

# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
1.1	Qualche esempio di uso della statistica . . . . .	5
1.2	Fenomeni individuali e collettivi . . . . .	6
1.3	Statistica descrittiva ed inferenza statistica . . . . .	6
1.4	Cenni storici . . . . .	7
<b>2</b>	<b>Alcuni concetti iniziali</b>	<b>10</b>
2.1	Premessa . . . . .	10
2.2	Unità e collettivo . . . . .	10
2.3	Caratteri e modalità . . . . .	11
<b>3</b>	<b>La rilevazione statistica</b>	<b>13</b>
3.1	Introduzione . . . . .	13
3.2	Le indagini campionarie . . . . .	14
3.2.1	Campionamento casuale . . . . .	14
3.3	Il trascorrere del tempo nelle rilevazioni . . . . .	16
3.4	Rilevazione mediante questionario . . . . .	16
3.5	Codifica e memorizzazione . . . . .	17
3.6	Spoglio e tabulazione . . . . .	17
<b>4</b>	<b>Tipi di dati statistici</b>	<b>19</b>
4.1	Distribuzioni unitarie e distribuzioni di frequenza . . . . .	19
4.2	Distribuzioni semplici . . . . .	19
4.3	Distribuzioni in classi . . . . .	21
4.3.1	Ulteriori commenti sulla chiusura delle classi . . . . .	23
4.4	Caratteri trasferibili . . . . .	24
4.5	Fenomeni di stato e di movimento . . . . .	24
4.6	Distribuzioni doppie e multiple . . . . .	24
4.7	Distribuzioni di quantità . . . . .	25
4.8	Distribuzioni ponderate . . . . .	26
4.9	Serie storiche . . . . .	27
4.10	Serie territoriali . . . . .	27
<b>5</b>	<b>Prime analisi delle distribuzioni semplici</b>	<b>28</b>
5.1	Simboli matematici e cenni sulle sommatorie . . . . .	28
5.2	Frequenze relative e percentuali . . . . .	29

5.3	Densità media e uniforme distribuzione . . . . .	30
5.4	Frequenze cumulate e funzione di ripartizione . . . . .	31
<b>6</b>	<b>Rapporti e numeri indici</b>	<b>34</b>
6.1	Introduzione . . . . .	34
6.2	Rapporti di derivazione . . . . .	35
6.3	Numeri indici . . . . .	35
6.3.1	Indici a base fissa ed a base mobile . . . . .	35
6.3.2	Variazioni relative o percentuali . . . . .	37
<b>7</b>	<b>Rappresentazioni grafiche</b>	<b>38</b>
7.1	Introduzione . . . . .	38
7.2	Distribuzioni con caratteri qualitativi . . . . .	39
7.3	Distribuzioni in classi . . . . .	40
7.4	Rappresentazioni di serie storiche . . . . .	41
<b>8</b>	<b>Valori Medi</b>	<b>44</b>
8.1	Introduzione . . . . .	44
8.2	Condizioni di coerenza . . . . .	44
8.3	Definizione e calcolo di alcune medie . . . . .	46
8.3.1	Media aritmetica . . . . .	46
8.3.2	Mediana . . . . .	48
8.3.3	Media geometrica . . . . .	51
8.3.4	Media armonica . . . . .	52
8.3.5	Moda . . . . .	53
8.4	Interpretazione e scelta del valore medio . . . . .	53
8.4.1	Consistenza . . . . .	53
8.4.2	Monotonia . . . . .	54
8.4.3	Medie associative . . . . .	54
8.4.4	Le medie come centri . . . . .	57
8.5	Quartili e quantili . . . . .	58
8.6	Applicazioni delle medie ai rapporti statistici . . . . .	60
8.6.1	Variazioni medie relative . . . . .	60
8.6.2	Numeri indici complessi . . . . .	60
<b>9</b>	<b>Variabilità, concentrazione e asimmetria</b>	<b>63</b>
9.1	Introduzione . . . . .	63
9.1.1	Principi di coerenza . . . . .	64
9.2	Indici assoluti di variabilità . . . . .	65
9.2.1	Scostamenti medi . . . . .	66
9.2.2	Differenze medie . . . . .	68
9.2.3	Intervalli di variazione . . . . .	69
9.3	Indici percentuali di variabilità . . . . .	70
9.4	La concentrazione . . . . .	70
9.4.1	Curva di concentrazione . . . . .	71
9.4.2	Indici di concentrazione . . . . .	74
9.5	Indici di asimmetria . . . . .	75

<b>10</b>	<b>Interpolazione statistica</b>	<b>77</b>
10.1	Introduzione . . . . .	77
10.2	Interpolazione grafica ed analitica . . . . .	77
10.2.1	Interpolazione matematica . . . . .	78
10.3	Il metodo dei minimi quadrati . . . . .	79
10.3.1	Interpolazione mediante una retta . . . . .	79
10.3.2	Funzioni non lineari nella variabile indipendente . . . . .	82
10.3.3	Funzioni non lineari nei parametri . . . . .	84
<b>11</b>	<b>Dipendenza e associazione</b>	<b>86</b>
11.1	Introduzione . . . . .	86
11.2	Profili delle distribuzioni condizionate . . . . .	87
11.3	Dipendenza e indipendenza . . . . .	88
11.4	Dipendenza in media . . . . .	92
<b>12</b>	<b>Regressione e concordanza</b>	<b>95</b>
12.1	Introduzione . . . . .	95
12.2	Distribuzioni unitarie e di frequenza . . . . .	96
12.3	Regressione lineare . . . . .	97
12.3.1	Distribuzioni di quantità e regression lineare . . . . .	101
12.4	Regressione non lineare . . . . .	103
12.5	Concordanza e discordanza . . . . .	105
<b>13</b>	<b>Probabilità</b>	<b>107</b>
13.1	Introduzione . . . . .	107
13.2	Incertezza, spazio degli eventi, probabilità . . . . .	108
13.3	Alcune nozioni di algebra degli eventi . . . . .	109
13.4	Probabilità . . . . .	111
13.5	Probabilità condizionata . . . . .	113
13.6	Indipendenza di eventi . . . . .	114
13.7	Formula di Bayes . . . . .	115
<b>14</b>	<b>Variabili casuali discrete</b>	<b>117</b>
14.1	La nozione di variabile casuale . . . . .	117
14.2	Variabili casuali discrete . . . . .	117
14.3	Media e varianza di una v.c. discreta . . . . .	119
14.4	Variabile casuale di Bernoulli . . . . .	121
14.5	Variabile casuale binomiale . . . . .	122
14.6	Variabile casuale di Poisson . . . . .	124
<b>15</b>	<b>Variabili casuali continue</b>	<b>126</b>
15.1	Introduzione . . . . .	126
15.2	Variabili casuali continue . . . . .	126
15.3	Media e varianza di una v.c. continua . . . . .	127
15.4	Variabile casuale normale o di Gauss . . . . .	127
15.5	Variabili casuali gamma e chi-quadrato . . . . .	130
15.6	Variabile casuale $t$ di Student . . . . .	132

<b>16 Variabili casuali multidimensionali</b>	<b>134</b>
16.1 Disuguaglianza di Tchebycheff . . . . .	134
16.2 Variabili casuali multidimensionali . . . . .	134
16.3 Variabili casuali indipendenti . . . . .	137
16.4 Combinazioni lineari di variabili casuali . . . . .	138
16.5 La legge dei grandi numeri . . . . .	140
16.6 Il teorema del limite centrale . . . . .	141
<b>17 La stima puntuale e per intervallo</b>	<b>144</b>
17.1 Inferenza statistica . . . . .	144
17.2 Statistiche e stimatori . . . . .	145
17.3 Proprietà degli stimatori . . . . .	146
17.4 Distribuzioni campionarie . . . . .	148
17.5 Intervalli di confidenza . . . . .	150
17.5.1 Intervallo di confidenza per $\mu$ . . . . .	151
17.5.2 Intervallo di confidenza per $\sigma^2$ . . . . .	153
17.5.3 Intervallo di confidenza per $p$ . . . . .	155
17.6 Problemi di determinazione dell'ampiezza campionaria . . . . .	156
<b>18 Verifica di ipotesi</b>	<b>158</b>
18.1 Introduzione . . . . .	158
18.2 Nozioni generali . . . . .	159
18.3 Verifica di ipotesi su $\mu$ . . . . .	159
18.4 Verifica di ipotesi per $\sigma^2$ . . . . .	165
18.5 Verifica di ipotesi su $p$ . . . . .	166
18.6 Potenza del test . . . . .	168

# Capitolo 1

## Introduzione

### 1.1 Qualche esempio di uso della statistica

La **Statistica** si occupa dei principi scientifici su cui dovrebbe basarsi la raccolta, l'elaborazione e l'utilizzazione delle informazioni concernenti cosiddetti *fenomeni collettivi*. Cominciamo considerando alcuni esempi.

**Dati anagrafici.** In ciascun comune vengono raccolte informazioni concernenti le nascite, i matrimoni, le morti e i cambiamenti di residenza di cittadini italiani. In particolare, nel caso delle nascite vengono acquisite varie notizie fra cui la data dell'evento, il peso ed il sesso del neonato, l'età e la professione dei genitori. Contando il numero di nascite avvenute in una serie di anni successivi e confrontando i valori ottenuti con l'ammontare della popolazione residente, è possibile studiare l'andamento della natalità. Tutto questo, oltre a risolvere problemi pratici, come ad esempio prevedere quale sarà fra alcuni anni il numero di studenti iscritti alle scuole elementari o il numero dei pensionati in Italia, può anche servire di base per la formulazione e la verifica di teorie generali sullo sviluppo delle popolazioni umane.

**Indagini ISTAT sulle forze di lavoro.** Con periodicità trimestrale viene selezionato un campione di famiglie e sui membri di queste si raccolgono notizie circa la condizione lavorativa (occupato, disoccupato, pensionato, casalinga, studente o altro), l'eventuale professione ed il ramo di attività in cui questa si esercita, l'eventuale tempo trascorso dopo la perdita dell'ultima occupazione ed altro. Queste informazioni servono di base per valutare l'andamento della disoccupazione a livello nazionale e ad istituire confronti fra regioni o fra rami di attività economica.

**Controllo di qualità industriale.** In alcuni moderni processi industriali viene rilevato automaticamente il peso e alcune altre caratteristiche da cui dipende la qualità dei singoli pezzi prodotti in serie. Lo studio di tali dati può fornire indicazioni tempestive circa un possibile deterioramento progressivo degli standard qualitativi. Gli stessi dati sono la base per individuare interventi in grado di minimizzare il grado di eterogeneità del prodotto.

## 1.2 Fenomeni individuali e collettivi

Per chiarire cosa si intenda per *fenomeno collettivo*, riprendiamo l'esempio dei dati anagrafici; si tratta di una attività che assolve a funzioni burocratiche e amministrative che richiedono l'accessibilità di dati nominativi (individuali) e statistici. Quando un impiegato riempie un modulo o compila un certificato di nascita svolge funzioni amministrative. È invece una attività specificamente statistica quella di conteggiare il numero di nati in ciascun giorno o in ciascun mese, magari distinti secondo il sesso o l'età della madre.

Una situazione analoga si verifica nelle rilevazioni contabili di una impresa: La trascrizione delle singole operazioni (vendita, pagamento di fatture o altro) in modo nominativo ha natura amministrativa. Quando invece consideriamo l'insieme delle operazioni di un certo tipo, ad esempio vendite, e le classifichiamo a secondo del tipo di pagamento, della residenza del cliente, o del mese in cui è avvenuta l'operazione e le analizziamo in complesso per valutare la salute dell'impresa o per fare previsioni di breve periodo, stiamo svolgendo attività di tipo statistico.

In sostanza, al burocrate le notizie anagrafiche interessano a livello **individuale** per documentarle così come sono mentre allo statistico interessano solo per giungere a valutazioni d'insieme relative a particolari gruppi di eventi omogenei rispetto al tempo, al luogo in cui si sono verificati, al sesso, l'età o la professione delle persone coinvolte, producendo quindi informazioni **collettive** che cioè riguardano contemporaneamente insieme più o meno ampi di casi individuali. Un esempio analogo è quello del medico: egli ha bisogno della singola scheda clinica per prendersi cura di ciascun paziente, tenendo conto di tutta la specificità individuale, mentre lo statistico cercherà piuttosto di capire se, per esempio, fra i pazienti con abitudini pressoché simili al fumo, quelli esposti più a lungo sono in media anche i casi più gravi. Tuttavia anche il medico, sia pure a modo suo, non potrà fare a meno di seguire un ragionamento statistico quando, nel consigliare una terapia, sceglierà quella che, in casi simili, si è dimostrata più spesso efficace.

## 1.3 Statistica descrittiva ed inferenza statistica

I problemi connessi alla corretta individuazione delle informazioni di interesse, alla raccolta ed alla loro presentazione sistematica sono, in un certo senso, le più importanti perché producono i **dati statistici** (chiamati semplicemente *statistiche* nel linguaggio corrente), cioè la materia prima dalla cui qualità dipendono tutte le operazioni successive. I problemi connessi con la rilevazione variano enormemente a seconda del contesto applicativo in cui si opera e con la dimensione dell'indagine. Una trattazione generale sarebbe fuori luogo in un corso elementare; ce ne occuperemo quindi solo in modo sommario nel Cap. 3. Aspetti più specifici connessi alla raccolta di dati *demografici*, *economici* o *sociali* sono oggetto di studio nei corsi di Demografia, Statistica Economica e Statistica Sociale.

Anche ammesso di disporre di dati statistici perfetti, non è affatto ovvio quali conclusioni se ne possano trarre. Supponiamo ad esempio di voler stabilire di quanto sono aumentati i prezzi al consumo in una certa città rispetto al mese precedente. Alcuni prezzi saranno aumentati in modo vistoso, altri saranno variati di poco, qualche altro potrebbe addirittura essere diminuito. Dall'esame di questi dati, persone diverse potrebbero giungere a conclusioni anche molto differenti, magari in buona fede. Analogamente l'esame dei

dati concernenti il fatturato di imprese industriali appartenenti a diversi settori può fornire impressioni contrastanti circa il livello di concentrazione raggiunto. In generale sarebbe opportuno disporre di strumenti concettuali per sottoporre ad un'analisi razionale le diverse tesi che potrebbero essere formulate sulla base degli stessi dati. La **statistica descrittiva** si occupa appunto del modo di sintetizzare i dati statistici per giungere a delle conclusioni scientificamente comparabili.

Quando il demografo cerca di sintetizzare i dati sulla natalità per giungere a confrontare la situazione delle varie regioni italiane mediante il calcolo di opportuni tassi di natalità, egli si avvale dei principi della statistica descrittiva; quando invece cerca di utilizzare i suoi dati per prevedere lo sviluppo della popolazione nei prossimi anni, egli sta andando oltre le informazioni a sua disposizione. Anche il tecnico che, sulla base dei dati raccolti su un piccolo campione di confezioni prodotte in serie, cerca di valutare la qualità dell'intera partita, sta in certo senso giocando d'azzardo, cioè facendo affidamento su una *presumibile regolarità dei fenomeni collettivi*. Non è sostanzialmente diverso il comportamento del medico che valuta l'efficacia di una terapia da applicare ad un nuovo paziente sulla base dei risultati ottenuti in passato applicando la stessa terapia a pazienti simili. Per molti di questi problemi, l'esperienza, l'intuito ed una certa dose di buon senso consentono di individuare delle soluzioni ragionevoli. L'**inferenza statistica** fornisce degli schemi concettuali in cui inquadrare tali modi di procedere per individuare sotto quali ipotesi razionali ciascuno di essi usa in modo ottimale le informazioni disponibili.

## 1.4 Cenni storici

La concezione di fondo della *statistica* ed i metodi che a questa si ispirano vengono presentati in questo corso in modo sostanzialmente diverso da come avveniva nel secolo scorso quando la materia venne introdotta ufficialmente nell'insegnamento universitario italiano; c'è quindi da aspettarsi che essa continui ad evolversi magari anche più rapidamente in futuro. Nel seguito viene fornito un quadro riassuntivo dei principali sviluppi che hanno condotto alla moderna concezione della Statistica.

Le radici della *statistica*, come è stata definita sommariamente in questo capitolo, si possono rintracciare nello spirito che ha animato lo sviluppo della scienza moderna con l'osservazione sistematica dei fenomeni naturali. Secondo alcuni, ad esempio, il modo come Galilei utilizzava i risultati delle sue osservazioni per formulare le sue teorie o per dimostrare la falsità di quelle precedenti, può essere considerato una anticipazione della statistica moderna.

Un piccolo libricino, pubblicato a Londra nel 1662 da John Graunt, è considerato da molti una pietra miliare nello sviluppo della statistica. RegISTRAZIONI attendibili delle nascite e delle morti venivano tenute da alcuni decenni nelle parrocchie londinesi. Graunt concepì l'idea (rivoluzionaria per i suoi tempi) di contare il numero di morti distintamente secondo il sesso, la classe d'età e la causa di morte a seconda dell'anno in cui si era verificato il decesso; analogamente per le nascite ed i matrimoni. In questo modo, delle informazioni di natura amministrativa venivano trasformate in dati statistici. Egli inoltre assumeva implicitamente che le frequenze ottenute in questo modo fossero determinate dal concorso di cause di fondo sostanzialmente stabili a meno di piccole perturbazioni dovute a fattori occasionali. Cercò quindi di approssimare e ricondurre a dei rapporti semplici i suoi dati in modo da far emergere presumibili regolarità. Notò, ad esempio, che il numero di coloro che,

secondo la dichiarazione dei parenti, erano deceduti per cause naturali variava di poco da un anno all'altro ma si raddoppiava in occasione di pestilenze; concluse quindi che, quasi certamente, un gran numero di morti di peste venivano spacciati per morti naturali. Un altro esempio istruttivo del modo di procedere di Graunt può essere il seguente. Avendo notato dai registri di alcune parrocchie che durante un anno si avevano circa 3 morti ogni 11 famiglie, dai 23 mila decessi avvenuti a Londra nel 1660 stabilì che ci dovevano essere circa 84 mila famiglie, cioè un valore simile a quello a cui arrivava partendo invece dal numero delle nascite.

La corrente di pensiero che si ispirava alle ricerche di Graunt venne denominata **aritmica politica** e divenne molto popolare anche al di fuori dell'Inghilterra. Ad esempio, verso la fine del 1700, molti studiosi francesi cercavano di stimare la popolazione totale sulla base del rapporto fra numero di abitanti e nascite annuali; tale rapporto era circa pari a 26 in un ampio campione di parrocchie dello stato in cui si era proceduto al conteggio. Come è facile immaginare, questo modo di procedere, che oggi chiameremmo *induttivo*, può condurre a conclusioni molto azzardate e magari arbitrarie se non si basa su una buona dose di intuizione e senso critico.

L'idea che il *calcolo delle probabilità* potesse fornire una base razionale per valutare l'accuratezza delle conclusioni ottenute con i metodi degli aritmetici politici, sembra emergere agli inizi del '700 con Bernoulli. Sembra che egli sia stato fra i primi a tentare di estendere le applicazioni della probabilità dal terreno artificioso dei giochi d'azzardo a quello molto più complesso dei fenomeni naturali. Durante la prima metà del secolo scorso, le scoperte di Laplace, Gauss ed altri dettero un forte impulso alla raccolta ed all'analisi sistematica di dati statistici. La figura fondamentale di questo periodo è Quetelet: secondo alcuni egli sarebbe il vero fondatore della statistica moderna, mentre secondo altri egli si sarebbe limitato a divulgare ed applicare, in modo non sempre corretto, il calcolo delle probabilità ai fenomeni sociali. A titolo di esempio vale la pena di ricordare il concetto di *uomo medio* introdotto da Quetelet in un suo libro del 1835 e diventato poi molto popolare. Si trattava di un individuo immaginario che aveva le caratteristiche fisiche (statura e peso) e morali (tendenza al crimine, all'ubriachezza o a contrarre certe malattie) pari ai valori medi effettivamente riscontrabili all'interno di una popolazione. Questa concezione si basava sulla constatazione, apparentemente paradossale, che se si considerava la situazione di uno Stato in anni successivi, non solo restava quasi invariata la statura media dei militari di leva, ma persino la percentuale di suicidi dell'anno in corso poteva essere prevista con buona approssimazione sulla base dei dati relativi agli anni precedenti, una circostanza che pareva contraddire il principio del libero arbitrio. A prescindere dalla controversia circa i meriti di Quetelet, il rapporto fra statistica e probabilità si è rivelato particolarmente fecondo ed è alla base dei più importanti progressi che hanno condotto alla statistica moderna.

Occorre infine menzionare un filone di ricerche, denominato **statistica universitaria**, che trae origine dagli antichi libri di viaggio dove venivano fornite informazioni e curiosità sui paesi stranieri. Con questa disciplina, sviluppatasi nelle università tedesche tra il '600 ed il '700 soprattutto per merito di Conring ed Achenwall, ci si proponeva di descrivere in modo sistematico le principali caratteristiche di uno stato. Benché sembri che il termine *statistica* sia stato introdotto per la prima volta proprio in questo contesto, i metodi della *statistica universitaria* erano più vicini a quelli della storia e della geografia che a quelli della statistica moderna. Tuttavia questa corrente di pensiero era ancora predominante in

Italia agli inizi del secolo scorso. Ad esempio, uno dei testi più autorevoli dell'epoca, scritto da Melchiorre Gioja, iniziava definendo la statistica come *la descrizione economica delle nazioni in un'epoca determinata*: essa addita le fonti della loro ricchezza, i metodi con cui le distribuiscono, . . . La statistica comprende dunque *quella somma di cognizioni relative ad un paese che nel corso giornaliero degli affari possono essere utili a ciascuno o alla maggioranza dei suoi membri, od al governo, . . .* Si tratta di una definizione che descrive abbastanza bene quelle che sono anche oggi le attività principali dell'Istituto Nazionale di Statistica (ISTAT) piuttosto che la *statistica* come disciplina scientifica.