

Introduzione

Le **statistiche di mortalità** per causa forniscono elementi informativi importanti per **confronti geografici e temporali** del rischio di morte per una causa, per gruppi di cause o per la totalità delle cause, costituendo uno strumento utile per la conoscenza dello stato di salute di una popolazione. Sono essenziali per la programmazione sanitaria e l'organizzazione delle attività di assistenza e di prevenzione. Rappresentano forse le più importanti statistiche sanitarie correnti, per la loro diffusione a livello mondiale e per l'uniformità dei sistemi di classificazione e codifica basati sulla classificazione internazionale delle malattie e dei problemi sanitari correlati (International Statistical Classification of Diseases and Related Health Problems), attualmente giunta alla decima edizione (ICD-10).

Con la descrizione della mortalità si rappresenta la distribuzione del **rischio di morte per causa** attraverso l'analisi e il confronto geografico, in funzione di variabili di genere, di stato civile, di istruzione o di altre caratteristiche. L'analisi e il confronto temporale dei dati ha lo scopo di valutare l'andamento nel tempo della mortalità per causa, ma non è oggetto di presentazione in questa relazione; per il momento ci si limita all'analisi geografica facendo una sorta di fotografia istantanea della mortalità nella ASL nel **periodo 2012-14** confrontando i dati con quelli dell'Italia del 2012. A tal fine sono stati messi insieme tutti i casi del triennio, modalità che consente di aumentare la precisione delle stime degli indicatori utilizzati per la descrizione della mortalità.

Le statistiche vengono prodotte, allo stesso modo con cui vi provvede l'Istat, attraverso l'indagine sui decessi e sulle cause di morte basate sulle **schede "Istat D4"** compilate dal medico curante o dal medico necroscopo. Ai medici vengono indicate, sul retro dei modelli, le regole per la buona certificazione, quali ad esempio rispettare la consequenzialità nella descrizione del processo morboso lasciandogli ampia libertà di compilare il certificato nella maniera che ritengono più opportuna e più vicina alla realtà del quadro patologico analizzato. Finora la formazione specifica in materia, carente fin dai corsi di studi universitari, è stata lasciata alla libera iniziativa dei singoli medici.

Si ricorda che il certificato internazionale di morte presentato nei testi dell'ICD-10 si differenzia molto dal certificato italiano in quanto non presenta parti diversificate per le morti accidentali e segue una logica di lettura contraria: sul modello internazionale si parte dalla causa ultima e si arriva alla causa prima, sul certificato italiano si parte dalla causa iniziale e si finisce sulla causa terminale.

Le statistiche di mortalità per causa si basano sul concetto di causa iniziale di morte così come definito dall'Oms *"la malattia o il traumatismo che avvia il concatenamento degli eventi morbosi che conduce direttamente alla morte"*, oppure *"l'insieme delle circostanze dell'accidente o della violenza che hanno provocato la lesione traumatica mortale"*. Il sistema di classificazione, l'**ICD-10**, è organizzato in **ventuno settori** (individuati da numeri romani), ognuno dei quali contiene un sistema di codici alfanumerici articolati per **categorie** (codice a tre cifre) e **sottocategorie** (codici a quattro cifre) a cui possono essere associate le diverse malattie: circa **6500 voci che classificano circa 2000 malattie**. L'ICD-10 è una **classificazione dinamica** aggiornata periodicamente, le revisioni, sia relative ai codici sia alle regole, presentano annualmente modifiche minori e ogni 3 anni modifiche maggiori. Si affianca a questa classificazione anche il *Manuale di codifica automatica Acme* che viene aggiornato annualmente, scaricabile dal sito del Cdc americano. La produzione delle statistiche di mortalità è pertanto basata, anche per la nostra ASL, sulla scelta di una causa, fra tutte quelle riportate nella scheda Istat-D, ovvero la **causa iniziale di morte**, individuata in base ai criteri di codifica e di selezione stabiliti dall'ICD-10. Negli ultimi anni si stanno producendo anche dati relativi a cause multiple di morte, modalità ancora in via di sperimentazione.

Per descrivere appropriatamente la mortalità e fare le relative valutazioni si devono conoscere gli elementi tecnici sottostanti alla specifica causa indagata; così ad esempio una mortalità elevata può essere legata alla grande diffusione di una malattia ma anche a una sua diffusione minore legata però ad un alto tasso di letalità

(si considerino ad esempio i valori di mortalità per Aids nel culmine dell'epidemia e prima dell'introduzione di terapie efficaci).

Il processo sistematico di codifica e di selezione della causa iniziale di morte nella ASL di Sanluri è iniziato solo nel 2014, la relazione sulla mortalità presenta i dati relativi ai decessi avvenuti nel triennio 2012-14 effettuando un confronto geografico con i dati del 2012 relativi all'Italia (ultimi dati resi disponibili dall'Istat), non vengono presentati report su confronti temporali che sarebbe auspicabile siano prodotti in futuro per poter valutare i trend.

L'intavolazione dei dati per la descrizione della mortalità è stata effettuata utilizzando la classificazione mutuata dalle Shortlist europee. Per un problema di intavolazione dei dati il report è carente delle statistiche sulle morti accidentali.

Gli indicatori statistici utilizzati, articolati con una ottica di genere, sono: **d** (numero dei decessi), **Quozienti mortalità** (X 10000), **Rapporti Proporzionali di Mortalità (RPM)**, ovvero la % della causa indagata rispetto al gruppo di appartenenza o alla totalità delle cause, **SMR** (Rapporto Standardizzato di Mortalità) e relativi **Limiti Fiduciali**, **Rango della causa** sul quoziente di mortalità.

La **popolazione standard** utilizzata per la standardizzazione degli SMR, con il metodo indiretto, è la **popolazione di Italia a metà periodo 2012**.

Solo per la mortalità generale sono stati calcolati i **Tassi standardizzati**, con il metodo diretto.

Completano la descrizione della mortalità **quadri sinottici** con tabelle riepilogative che riportano anche il risultato della **valutazione dei singoli SMR**, evidenziando quelli che hanno valore **alto o basso** e la relativa significatività, con un livello prefissato di probabilità della statistica α uguale al **5%**.

Il Responsabile Re.N.Ca.M

NOTE METODOLOGICHE PER LA LETTURA E LA VALUTAZIONE CRITICA DEI DATI

a) Generalità

La lista delle cause di morte utilizzata per presentare i dati di mortalità si compone di categorie tra loro mutuamente esclusive, selezionate a partire dai raggruppamenti previsti dalla European Shortlist of causes of death – Revisione 2012, in tal modo gli utilizzatori possono effettuare confronti con altri livelli rispetto a quelli considerati.

L'ICD-10 utilizzata per la codifica si compone dei seguenti settori:

SETTORI ICD-10	DESCRIZIONE
I	Alcune malattie infettive e parassitarie (A00-B99)
II	Tumori (C00-D48)
III	Malattie del sangue e degli organi ematopoietici ed alcuni disturbi del sistema immunitario (D50-D89)
IV	Malattie endocrine, nutrizionali e metaboliche (E00-E90)
V	Disturbi psichici e comportamentali (F00-F99)
VI	Malattie del sistema nervoso (G00-G99)
VII	Malattie dell'occhio e degli annessi oculari (H00-H59)
VIII	Malattie dell'orecchio e dell'apofisi mastoide (H60-H95)
IX	Malattie del sistema circolatorio (I00-I99)
X	Malattie del sistema respiratorio (J00-J99)
XI	Malattie dell'apparato digerente (K00-K93)
XII	Malattie della cute e del tessuto sottocutaneo (L00-L99)
XIII	Malattie del sistema oste muscolare e del tessuto connettivo (M00-M99)
XIV	Malattie dell'apparato genitourinario (N00-N99)
XV	Gravidanza, parto e puerperio (O00-O99)
XVI	Alcune condizioni morbose che hanno origine nel periodo perinatale (P00-P96)
XVII	Malformazioni e deformazioni congenite, anomalie cromosomiche (Q00-Q99)
XVIII	Sintomi, segni e risultati anormali di esami clinici e di laboratorio, non classificati altrove (R00-R99)
XIX	Traumatismi, avvelenamenti ed alcune altre conseguenze di cause esterne (S00-T98)
XX	Cause esterne di morbosità e mortalità (V01-Y98)
XXI	Fattori influenzanti lo stato di salute e il ricorso ai servizi sanitari (Z00-Z99)

Le schede di morte codificate sono quelle di **1337** donne e **1536** uomini relative ai deceduti nel triennio 2012-14 residenti anagraficamente in uno dei comuni appartenenti alla ASL.

Il **livello di completezza dei dati**, anche se valutato su una sola fonte (morti secondo Istat), si può stimare intorno al **99,8%**. La scelta di utilizzare la popolazione residente nel triennio 2012-14 (popolazione a rischio), come denominatore su cui calcolare i rischi di morte, comporta il vantaggio di migliorare le stime delle statistiche, ma ha lo svantaggio, ovviamente, di far perdere le informazioni sull'andamento temporale all'interno del periodo.

Un controllo esterno indipendente, per valutare la qualità della codifica sulle schede ASL 2012-14 ha dimostrato una **corretta codifica (secondo l'ICD-10) del 90% alla terza cifra** (quella utilizzata per la

presentazione dei dati statistici) e **dell'83% alla quarta cifra** dei codici ICD. Il controllo di qualità della codifica, come da relazione agli atti, è stato effettuato con la seguente modalità: sono state estratte 100 schede da quelle custodite con criterio temporale, in particolare dei quattordici faldoni numerati progressivamente dal n. 1 al n. 14 è stato selezionato, con una procedura at random (utilizzando l'apposita funzione di Excel), il n. 13 e da quest'ultimo sono state estratte le 100 schede sottoposte a verifica da un codificatore esterno.

La codifica delle informazioni sanitarie è un processo complesso. Devono prima essere codificate tutte le patologie presenti nella scheda e successivamente devono essere applicate le regole per individuare il codice della causa iniziale di morte. Le schede sono state codificate in un primo tempo da due operatori medici, mentre in un secondo momento solo uno di essi ha effettuato una nuova verifica, su tutte le schede, con eventuale correzione di codici ritenuti errati. Pertanto, di fatto è come se un unico osservatore ha codificato tutte le schede. Il medesimo osservatore successivamente ha effettuato la selezione della causa iniziale (una fra quelle appartenenti ai settori I-XX), per l'intavolazione dei dati, utilizzando la procedura automatica fornita dal software ACME/TRANSAX aggiornato al 2012. I record sono stati inseriti nel DB Access ed elaborati con il foglio elettronico Excel.

Per quanto attiene ad alcune verifiche, fatte per evidenziare elementi di criticità sulla qualità della certificazione, si osserva quanto segue. Su un controllo fatto su 1666 schede (sul totale di 2783 codificate) per valutare la concordanza della causa indicata dal medico certificatore al punto 1 della scheda Istat con la causa selezionata dal software con la procedura automatica, è risultata una **concordanza pari al 65%**. Inoltre, premesso che **i codici R (R00-R99)**, essendo relativi a: sintomi, segni e risultati anormali di esami clinici e di laboratorio, non classificati altrove, **NON dovrebbero, per definizione, essere ordinariamente utilizzati per l'individuazione della causa iniziale di morte**, si è andati a vedere quanto pesa fra i medici certificatori della ASL l'utilizzazione dei codici R confrontandoli con il livello nazionale. L'analisi dimostra, nel periodo analizzato, *un eccesso nella ASL di cause di morte con codici R (statisticamente significativo nelle donne)* rispetto ai valori di Italia 2012.

Completano la descrizione della mortalità tabelle riepilogative finali che accanto al **numero di morti (d)** e ai **quozienti di mortalità x 10000** riportano anche: i **ranghi**, per genere e riferimenti geografici; le stime degli **SMR**, accompagnati dai loro **limiti fiduciali**; gli **osservati** e gli **attesi**; le **risultanze della valutazione** delle stime degli **SMR**. Per le cause più importanti la descrizione statistica è più analitica riportando i **Rapporti Proporzionali di Mortalità, grafici con andamento della mortalità** in funzione dell'età, ecc.

La **popolazione standard** che viene normalmente utilizzata per i calcoli delle stime degli SMR per i confronti nazionali è la popolazione al censimento del 2001, in realtà possono essere scelte anche altre popolazioni. Quella utilizzata per le nostre statistiche è quella relativa **all'Italia 2012**. La struttura della popolazione per età dell'Italia del 2012 è sicuramente più vicina a quella della ASL, nel periodo considerato, rispetto a quella del censimento del 2001. Ma la scelta della popolazione standard è relativamente libera, ad esempio Cislaghi propone di utilizzare una "**popolazione tipo italiana**" (Epidemiol Prev 2009; 33 (4-5), Periodo: luglio-ottobre, pagine: 190-192).

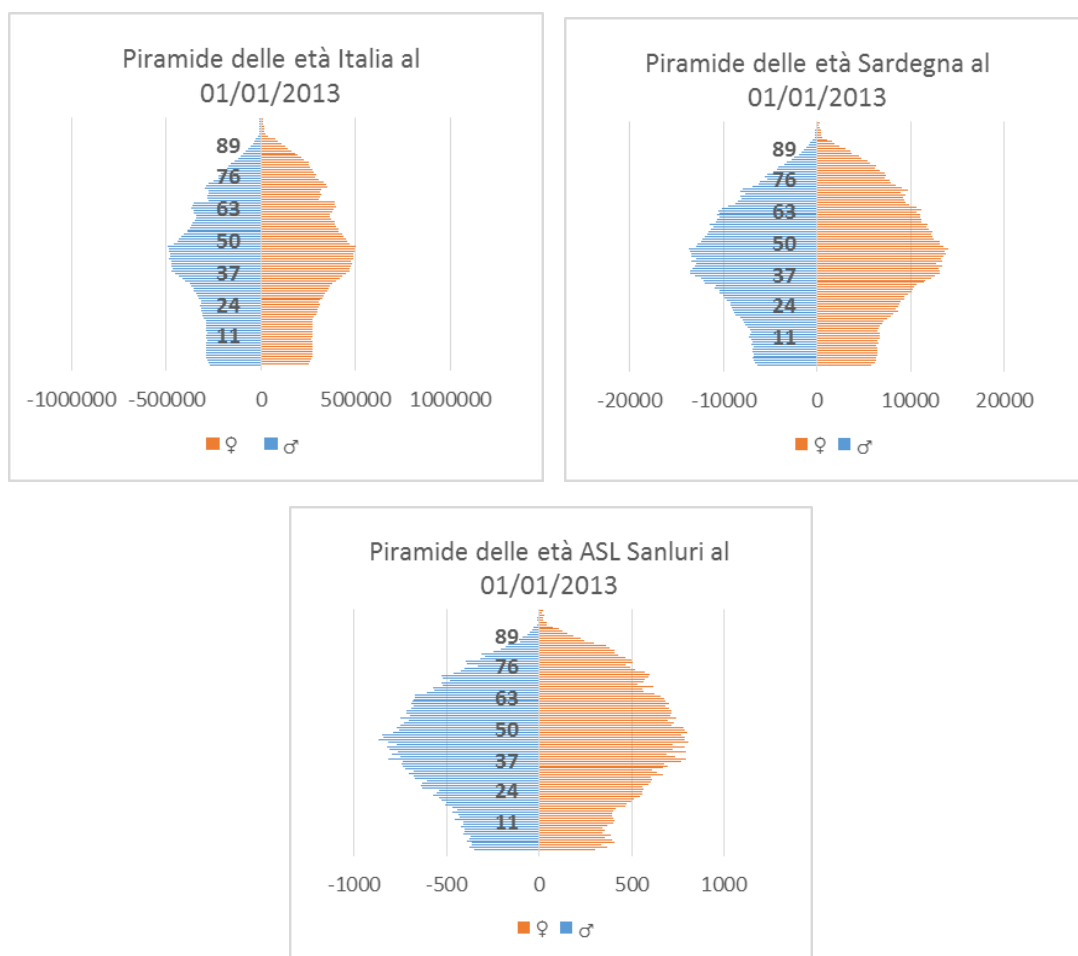
Per la descrizione della mortalità è necessario avere un minimo di conoscenza della struttura della popolazione per età in quanto essa è il denominatore, ovvero la popolazione a rischio su cui calcolare il rischio di morte per una o più cause.

Di seguito si riportano **le piramidi delle età di Italia, Sardegna e ASL** al 01/01/2013, che consentono di visualizzare le differenze di struttura per età di tali popolazioni.

Sono anche riportati i relativi **indici di vecchiaia**, per genere, di ASL, Sardegna e Italia, sempre al 01/01/2013, che contribuiscono a dare una idea della struttura delle relative popolazioni.

L'indice di vecchiaia è dato dal rapporto tra la popolazione over 65 anni e quella 0-14 rapportato a 100.

Gli indici **dimostrano un invecchiamento della popolazione della ASL, maggiore di quello della Sardegna che a sua volta è maggiore di quello dell'Italia.**



Indice di vecchiaia per genere al 01/01/2013

AMBITO GEOGRAFICO	GENERE		
	♂	♀	♂ + ♀
Italia	125,4	178,9	151,4
Sardegna	142,1	198,1	169,2
ASL Sanluri	162,3	221,3	190,8

b) Indicatori statistici

Di seguito si riportano gli indicatori statistici utilizzati per la descrizione della mortalità illustrandone sinteticamente il loro significato.

1. La statistica **d**, ovvero il **numero assoluto dei morti** per una causa, per un gruppo di cause o per la totalità delle cause (i valori potrebbero essere calcolati anche per fasce di età). È un numero che si ottiene con l'operazione elementare del conteggio dei morti (i **casi osservati**). Il suo significato è quello di dare una prima informazione sulla entità della causa in considerazione, nella realtà territoriale considerata che nel nostro caso è la ASL. I valori di **d non ci danno la possibilità di confrontare direttamente i valori riferiti a ambiti differenti** (ad esempio, ASL e Italia hanno valori di **d per tutte le cause per anno**, nei periodi considerati, uguali rispettivamente a circa **1000** e circa **600.000**; *nulla si può concludere dal confronto solo di questi due valori*). Per poter descrivere la frequenza e la distribuzione degli eventi morte è necessario ricorrere ad apposite misure, dette appunto "misure epidemiologiche" (*Rapporti,*

Proporzioni, Tassi), attraverso le quali è possibile osservare e quantificare i più importanti indicatori epidemiologici (*Incidenza, Prevalenza*) utili per la descrizione del fenomeno. Il numero assoluto da solo, infatti, non basta per misurare un fenomeno sanitario, per questo sono più utili le statistiche presentate di seguito.

2. La statistica **quoziente di mortalità** (detto impropriamente tasso grezzo, in quanto ha dei limiti che poi presenteremo), ovvero il valore del **rapporto tra il numero dei morti**, per una causa, per un gruppo di cause o per la totalità delle cause, e la **popolazione a rischio rapportata in genere a 10.000**; consente di confrontare i valori calcolati per differenti ambiti (ad esempio, i valori per *tutte le cause*, nei periodi considerati, sono per **ASL e Italia** uguali *rispettivamente* a **94,9** e **103,0**, i valori consentono di concludere per un quoziente di mortalità maggiore dell'Italia).
3. La statistica **Rapporto Proporzionale di Mortalità (RPM)**, ovvero il **rapporto percentuale fra i casi di morte**, per una causa o per un gruppo di cause, e il **totale dei casi del gruppo o di tutte le cause**. Fornisce informazioni sul **“peso”** di cause singole o gruppi di cause rispetto al riferimento considerato (totale di gruppo o totale generale); i valori dei RPM calcolati per ambiti di riferimento differenti possono essere confrontati direttamente fra loro, anche se offrono informazioni parziali sul fenomeno della mortalità per la causa considerata. Ad esempio il RPM per **tutti i tumori maligni** rispetto a tutte le cause nella **ASL** è uguale, totale maschi e femmine nel periodo considerato, a **28,75**; indica che *il 28,75% di tutte le cause di morte nella ASL è dovuto a tumori maligni*.
4. Le statistiche **Tasso standardizzato** (i valori possono essere calcolati per fasce di età) e il **Rapporto Standardizzato di Mortalità** (acronimo **RSP**, nei paesi anglosassoni **SMR**: Standardized Mortality Ratio); *per condizioni non fatali, il Rapporto Standardizzato di Mortalità a volte è riportato come Rapporto Standardizzato di Morbosità*. Entrambi, seppur differenti, hanno in comune l'applicazione di metodi di calcolo che consentono un **aggiustamento** dei loro valori **in modo da eliminare l'influenza delle differenti strutture per età delle popolazioni su cui sono calcolati**. Infatti, come è noto, la mortalità cresce al crescere dell'età e la standardizzazione, diretta per calcolare i tassi standardizzati o indiretta per calcolare gli SMR, rappresenta un metodo adeguato per avere la disponibilità di valori che, essendo indipendenti dalla struttura per età della popolazione, si prestano a confronti diretti. Il **Tasso standardizzato ASL (X 10000)** è stato calcolato solo per la **mortalità generale** (*valori nel triennio considerato, ♀: 70,2; ♂: 104,8; ♂ + ♀: 86,9*), per tutte le altre cause o per gruppi di cause i dati sono presentati utilizzando il SMR.

La **popolazione “standard”** utilizzata è quella relativa all'**Italia 2012**. *Per i confronti pertanto si possono presentare solo i valori delle stime degli SMR della ASL, in quanto il valore di riferimento di Italia 2012 è posto uguale a 100 per le varie cause*. Lo **SMR** concettualmente deve essere considerato un **indicatore comparativo di rischio**, nel nostro caso del **rischio morte** per la causa considerata, per un gruppo di cause o per la totalità delle cause. Per calcolare il valore occorre avere il numero dei *decessi attesi rispetto ad una popolazione standard*, vale a dire il numero di decessi che ci si aspetterebbe di avere nella popolazione considerata se essa avesse la stessa composizione, struttura per età, della popolazione standard (nel nostro caso Italia 2012). Pertanto, per calcolare i decessi attesi è necessario conoscere la distribuzione per gruppi di età della popolazione considerata e i tassi di mortalità età specifici della popolazione Standard. *Il risultato è quello di avere il totale delle morti attese ponderate per età*.

Il **valore dello SMR** per la causa considerata è dato dal **rapporto tra il numero degli osservati e il numero degli attesi rapportato a 100** (per evitare di avere a che fare con numeri piccoli). Lo SMR in quanto tale è una grandezza fittizia, costruita per indicare la percentuale dello scostamento del numero di morti osservate rispetto a quelle attese. Ad esempio, il valore dello SMR Asl per il tumore dello stomaco pari a 60 significa che la mortalità è inferiore del 40% (rispetto alla popolazione standard: Italia 2012), il valore dello SMR Asl per le epatiti virali pari a 181,3 significa che la mortalità è maggiore del 81,3% (sempre rispetto alla popolazione standard).

5. **I limiti fiduciali (LCI e LCS) dello SMR:** non è sufficiente calcolare i valori degli SMR (o dei Tassi Standardizzati) per effettuare confronti, nel nostro caso fra ASL e Italia, e trarre conclusioni sulle eventuali differenze dei valori. Per un primo confronto fra i valori è necessario accompagnarli con i valori dei loro intervalli di Confidenza (Limiti di Confidenza Inferiore e Superiore), in modo da considerare le influenze della variabilità casuale. **Il valore reale di un SMR per una causa non è conosciuto**, quando lo si calcola non si fa altro che ottenere un valore che, infatti, si dice essere una “stima” del valore reale; stima più o meno vicina al valore reale in funzione della “precisione” della stima stessa. La stima dello SMR sarà tanto più accurata quanto più grande è il numero dei casi osservati. I Limiti di Confidenza sono due: uno inferiore (LCI) e l’altro superiore (LCS), rappresentano l’intervallo di valori entro cui è compreso il **valore reale dello SMR**. Tanto più piccolo sarà l’intervallo tra i due limiti (che come per lo SMR dipende dalla grandezza del numero dei casi osservati) tanto più precisa è la stima del suo valore reale compreso in tali limiti. Infine si deve considerare che i valori dei limiti di confidenza dipendono dal livello prefissato di probabilità che si intende ottenere (di solito posto pari al 95%) per farvi ricadere il valore reale dello SMR.

Per il calcolo dei limiti di confidenza delle stime degli SMR ASL sono stati utilizzati due metodi: a) quando il numero dei **morti osservati è inferiore a 100 è stato utilizzato il test esatto di Fisher** (anche se per un numero superiore a 5 di morti osservate potrebbero essere ugualmente usati altri metodi di approssimazione che sono accurati quasi quanto i test esatti);

b) quando il numero dei **morti osservati è maggiore di 100 il metodo e la formula di calcolo derivano dalla presunzione di una distribuzione “normale”** del rischio di morte per la causa o il gruppo di cause considerate. Ma per capire questo concetto essenziale è necessario soffermarsi un attimo, fare un passo indietro nel tempo e andare idealmente in una località tedesca per conoscere meglio colui che ci ha lasciato in eredità qualcosa di straordinario e immenso che ci aiuta a capire meglio il problema.

La “normalità” del principe della matematica e la variabilità delle stime delle statistiche

Dobbiamo andare a Gottinga, cittadina tedesca sede di una prestigiosa e antica scuola di matematica, patria di illustri matematici e punto di riferimento della plurisecolare competizione tra le due scuole europee: tedesca e francese.



In tale città ha operato Johann Carl Friedrich Gauss dove studierà nella stessa Università dal 1795 al 1798, vi soggiognerà e lì morirà nel 1855. Il padre prefigurava per lui un destino da muratore, la madre ne intuì le doti, che manifestò fin dalle scuole elementari; presto venne preso sotto l’ala protettrice del Re di Hannover di cui divenne amico. Gauss era di origini estremamente modeste. Il nonno era contadino; il padre lavorava a giornata come giardiniere e muratore. Gauss frequentò la scuola locale più povera. Un famoso aneddoto, riportato dalla scuola, è probabilmente vero molto più di altre storie simili. ...*Un giorno il maestro, per prendersi mezz’ora di pausa, assegnò alla classe il compito di sommare i primi cento numeri*¹. Poco dopo, Gauss posò la sua lavagnetta sul tavolo del maestro, esclamando, *Ligget se!*, che nel dialetto contadino di

¹ La **somma di Gauss** ha un certo fascino come ricorda l’iraniana **Maryam Mirzakhani**, prima donna vincitrice della medaglia Fields (il Nobel della matematica), nell’intervista al quotidiano Repubblica del 14/08/2014 nel servizio Le Signore dei Numeri: “*Volevo fare la scrittrice scoprii la somma di Gauss e incontrai la bellezza*”.

quelle parti significa “Ecco”. Gauss aveva elencato a mente i numeri nell’ordine (1, 2, 3, .., 100), poi nell’ordine inverso (100, 99, 98, ...1) e quindi aveva sommato un numero della prima lista con quello corrispondente dell’altra ($1 + 100 = 101$, $2 + 99 = 101$, $3 + 98 = 101$, ...). Aveva ottenuto cento volte il risultato 101, e poiché ogni numero era stato contato due volte, per ottenere la risposta giusta era necessario dividere a metà quella somma: cinquanta volte 101, che fa 5050. Facile quando ve lo spiegano, ma non è un metodo che verrebbe in mente a qualunque ragazzino di 10 anni. La fortuna di Gauss fu che gli insegnanti riconobbero le sue capacità e furono disposti a sopportare qualche sacrificio per incoraggiarle. Fu una fortuna ancora maggiore vivere nel piccolo ducato tedesco di Brunswick governato dal duca Carlo Guglielmo Ferdinando... Valoroso soldato per tutta la vita, era feldmaresciallo dell’esercito prussiano ed era stato a capo della forza congiunta prussiano-austriaca che i francesi arrestarono a Valmy il 20 settembre del 1792².

Noto, ancora in vita, come **il principe della matematica**, “la regina delle scienze” come lui stesso la definì, sulle orme di Eulero. Persino dalla firma si può riconoscere un personaggio particolare. Gauss può a buon diritto essere considerato il più grande matematico di tutti i tempi. In vita era noto come princeps mathematicorum –il principe dei matematici- e alla sua morte il Re di Hannover³, Giorgio V, commissionò una medaglia commemorativa in suo onore, con quel titolo inciso.

Gauss pubblicò pochissimi scritti. Il suo sigillo personale recava impresso un albero con pochi frutti e il motto, ***Pauca sed matura, “pochi, ma maturi”*** (IN L’OSSESSIONE DEI NUMERI PRIMI, Bernhard Riemann e il principale problema irrisolto della matematica, Jhon Derbyshire. *Le Scienze* 2009, Edizione speciale per il mensile). Si diceva che fosse come la volpe che cancella le tracce con la coda per non essere seguita.



Studiò, tra l’altro, il **comportamento degli errori**. Inventò il metodo dei minimi quadrati, che tende a ridurre al minimo gli errori di misurazione, usandolo per calcolare l’orbita del pianeta Cerere solo con poche osservazioni empiriche sul suo moto.

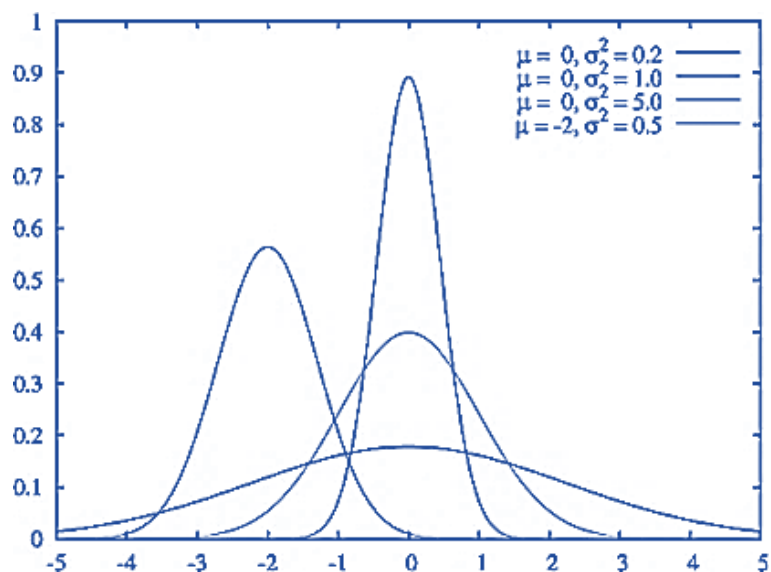
Tuttavia il lavoro più importante in questo ambito fu la scoperta della **variabile casuale normale**, detta anche gaussiana. Essenziale e superba come vuole la sua denominazione: **curva di normalità**. La curva descrive il comportamento e l’entità degli *errori di misurazione*⁴. L’eredità lasciata, considerato che stiamo parlando dell’evento morte, si potrebbe dire che è un’eredità che ha consegnato Gauss all’immortalità.

Ed ecco la sua curva, di straordinaria bellezza.

² La **battaglia di Valmy** è stata una delle quindici battaglie decisive della storia del mondo, con essa iniziarono le guerre napoleoniche che impegnarono l’Europa per i successivi 23 anni, dopo con il trattato di Vienna del 1815 l’Europa visse un periodo di pace di cento anni. “*Le quindici battaglie decisive nella storia del mondo, Autore: Eduard Creasy, editore: Editori & Associati -1991.*”

³ **Hannover** era uno dei quattro regni, con circa un milione di abitanti, della Germania del XVIII secolo che era formata da circa trecento entità, la maggior parte minori, tra le quali le più importanti erano la Prussia, con circa 15 milioni di abitanti, e la Baviera con oltre 2 milioni. I Re di Hannover per ragioni dinastiche erano anche i Re di Inghilterra e si chiamavano tutti Giorgio, Giorgio IV salì al trono nel 1826.

⁴ Nella misurazione di una qualsiasi grandezza si ottengono valori soggetti a **errori di misurazione** che comportano una incertezza della relativa misura. Il risultato della misurazione pertanto non è un unico valore bensì l’insieme dei valori probabili che assume l’oggetto della misurazione. Gli errori possono essere dovuti a errori sistematici dello **strumento** o a errori sistematici dell’**osservatore**. L’**errore casuale o statistico**, influenza la precisione del risultato e non è eliminabile, è un errore di misurazione che può incidere con la stessa probabilità in aumento o in diminuzione sul valore misurato. Ne deriva che una serie ripetuta di misurazioni comporta la progressiva riduzione dell’errore casuale, poiché i singoli scostamenti si annullano reciprocamente.



Distribuzioni gaussiane degli errori con **medie μ** uguali a: 0, 0, 0, -2 e **varianze σ^2** uguali a: 0.2, 1.0, 5.0, 0.5
(Wikipedia)

La variabile normale è una delle più importanti variabili casuali, molto utile nel modello probabilistico, in medicina facilita la spiegazione della causalità delle malattie croniche degenerative ad eziologia multifattoriale, e quindi la stragrande maggioranza delle malattie. Una qualsiasi distribuzione statistica di probabilità può essere descritta dai parametri che misurano la **tendenza centrale**: la **media**⁵, la **moda**⁶ e la **mediana**⁷: nel caso della “curva normale” o “curva di Gauss” i tre parametri coincidono; questo è il dato da rilevare, rispetto alle distribuzioni non Gaussiane. Gli statistici ricorrono poi a parametri più complessi, che misurano la **variabilità** della distribuzione con **indici di dispersione**: **scarto medio** (media aritmetica degli scarti assoluti: valori assoluti di differenza tra il dato e la media), **varianza**⁸ e **scarto quadratico medio (o Deviazione Standard**⁹, DS o σ), che rappresenta l’indice più significativo e più conosciuto per la misura della variabilità. La *curva di Gauss o gaussiana o normale* per la sua forma è nota anche come Campana di Gauss¹⁰. La distribuzione normale è considerata il riferimento delle distribuzioni di probabilità continue che si incontrano spesso nelle applicazioni pratiche. La forma della distribuzione normale (più o meno schiacciata, più o meno dilatata) dipende da due parametri, la media μ e la varianza σ^2 (estraendo la **radice quadrata** della varianza si ottiene la **Deviazione standard (DS)**).¹¹ Conoscere il tipo di distribuzione di una variabile (ad esempio altezza, pressione arteriosa, glicemia, ecc.), aiuta molto quando si vogliono descriverne sinteticamente i dati (molti) raccolti sulla variabile stessa.

**Area sottostante alla curva = 1 = 100%;
la probabilità p varia da 0 a 1.**

p

⁵ La **media aritmetica semplice**, matematicamente è data dal rapporto tra la somma dei valori delle singole osservazioni con il numero totale delle osservazioni, esprime il valore che è più probabile trovare nella distribuzione normale; in tale distribuzione è uguale alla moda e alla mediana.

⁶ La **moda** è il valore più grande dell’insieme delle osservazioni, nella distribuzione normale ha lo stesso significato ed è uguale alla media.

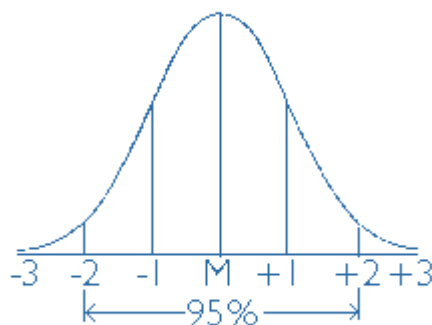
⁷ La **mediana** è il valore centrale dell’insieme dei dati ordinati (il 50% dei dati saranno inferiori e il 50% superiori), nella distribuzione normale è uguale alla media e alla moda.

⁸ La **varianza** è il quadrato della Deviazione Standard.

⁹ La **Deviazione Standard** è la radice quadrata della varianza.

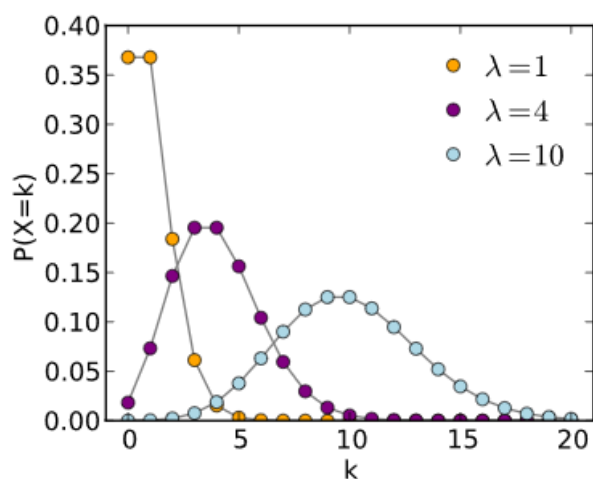
¹⁰ Detta anche curva degli errori o ogiva, è una *distribuzione teorica aleatoria continua di densità di probabilità* (asse Y).

¹¹ La particolarità della curva consiste, oltre all’aver gli indici di posizione (media, moda, e mediana) uguali, nel fatto che il 68% dei valori delle osservazioni (della variabile indipendente, asse x: ad esempio i valori di peso, di altezza, di glicemia, ecc.), è ricompreso in un segmento di ascissa tra la media e ± 1 DS, mentre il 95% tra la media e ± 2 DS.



Perché si possano applicare le proprietà che derivano dalla curva gaussiana è necessario che la distribuzione della variabile considerata abbia un andamento gaussiano, ovvero come quella della curva che la rappresenta. Per quanto attiene alla mortalità è **ammesso che la distribuzione della variabile numero di morti per causa abbia un andamento gaussiano quando il numero di casi osservati è > 100**. In questa eventualità è possibile calcolare i Limiti Fiduciari, a livello di un valore prefissato di probabilità che di solito è posto al 5%, basandosi sulle proprietà della curva di Gauss.

Invece, quando gli **eventi** sono **rari**, e li consideriamo tali quando il numero dei morti per causa osservati è < **100**, la distribuzione della probabilità ha un andamento che può essere descritto meglio da un altro tipo di distribuzione: la **distribuzione di probabilità di Poisson** (o poissoniana), **distribuzione di probabilità discreta, più adatta a rappresentare gli eventi rari**.



Distribuzione di probabilità di Poisson

Per le ragioni sopraesposte per calcolare le stime dei valori degli SMR della ASL sono state usate due modalità di calcolo, differenti a seconda che si ipotizzasse un andamento della distribuzione di probabilità di rischio di morte per la causa di tipo gaussiano o del tipo di Poisson.

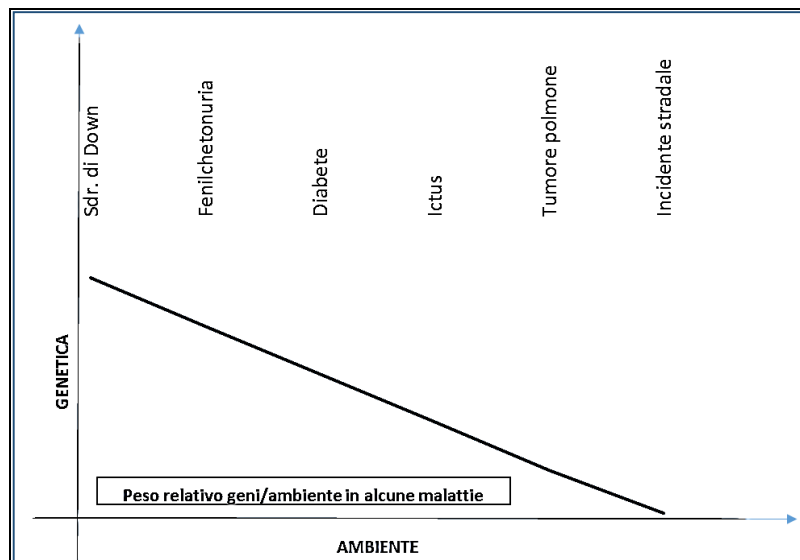
Per gli eventi rari, nel nostro ambito nell'eventualità che i morti osservati nella ASL per la causa considerata siano < 100, esistono diverse formule che consentono il calcolo dei limiti fiduciali per distribuzioni tipo Poisson; la formula utilizzata per i dati ASL è il **test esatto di Fisher**, con un livello prefissato di α al 5%.

Una volta calcolate le stime degli SMR per le varie cause di morte, la **valutazione** dei singoli **SMR** inizia con lo stabilire se il valore è **alto o basso**, e in che percentuale si scosta dal valore ottenuto sulla popolazione standard di riferimento, si prosegue con la valutazione del punto in cui ricade il **valore dello standard (Italia: 100)** rispetto ai limiti fiduciali dello SMR ASL; **se 100 è fuori dall'intervallo** (maggiore del LCS o minore del LCI) si può affermare l'esistenza di una **differenza significativa** (con un errore a livello del 5%) in quanto la differenza non può essere spiegata (con un livello posto al 95% di probabilità) dalla variabilità casuale. In pratica quando si afferma che si è trovata una differenza significativa, i metodi usati

consentono di controllare gli effetti della variabilità casuale, offrendoci l'opportunità di affermare con il 95% di probabilità che l'affermazione sia vera, o al contrario con un 5% di errore che l'affermazione sia falsa, e pertanto le differenze non sono dovute al caso.

È molto importante considerare che qualora non si possa concludere per l'esistenza di una differenza significativa, non si può escludere che vi siano **altre differenze** (*non dovute alla variabilità casuale*). In pratica la conclusione per una condizione di differenza significativa è relativa solo alla valutazione della variabilità casuale e non ad altre fattispecie di meccanismi causali che possono generare differenze; in altri termini **non poter concludere per una differenza significativa non equivale alla certezza di assenza di differenze per cause diverse dalla variabilità casuale**.

Questo è bene sia sempre considerato quando si vogliono trarre ipotesi di lavoro, che possono essere suggerite anche dagli studi di mortalità, per approfondimenti, con altri tipi di studio più adatti per indagare la causalità delle malattie nelle popolazioni a rischio cui riferire l'evento morte o l'evento malattia. Come afferma l'epidemiologo Kenneth J. Rothman¹² si deve ricordare che: *“Sebbene possa sembrare esagerato affermare che il 100% di ogni malattia è causato contemporaneamente sia da fattori ambientali che genetici, questa è comunque una buona approssimazione della realtà. Può essere non intuitivo tale concetto, perché la maggior parte delle volte non ci è possibile controllare molte delle cause, e quelle per cui è fattibile il controllo tendenzialmente o sono le ambientali o le genetiche, ma generalmente non entrambe. ...”*.



Nel grafico sono rappresentate alcune condizioni morbose associabili al peso dell'ambiente in senso lato o alla genetica



¹² **Kenneth J. Rothman**: Professore di Medicina Preventiva e Epidemiologia alla Boston University School of Medicine, ha ricevuto l'Abraham Lilienfeld Award della American Public Health Association per il 2002, riconoscendo l'eccellenza nell'insegnamento della epidemiologia nel corso di una carriera, è membro della Società Internazionale per la Farmacoepidemiologia e dell'American College of Epidemiology. I suoi lavori nel campo della ricerca epidemiologica spaziano dalla teratogenicità della vitamina A, agli effetti sulla salute dell'uso del telefono cellulare e potenziali fattori di rischio ambientale per il cancro, alle malattie cardiovascolari e difetti di nascita, epidemiologia ambientale, ecc.

Kenneth J. Rothman, Boston University

E di questo si deve tener conto quando pensiamo al rischio e alla cosiddetta popolazione a rischio: **la residenza in sé non è la causa delle malattie**, le cause devono essere ricercate nei fattori di rischio sottostanti alla residenza, ovvero nei rischi che ha sperimentato la popolazione (il denominatore cui sono riferiti i casi indagati) per l'esposizione complessiva a un insieme di fattori genetici e ambientali, intesi sia come fattori propri dell'ambiente fisico (aria, acqua, alimenti), sia come fattori "psicosociali" (socio-economici, lavorativi, extra lavorativi, comportamentali, ecc.), che incontrandosi e interagendo con l'organismo umano generano l'insorgenza della stragrande maggioranza delle malattie, ovvero di quelle che sono connotate da meccanismi di azione a eziologia multifattoriale.