

2 - INDICATORI STATISTICI UTILIZZATI

Le precedenti fasi della raccolta e della codifica dei dati di mortalità sono le tappe indispensabili per tradurre l'informazione pervenuta in sintesi numeriche rappresentative del fenomeno. E' necessario ricordare a tal proposito che col termine di "dato" si intende l'immagine elementare dell'evento, con "indice" si intende la misura quantitativa a cui quell'evento si riferisce e con "indicatore" si vuole intendere l'informazione strutturata di quel fenomeno per un suo utilizzo operativo. Sulla scorta degli indicatori ottenuti, quindi, è possibile effettuare gli studi di mortalità di tipo descrittivo inferenziale, quali quelli illustrati nel presente lavoro. A tal fine si presenta assai utile riportare sinteticamente il significato dei principali indicatori statistici che sono stati utilizzati in questo studio sulla mortalità.

2.1. Le misure epidemiologiche

Un primo semplice indicatore è rappresentato dal numero di eventi e, per quanto riguarda la mortalità, dal Numero di Decessi Annuì (NDA):

NDA (NUMERO DECESSI ANNUI)

Questa funzione statistica esprime il numero assoluto di decessi annui totale e per singole fasce quinquennali d'età e per sesso.

Ma per poter descrivere la frequenza e la distribuzione degli eventi 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, Mortalità). Il numero degli eventi da solo, infatti, non basta per poter misurare un fenomeno sanitario. In Epidemiologia è necessario rapportare sempre il numero degli eventi (numeratore) alla popolazione di riferimento, che rappresenta il denominatore di questo rapporto. Ai fini dello studio sulla mortalità, in questo lavoro saranno utilizzati i Tassi Proporzionali e Grezzi:

TP (TASSI PROPORZIONALI)

Il tasso proporzionale esprime la percentuale delle singole cause di morte sulla totalità delle stesse.

TGS (TASSI GREZZI SPECIFICI)

Il tasso grezzo specifico esprime i decessi annui per 100.000 abitanti suddivisi per sesso e per fasce d'età quinquennali. Risulta dal rapporto tra il numero di decessi per ogni classe quinquennale di età e la popolazione residente della corrispondente classe. E' espressa dalla seguente formula:

$$TGS_i = n_i / P_i \times 100.000$$

dove "n" e "P" rappresentano il numero di decessi annui e la popolazione, mentre "i" sta per i-esima classe d'età.

Nelle indagini epidemiologiche, però, diventa fondamentale conoscere la composizione per sesso e fasce d'età della popolazione studiata, poiché la loro diversa composizione influisce sull'andamento dei fenomeni sanitari. Quando, infatti, un fenomeno sanitario è fortemente correlato con l'età o con il sesso (come ad esempio l'occorrenza dei tumori, più frequente tra gli anziani), la semplice osservazione dei Tassi Grezzi Specifici potrebbe condurre a conclusioni errate. In una popolazione più invecchiata, ad esempio, la mortalità per tumori potrebbe apparire maggiore rispetto ad un'altra popolazione più giovane, e non tanto perché ci siano realmente più tumori nella prima popolazione rispetto alla seconda, quanto piuttosto perché nella prima popolazione ci sono più anziani (e dunque più tumori correlati ad essi) rispetto all'altra popolazione. Per annullare il suddetto fattore di distorsione, in Epidemiologia si ricorre alla tecnica della "Standardizzazione", applicando i tassi specifici per età della popolazione in osservazione alle classi di età di

una popolazione standard di riferimento (di solito la popolazione italiana dell'ultimo censimento), al fine di rendere confrontabili i tassi osservati anche in città caratterizzate da forti differenze demografiche (ad esempio le regioni del nord Italia, fortemente invecchiate per via del calo delle nascite, e le regioni del sud molto più giovani). Nel presente studio, dunque, si è fatto largo ricorso ai Tassi Standardizzati di Mortalità:

TSI (TASSI STANDARDIZZATI sulla popolazione ITALIANA)

I TSI (Tassi Standardizzati sulla popolazione Italiana) esprimono il numero di decessi annui per 100.000 abitanti standardizzati per fasce d'età. La standardizzazione diretta ricava i TSI totali dalla somma dei decessi attesi per fasce quinquennali d'età. In altri termini si applicano i tassi specifici per età della popolazione in osservazione alle classi d'età della popolazione standard di riferimento e, così, si ottiene il numero di decessi attesi nella popolazione di riferimento se questa morisse con la stessa probabilità (o tasso) della popolazione in studio. Sommando i totali per sesso dei decessi attesi e dividendoli per la numerosità totale (e per sesso) della popolazione standard, si ottengono i tassi standardizzati diretti per maschi e femmine nella popolazione in osservazione. I TSI, pertanto, sono espressi dalla seguente formula:

$$\text{TSI} = [s (\text{TGSi} \times \text{Pi} / 100.000)] / (s \text{Pi} \times 100.000)$$

dove "P" rappresenta la popolazione italiana standard ed "i" la "i-esima" classe d'età. Per elaborare i TSI sono stati utilizzati i dati del censimento 2001 della popolazione italiana.

Conoscere un Tasso, ancorchè standardizzato, non basta tuttavia ad avere una esatta conoscenza dell'entità del fenomeno osservato, se non si confronta il dato osservato negli studi locali con un elemento di paragone che può essere rappresentato dal corrispondente dato della media nazionale, regionale, provinciale, e così via. Negli studi di mortalità l'indicatore comparativo più utilizzato è il Rapporto Standardizzato di Mortalità.

RSM (RAPPORTO STANDARDIZZATO DI MORTALITA')

Il rapporto standardizzato di mortalità è un indicatore comparativo di rischio.

Per calcolare il RSM occorre ricavare innanzitutto i decessi attesi rispetto ad una popolazione standard, vale a dire il numero di decessi che ci si aspetterebbe nella popolazione in studio se quest'ultima sperimentasse la mortalità specifica della popolazione standard.

Per calcolare i decessi attesi è necessario, dunque, conoscere la distribuzione per gruppi di età della popolazione studiata, e i tassi di mortalità per i corrispondenti gruppi di età di una popolazione di confronto o popolazione "Standard" (in questa applicazione ci si è riferiti alla popolazione nazionale). Moltiplicando questi due valori per ogni fascia d'età e poi sommando i risultati si ottiene il totale delle morti attese, ponderate per età.

Il rapporto tra i decessi realmente osservati e quelli attesi, calcolati come sopra, dà il RSM, che è espresso dalla seguente formula:

$$\text{RSM} = \text{Morti osservate} / \text{Morti attese} \times 100.$$

In ultima analisi il RSM indica la percentuale dello scostamento del numero di morti osservate da quelle attese.

Ritornando alle Misure Epidemiologiche, occorre aggiungere che in Epidemiologia è necessario anche organizzare i dati raccolti e poi valutarli, ricorrendo ai metodi della Statistica: gli Indici di Tendenza Centrale (le Medie), gli Indici di Variabilità (Devianza, Varianza, Deviazione Standard, Errore Standard, Limiti Fiduciali) e le Prove di Significatività Statistica (t di Student, Analisi della Varianza, Errore Standard di Proporzioni o Test z, Chi Quadrato). Per l'elaborazione del presente studio è stato necessario ricorrere molto spesso agli Indici di Variabilità, ed in particolare ai Limiti Fiduciali, nonché ad alcune prove di Significatività Statistica (Test z)

2.2. Gli Indici di Variabilità

Le medie da sole non ci dicono se i valori esaminati sono molto discordanti fra di loro o sono, piuttosto, vicini alla media stessa, presentando scarti piccoli. In Epidemiologia è utile descrivere il grado di scostamento dalla media nell'ambito di una serie di osservazioni, ricorrendo agli Indici di Variabilità. La **Devianza** è la

sommatoria del quadrato degli scarti fra ciascun valore (x) che concorre alla media e la media stessa (m). La **Varianza** si ottiene dividendo la Devianza per il numero n di osservazioni (in realtà n-1, che esprime meglio il concetto di gradi di libertà del campione esaminato), ottenendo una vera media del quadrato degli scarti. La **Deviazione Standard** si ottiene estraendo la radice quadrata della Varianza, in modo da ottenere non la semplice media del quadrato degli scarti, bensì la media degli scarti. La Deviazione Standard, pertanto, esprime la media degli scostamenti dalla media (m) e permette di osservare la dispersione dei valori intorno alla stessa media (m). Non sempre è possibile, infatti, studiare medie ed altre misure direttamente dalla popolazione di riferimento. Il più delle volte si studiano campioni rappresentativi della popolazione di riferimento. La Deviazione Standard in realtà è riferita solamente al campione. L'indice di variabilità che mette in relazione i valori osservati con la popolazione dal quale è stato estratto il campione è l'**Errore Standard**. Tanto più piccolo è l'Errore Standard, quanto maggiore è il grado di certezza che la Media Campionaria si avvicini a quella reale della popolazione di riferimento dalla quale è stato estratto il campione. L'Errore Standard è espresso dal rapporto tra la Deviazione Standard e la radice quadrata della Numerosità del campione (n), secondo la formula: E.S. = Dev.Stand. / Rad.Q. n. Ovviamente più piccola è la Deviazione Standard (cioè la variabilità interna del campione), o più grande è la Numerosità del Campione, e più piccolo sarà l'Errore Standard. I concetti su esposti servono in realtà ad introdurre quello dei **Limiti Fiduciali**.

LIMITI FIDUCIALI dei TSI al 95%

I Limiti fiduciali, o intervalli di confidenza al 95% del TSI esprimono l'intervallo di valori entro il quale sarà compreso il valore reale del TSI con una probabilità del 95%.

E' dato dalla seguente formula:

$$LF_{95\%} = +/- 1,96 \times s / \text{Rad.Q. } n$$

dove "s / radice di n" rappresenta l'errore standard. Più in particolare sigma è la deviazione standard ed n la numerosità del campione. In altri termini più numeroso è il campione, più piccolo è l'errore standard e più ristretti saranno gli intervalli di confidenza e, dunque, più significativa sarà la stima del TSI.

Ma dinanzi ad una proporzione (come nel caso dei TSI) la base dell'analisi è data dalla distribuzione binomiale. Tuttavia quando la numerosità è sufficientemente grande si può usare la distribuzione normale gaussiana come alternativa alla binomiale, applicando la formula della binomiale approssimata per mezzo della distribuzione gaussiana:

$$LF_{95\%} = p +/- 1,96 \times \text{Rad.Q. } p(1-p)/n$$

dove p rappresenta la proporzione a 100.000 osservata (TSI) ed n la numerosità della popolazione studiata moltiplicata per gli anni di osservazione. I valori totali dei Limiti Fiduciali vengono visualizzati nel formato TSI +/- LF.

Nelle sedi neoplastiche dove la numerosità del campione risulta inferiore a 9 casi i LF non vengono calcolati.

In Epidemiologia, infatti, non basta calcolare i valori dei Tassi Standardizzati di Incidenza o di Mortalità o i Rapporti Standardizzati di Mortalità, già illustrati in precedenza. Spesso infatti occorre accompagnare i dati con l'indicazione degli Intervalli di Confidenza (o Limiti Fiduciali), che esprimono l'intervallo di valori entro cui ricadrà il reale valore dell'Incidenza o della Mortalità della malattia studiata, nell'ambito della popolazione da cui è stato tratto il campione osservato. I Limiti Fiduciali derivano dal prodotto tra l'Errore Standard ed un preciso coefficiente legato al livello di probabilità che si intende ottenere (per un livello del 95% di certezza, ad esempio, il coefficiente è 1,96, per un livello del 99% il coefficiente sale a 2,57).

LIMITI FIDUCIALI dei RSM al 99%

I limiti fiduciali al 99% esprimono l'intervallo entro cui ricadrà il valore reale del RSM e vengono calcolati secondo la seguente formula:

$$L.F.99\% = RSM +/- 2,58 \times 100 / \text{Rad.Q. } A \times Y$$

dove "A" rappresenta i decessi attesi ed "Y" gli anni di osservazione. Dunque, più numeroso è il campione, o maggiori sono gli anni di osservazione, e più ristretti saranno gli intervalli di confidenza. Quando il numero di decessi attesi non supera i 9 casi i LF non vengono calcolati.

I Limiti Fiduciali possono accompagnare, oltre i valori di un Tasso Standardizzato, anche quelli di un Rapporto Standardizzato di Mortalità. In questo studio si è ricorso ai L.F. 99% dei RSM osservati.

2.3. Le Prove di Significatività Statistica

L'osservazione epidemiologica porta spesso a riscontrare delle differenze nelle medie, nelle frequenze e in altre caratteristiche di gruppi di popolazioni messi a confronto. Di fronte a queste evenienze è doveroso chiedersi se le differenze riscontrate sono dovute al puro caso o se scaturiscono da reali differenze delle popolazioni messe a confronto.

L'ipotesi che le differenze osservate siano dovute al puro caso è definita in Epidemiologia "**Ipotesi di Nullità**". Dunque, per poter accertare la significatività statistica di una osservazione occorre "rifiutare l'ipotesi di nullità". Per far ciò si confrontano i valori ottenuti attraverso le Prove di Significatività con appositi valori "critici" predeterminati: se i valori ottenuti superano i valori critici la differenza osservata è statisticamente significativa e l'ipotesi di nullità è rifiutata.

Le prove sono collegate a ben precisi Livelli di Significatività (di solito il 5% o l'1%), mentre la probabilità che le differenze osservate siano dovute al caso è espressa dal valore di "p": una prova condotta con un $p < 0.01$ sta a significare che la probabilità che la differenza osservata sia dovuta al caso è uguale a 1 su 100 (valore altamente significativo), una prova condotta con un $p < 0.05$ sta a significare che la probabilità che la differenza osservata sia dovuta al caso è uguale a 5 su 100 (valore statisticamente significativo). Di solito un valore che si attesta al di sotto di $p < 0.05$ è considerato statisticamente non significativo (ns).

Le Prove di Significatività Statistica più largamente utilizzate sono:

- ❑ *Il t di Student e l'Analisi della Varianza nel caso di confronto tra Medie*
- ❑ *Il Test z nel caso di confronto tra Proporzioni o Rapporti*
- ❑ *Il Chi quadro nel caso di confronto tra Frequenze.*

Nel presente studio, al fine di vagliare la significatività statistica dei Rapporti Standardizzati di Mortalità osservati per le singole cause di morte si è fatto ricorso al Test z.

TEST Z DEL RAPPORTO STANDARDIZZATO DI MORTALITÀ

Il Test z viene calcolato dividendo la differenza tra due proporzioni osservate (numeratore) e dividendola per l'errore standard della differenza tra le due proporzioni (denominatore). Il valore ottenuto viene confrontato con un apposito valore critico, superato il quale la differenza osservata è da intendersi statisticamente significativa. E, dunque, maggiore è la differenza osservata tra le due proporzioni (ed al contempo minore è l'errore standard della differenza tra le due proporzioni), e maggiore sarà il valore di z e la significatività delle differenze. La formula del test z, pertanto è la seguente:

$$z = (p_1 - p_2) / \text{Rad}Q [p_1(1-p_1)n_1] + [p_2(1-p_2)/n_2]$$

dove p_1 e p_2 sono le due proporzioni messe a confronto, n_1 ed n_2 sono le numerosità dei campioni delle due proporzioni ed infine $1-p$ rappresenta il complemento a uno della proporzione.

Applicando la formula su esposta al nostro caso, il Test di significatività attraverso il quale si valuta la solidità statistica di un RSM è dato dalla formula:

$$Z = (\text{RSM} - 100) \times \text{Rad.Q. } A \times Y / 100$$

Dove "A" indica i casi attesi e "Y" gli anni di osservazione. Più in particolare viene riportato il livello di significatività del Test Z attraverso il valore "p", che esprime la probabilità che le differenze osservate siano dovute al puro caso. Quando il risultato ottenuto dal test di significatività supera il valore critico di 2,58 la differenza tra i decessi osservati e quelli attesi nel RSM è altamente significativa con un $p < 0,01$. Quando il risultato ottenuto supera il valore di 1,96, ma non quello precedente la differenza tra osservati ed attesi è statisticamente significativa con un $p < 0.05$. Al contrario quando il valore di Z non supera il livello critico di 1,96 la differenza osservata non è significativa (n.s.).

I dati relativi ai Rapporti Standardizzati di Mortalità con i corrispondenti valori dei limiti fiduciali e dei test di significatività statistica sono stati rapportati ai tassi di mortalità regionali e nazionali, e suddivisi in due gruppi di cinque caselle per ogni causa di morte.

Nella prima casella delle tabelle riepilogative, sono elencati i decessi osservati nel corso degli anni selezionati. Nella seconda casella sono riportati i decessi attesi, ricavati dai tassi età specifici della popolazione italiana e regionale (Annuario ISTAT 1994), e moltiplicato per il numero di anni di osservazione. Nella terza casella sono riportati i RSM. Nella quarta casella sono riportati i limiti fiduciali al 99%, che esprimono l'intervallo entro cui ricadrà il valore reale del RSM. Nella quinta casella, infine, viene riportato il risultato del test di significatività relativo al RSM attraverso le tre opzioni di $p < 0.01$ (valore altamente significativo), $p < 0.05$ (valore statisticamente significativo) e di n.s. (valore statisticamente non significativo).

2.4. Altri Indicatori epidemiologici utilizzati

Oltre agli indicatori precedentemente illustrati, si presenta particolarmente utile negli studi di mortalità un indicatore di costo sociale denominato PYLL, sigla inglese con la quale si indicano i Potential Years of Life Lost, vale a dire gli anni potenziali di vita persa, che esprimono la valutazione del "peso" delle varie cause di decesso in termini di mortalità prematura.

PYLL (ANNI POTENZIALI DI VITA PERSI)

Il PYLL (Potential Years of Life Lost) esprime il numero di anni che rimarrebbero da vivere fino a 65 anni per ciascun decesso riferito alla popolazione residente in quella particolare classe d'età, compresa tra 0 e 65 anni. E' espresso dalla seguente formula:

$$\text{PYLL} = \sum d_i (65 - h_i)$$

dove "h i" rappresenta l'età di ogni singolo decesso, e "d i" l'insieme dei decessi relativi alla causa in studio nella i-esima classe d'età.

Nella presente funzione sono riportati i seguenti valori:

- il PYLL-t o PYLL-tasso per mille
[PYLL-t = (PYLL/anni osserv.) x 1000 / Pop.Distr]
- il PYLL percentuale sulla totalità delle cause di morte
(PYLL% = PYLL- singola causa di morte x 100 / PYLL TOT. causa di morte).

Attraverso il PYLL è possibile valutare con accuratezza il maggior peso sociale rivestito da alcune cause di morte (ad esempio gli incidenti stradali) o di alcuni tumori (ad esempio le leucemie infantili) rispetto ad altre cause di morte, in relazione alla capacità di indurre morti premature nella comunità.

Quelli sopra illustrati sono gli Indicatori Statistici ed Epidemiologici che l'equipe del RTP ha ricavato dalla elaborazione dei dati al fine di poter studiare la mortalità osservata in provincia di Siracusa negli anni dal 1995 al 2002.

Prima di procedere alla esposizione dei risultati è necessario, tuttavia, soffermarsi su alcune informazioni riguardanti il territorio oggetto dello studio: vale a dire la Provincia di Siracusa.