

Keep Calm and EnGioI Statistics - N°1

Andrea Sansone & Angelo Cignarelli

Febbraio 2017

Introduzione all'introduzione

Gli scienziati hanno calcolato che le reali probabilità di esistenza di una cosa effettivamente assurda come questa siano una su un miliardo. I maghi, invece, hanno calcolato che le probabilità stimate una su un miliardo si avverano nove volte su dieci.

— Terry Pratchett, *Morty l'Apprendista*

SE STATE LEGGENDO QUESTE RIGHE, è verosimile che abbiate ricevuto un invito tramite mail, Facebook, o un avviso da qualche collega in merito a una newsletter di statistica spiegata in termini popolari, alla portata di tutti e soprattutto senza sborsare un quattrino. Certo, è possibile che siate finiti per caso in possesso di questo file, ma sinceramente riteniamo abbastanza improbabile che sia stato solo un caso¹.

Probabilmente vi starete domandando il perché di questa newsletter. Ci sono diverse risposte: la prima è che siamo senza dubbio due narcisisti assolutamente privi di vergogna² ma la più probabile è che, per quanto vi possa sembrare assurdo, *ci piace* fare statistica e divulgare, divulgare forte. Chiacchierando e scambiandoci appunti, ci siamo resi conto che sarebbe potuto tornare utile unire tutto questo materiale (invece di conservarli stile Gollum), con l'obiettivo di realizzare qualcosa più comprensibile, più interessante e, perché no, più divertente per tutti quelli che ritengono che la statistica sia una materia imperscrutabile.

Riassumere in pochi appuntamenti mensili il contenuto di dozzine di trattati di statistica, rendendoli al contempo di facile comprensione, ci sembra ovviamente un'impresa titanica. Noi faremo del nostro meglio per scrivere un testo che possa aiutare *concretamente* studenti di medicina, medici in formazione specialistica, biologi, biotecnologi, dottorandi o semplici curiosi ad acquisire capacità di base per raccogliere correttamente i dati, per analizzarli, per utilizzarli nella preparazione di una tesi o, perché no, nello scrivere un articolo scientifico.

Non siamo statistici, anche se forse, in un'altra vita³, lo diventeremo, e quindi è più che probabile⁴ che ci scappi qualche errore qua e là. I numeri non mentono, ma gli uomini possono sbagliare! (*PS: apparentemente, anche i software possono sbagliare. Quello che usiamo per scrivere questa newsletter, in particolare, ha deciso di ignorare ogni regola di grammatica per andare a capo. Abbiate pazienza, questi*

¹ Non entreremo per ora in merito a quanto possa essere probabile questo - diciamo decisamente poco, comunque.

² E ci mancherebbe altro. Siamo bellissimi.

³ Per chi non crede nella reincarnazione, togliere "in un'altra vita"

⁴ Di nuovo, non ci occuperemo di quantificare questa probabilità!



strafalcioni non sono colpa delle nostre maestre elementari!)

Preconcetti comuni nella statistica

Prima di iniziare realmente questa newsletter, vale la pena fare coming out e confessare quali sono le frasi comuni che ci fanno venire i tic nervosi.

”MA LA STATISTICA È DIFFICILE” è la prima, e più temibile. Già, statistica è difficile. *E invece anatomia patologica o biochimica FISCHI*. Però quando toccava preparare uno di questi esami abbiamo tutti scoperto che si può diventare dipendenti dalla Red Bull, mentre per dedicare qualche ora allo studio di statistica apriti cielo! La statistica in ambito bio-medico, fortunatamente, non è così terribile come sembra. I software più popolari (SPSS, STATA, R, etc...) consentono di fare in pochi istanti operazioni altrimenti molto dispendiose in termini di tempo ed energie; in più, i dati sono l'unica cosa che non mancano in ambito ospedaliero o in un laboratorio. Peso, altezza, cellule sane/cellule cancerose, punteggi dei più disparati indici o concentrazioni dei più sconosciuti ormoni... tutto a portata di mano, basta avere un PC, inserire tutto in un foglio Excel e poi decidersi ad analizzarli in maniera appropriata. Ci vogliono un po' di tempo e voglia per imparare almeno le basi per farlo: insomma, si può fare!⁵

”VABBÈ, MA IO RACCOLGO I DATI, POI LI PORTO ALLO STATISTICO E CI PENSA LUI” è un'altro buon modo per scatenare la furia omicida di un appassionato di statistica. Grosso modo è come dire “Ora vado a comprare un chilo di spaghetti, due uova, un po' di guanciale, una spolverata di pecorino, butto tutto a caso in un pentolone e poi tanto la carbonara si cuoce da sola”⁶. Se uno non sa a che servono i dati, o non sa come raccogliarli, o non sa come costruire un questionario, le cose andranno sicuramente male. E spesso anche uno statistico non è in grado di risolvere il problema, anche a fronte di un campione di popolazione immenso. Quasi tutti i libri di statistica raccontano la brillante storia del *Literary Digest* e del suo sondaggio sulle elezioni presidenziali americane del 1936 (*e il fatto che ancora si parli di una cosa risalente a 80 anni fa la dice lunga*)⁷. Il *Literary Digest* fece un sondaggio sulla popolazione americana in occasione delle elezioni che vedevano confrontarsi Alfred Landon e Franklin Delano Roosevelt: già nei precedenti turni elettorali il *Literary Digest* aveva indicato con una precisione quasi assoluta il vincitore delle elezioni con un largo anticipo. Nel 1936, oltre 10 milioni di individui - una cifra astronomica anche per gli standard attuali - furono intervistati, e circa 2.5 milioni di individui risposero. Secondo gli autori del *Di-*



5

⁶ Andrea ci tiene a sottolineare che ogni altro esempio culinario è stato brutalmente sevizato dall'introduzione dei robot da cucina.

⁷ Questa è una delle tre cose che si trovano in praticamente tutti i libri di statistica. Le altre le scoprirete a tempo debito.



gest, il vincitore sarebbe stato Landon... che invece fu sconfitto con una percentuale schiacciante. Perché? Semplicemente, l'errore fu nel campionamento: i soggetti intervistati erano gli abbonati del *Digest*, i possessori di automobili e i possessori di un telefono. Gli elettori di Roosevelt appartenevano alla popolazione esclusa dal sondaggio - e probabilmente anche la mancata risposta al questionario da parte di circa 3/4 degli intervistati ha avuto un suo peso. Morale della favola, onde evitare di fare la finaccia che ha fatto il *Digest* (fallito miseramente), conviene sapere cosa significa fare statistica se non altro per evitare di fare studi inconcludenti o, peggio, irrealistici.

”MA ALLA FINE BASTA CHE LA P SIA INFERIORE A 0.05, NO?” è la riprova che invece il vostro interlocutore qualcosa di statistica lo ha appreso... ma ha deciso di tagliare tutte le parti importanti intorno a questo semplice concetto. Sì, è sempre bello quando compare quel numerino, ma per far le cose per bene toccherà andare un po' più in profondità. Innanzitutto, dimentichiamoci che un test statistico (i.e. t di Student) possa essere sempre valido in ogni situazione: senza entrare troppo nel dettaglio (per ora), se i dati non mostrano un andamento *normale*⁸ usare dei test impropri porta a risultati scorretti, ed in quel caso, anche se la significatività statistica dovesse salvarvi, nulla potrà sottrarvi dal giudizio inclemente di chi *sa* di statistica e dalla conseguente figuraccia.

Inoltre, dire “p inferiore a zero punto zero cinque” è molto, molto poco specifico; la cosa più corretta sarebbe indicare, con un'adeguata approssimazione (tipo alla terza cifra decimale), il reale valore del vostro⁹ *p*. Pensatevi: vi sentireste più presi in giro da un commerciante che vi dice che il vostro oggetto concupito costa “meno di 50 euro”, o da uno che vi dice “esattamente 49.9 euro”? Esempi stupidi a parte, c'è stato recentemente un ***position statement*** (link) da parte della American Statistical Association finalizzato a rivalutare il ruolo del tanto adorato p statistico. Inoltre, il p-value non rappresenta l'entità dell'effetto: alcuni dati possono risultare molto significativi da un punto di vista statistico, ma irrilevanti da un punto di vista clinico o biologico¹⁰. Pertanto, la statistica non è solo p-value, ma anche di questo ne parleremo diffusamente in seguito.

Conclusioni

Siete giunti alla fine del primo “episodio” di *Keep Calm and EnGiol Statistics*. Come avrete notato, per ora non ci siamo addentrati più di tanto nella statistica vera e propria: abbiamo cercato di dare un approccio “soft” alla tematica, e proprio per questo abbiamo deciso di lasciare i primi concetti “veri” di statistica per un prossimo appun-

⁸ La curva normale è la curva gaussiana che sicuramente un vostro prof di statistica vi ha fatto vedere. Se non vi ricordate, è questa qui sotto. Ne parleremo approfonditamente in una prossima puntata

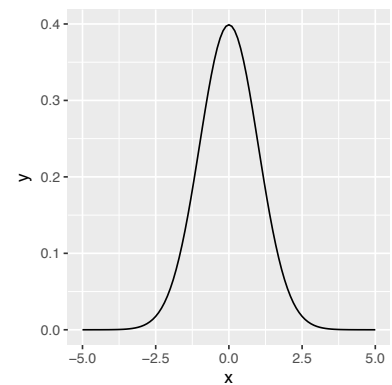


Figure 1: Curva Gaussiana

⁹ o della vostra, a seconda del vostro orientamento

¹⁰ Immaginate quanto possa essere rilevante l'aumento significativo di 1 grammo di peso corporeo o di 1% dell'espressione di un gene



tamento che, non temete, non tarderà ad arrivare. Abbiamo già steso una sorta di indice del lavoro e a breve, passo dopo passo, cercheremo di approfondire i vari concetti utili per la valutazione statistica, partendo dalle basi e arrivando (si spera) a parti più complesse. Probabilmente non riusciremo a rendere tutti perfetti statistici (del resto, non lo siamo neppure noi) ma speriamo di riuscire a farvi appassionare alla statistica o almeno di fornirvi le giuste chiavi per analizzare da soli, con spirito critico, i risultati dei vari *papers* con cui giornalmente tocca confrontarsi.

Vi invitiamo a fornirci i vostri commenti sulla pagina di Facebook di EnGioI ([link](#)) o via mail ai nostri indirizzi di posta elettronica (**Andrea** e **Angelo**).

Cercheremo di scrivere il prossimo “capitolo” nei prossimi giorni... Stay tuned!

NELLA PROSSIMA PUNTATA di *Keep Calm and EnGioI Statistics*, discuteremo di alcuni dei più annosi dilemmi del mondo moderno: come arroventare a puntino il ferro per punzonare le vostre vacche? Mentre per i vegetariani, come allacciarsi le scarpe con una mano sola? Le risposte a questi ed altri interessanti quesiti vi aspettano nel prossimo episodio di *Keep Calm and EnGioI Statistics*.

