pazienti Positivi**
 - Il **questionario** andrà tenuto sopra soglia con prevalenza 9:10. Per i valori noti la distribuzione delle risposte positive andrà rivista considerata la prevalenza nota, rispetto a quella utilizzata per il questionario dei casi negativi:
 - Disturbi Genetici (prevalenze 1:20, quindi 5 volte più probabile)
 - Madre celiaca di figlia (1:10, quindi 10 volte più probabile)
 - Madre celiaca di figlio (1:25, quindi 4 volte più probabile)
 - Il **POCT** avrà esito positivo se il questionario ha avuto esito negativo. Altrimenti sarà positivo con prevalenza 9:10, considerando sempre che sia inconclusivo con prevalenza 1:600.
 - Gli **esami del sangue** seguiranno una logica simile a quella per i casi negativi, ma saranno sempre presenti:
 - Prima si genererà un valore per le IGA totali. Nel caso di POCT inconclusivo avrà media .125 e varianza 1 sigma, altrimenti media 8 e varianza 2 sigma.
 - Nel caso di POCT inconclusivo si genererà un valore per le TTG Igg con media 14 e varianza 2 sigma ed il valore delle TTG Iga rimarrà mancante. Saranno scartati e rigenerati i valori nel range negativo.
 - Nel caso di POCT positivo si genererà un valore per la TTG Iga da una distribuzione gaussiana con media 24 e coda lunga a destra, ed il valore per le TTG Igg rimarrà mancante. Saranno scartati e rigenerati i valori nel range negativo.
 - La **biopsia** avrà esito positivo (classe 3a, 3b, 3c con distribuzione 1:3 uniforme)