

STATEMENT DELLA SIS SU COVID-19

Introduzione

La popolazione mondiale vive un momento eccezionalmente difficile che non riteniamo necessario commentare ulteriormente. La Società Italiana di Statistica (SIS) è una società scientifica costituita nel 1939 come ente morale e attualmente inclusa fra gli enti di particolare rilevanza scientifica, il cui scopo è promuovere l'attività ricerca per lo sviluppo delle scienze statistiche. In questo periodo di difficoltà, molti cittadini si rivolgono quotidianamente alla lettura di dati per comprendere ciò che sta accadendo, oltre che per avere un conforto e un sostegno delle volontà nello sforzo di uscire rapidamente dalla pandemia che ci ha colpiti. La lettura dei dati, e più ingenerale delle informazioni statistiche, a disposizione non è sempre immediata e univoca e necessita che, insieme ad essi, vengano forniti strumenti e chiavi di lettura per una loro corretta interpretazione e una piena comprensione dei fenomeni che descrivono. I rappresentanti della SIS sono chiamati in causa e, come cittadini responsabili, sentono urgente il compito di mettere a disposizione di tutti le proprie competenze.

Dati

Manca un coordinamento della progettazione dei dati statistici sul fenomeno, e forse in una situazione di emergenza non si poteva fare diversamente, ma per questo le informazioni non riescono a rappresentare adeguatamente lo sviluppo della pandemia. Ad esempio, i dati riportati nel sito web del Ministero della Salute e predisposti dall'Istituto Superiore della Sanità (ISS) sono aggregati e sono soltanto una illustrazione dei dati raccolti dai sistemi di sorveglianza per l'attuale emergenza. In questa forma sono insufficienti a descrivere e meglio comprendere il fenomeno per poter così fronteggiare adeguatamente l'emergenza e per il raggiungimento del comune obiettivo del bene pubblico. Il dato, in statistica, è di qualità se è aderente alla realtà concreta che si vuole misurare. La qualità inoltre non vale di per sé, ma è definita da caratteristiche di validità e di conformità del dato al tipo di analisi che viene effettuata.

Misure

Nel dibattito pubblico ogni giorno vengono discusse alcune cifre che non sono sufficienti per una valutazione statistica seria. Qualche esempio. Il numero giornaliero di nuovi casi non rappresenta accuratamente il numero di individui infettati poiché dipende in modo cruciale dalle modalità e dal numero dei tamponi effettuati e analizzati. Oltretutto spesso non viene specificata la variazione dei casi nel tempo in modo da avere anche la misura di quanti sono in entrata o in uscita dalla condizione di positività. È infine una misura inaccurata e non consente i confronti territoriali, poiché dipende dalle politiche sanitarie e dalla consistenza demografica delle regioni e degli Stati. Anche il numero giornaliero dei decessi va considerato con estrema cautela. Quotidianamente vengono conteggiati sicuramente, con un certo ritardo, solo i decessi di pazienti positivi al Covid-19 avvenuti nelle strutture sanitarie. Molto si è detto sulla relazione tra decessi "di" e "con" coronavirus e sulla corrispondenza con i dati ottenuti dall'Istat mediante la rilevazione dei certificati di morte (anche questi in naturale ritardo sui tempi dell'emergenza). I conteggi giornalieri di casi e decessi risultano altresì estremamente volatili e necessitano di interpolazioni per poter evidenziare tendenze evolutive. La stampa e le agenzie

governative mostrano spesso grafici dell'andamento cumulato del numero totale dei casi o dei decessi (su scala naturale o logaritmica). Va tenuto presente che tali grafici non aiutano a comprendere il fenomeno, e sono soltanto la premessa indispensabile sia per cercare di individuare le cause che lo provocano sia per scegliere eventuali modelli di previsione dell'evoluzione futura della pandemia. Sono oltretutto particolarmente utili se tali rappresentazioni sono dettagliate per Regione e/o per i domini territoriali di interesse per lo studio dell'epidemia.

Accesso ai dati

Pur consapevoli delle norme sulla privacy a tutela dei dati personali, è estremamente urgente poter disporre di almeno parte dei dati raccolti a livello individuale. Tali dati potrebbero essere forniti in forma anonima solo ai fini di ricerca, insieme ad altre informazioni quali età, data dei sintomi, data del tampone, data del ricovero, comune di residenza e di lavoro, solo per citarne alcune. Ciò perché solo con dati individuali e longitudinali è possibile comprendere appieno l'evoluzione dell'epidemia. A breve dovrebbe essere pronto il progetto della App per il tracciamento, elaborata dai tecnici del ministero dell'Innovazione per intercettare, curare e isolare i nuovi casi. Anche i dati provenienti dall'utilizzo di questa piattaforma potranno fornire spunti estremamente utili per la ricerca sulle caratteristiche della diffusione della malattia.

Conoscere l'epidemia

In Italia al momento con oltre due milioni tamponi effettuati (Ministero della Salute 3/5) non si è ancora in grado di proiettare le caratteristiche delle persone sottoposte a test sulla popolazione generale. Per effettuare generalizzazioni all'intera popolazione non occorrono campioni necessariamente di dimensione elevata, occorrono campioni congeniati sotto rigorosi criteri statistici. Un'alternativa diagnostica apparentemente meno accurata è costituita dai test sierologici che però possono dar luogo a molti falsi negativi. Tuttavia anche un test poco accurato può portare a risultati affidabili sull'immunità della popolazione se applicato ad un campione probabilistico e rappresentativo. Questo richiede di individuare con precisione la popolazione, l'unità statistica e l'unità di rilevazione. Uno dei passaggi sanitari fondamentali individuati dal ministro della Salute per avviare la fase 2 è la somministrazione di test sierologici ad un campione di 150 mila individui stratificato per età, regione di residenza e attività professionale. Gli esami serviranno a capire quanto e come ha circolato il virus nel nostro Paese e valutare il grado di immunizzazione. Questo permetterà di avere un elenco di riferimento (Master sample) da cui impostare studi longitudinali per il monitoraggio di gruppi di popolazione e/o focus su domini di studio che si rilevassero cruciali (aree territoriali focolai di epidemia).

Spiegare e prevedere l'evoluzione del fenomeno

I modelli statistici rappresentano uno strumento fondamentale per comprendere le cause dei fenomeni, per poterne valutare le conseguenze demografiche, economiche e sociali, oltre che per effettuare previsioni. Generalmente rappresenta una semplificazione della realtà, congeniata per evidenziare alcune caratteristiche essenziali dei fenomeni osservati tralasciandone altre giudicate trascurabili per la loro comprensione. Purtroppo, in questi giorni abbiamo assistito troppe volte all'uso inappropriato dei modelli statistici, così come alla raccolta dei dati senza criteri statistici inficiandone, di fatto, la qualità. Un modello poco buono trascura la qualità del dato utilizzato e interpola meccanicamente gli andamenti della curva epidemica, senza tener conto della sua specificità. Questo tipo di modello può anche dare

frutti discreti a fenomeno in corso (ad esempio dopo il passaggio del picco epidemico), ma difficilmente riesce a prevedere accuratamente i punti di svolta. Nei modelli statistici, non va dimenticato, la misura dell'incertezza, associata ai risultati e al processo produttivo dei dati su cui il modello si basa, è sempre contemplata. Dei modelli statistici esplicativi e previsivi continueremo ad averne bisogno per uscire da questa pandemia con buona pace di chi li critica a priori, ma purché essi siano ben congeniati, utilizzati con dati di buona qualità, e senza improvvisazioni, in altri termini senza fare gli “apprendisti stregoni”.