CAPITOLO SESTO –- STATISTICA INDUTTIVA

§1. Verifica di un’ipotesi: un caso particolare

Nell’enunciato di numerosi esempi ed esercizi proposti nei capitoli precedenti si è fatto riferimento a monete o dadi non truccati, nei quali i possibili esiti sono equiprobabili. Abbiamo tra l’altro imparato a risolvere problemi di questo tipo:

Effettuando 200 lanci di una moneta buona, qual è la probabilità di ottenere un numero di teste compreso tra 90 e 110? Oppure di ottenere almeno 95 teste? Oppure di ottenere meno di 100 teste? Lanciando 300 volte un dado buono, qual è la probabilità di ottenere 6 per un numero di volte inferiore a 50? O compreso tra 40 e 60?

La risoluzione di tali quesiti richiede un certo lavoro, ma, una volta introdotti gli strumenti adeguati, non comporta particolari difficoltà concettuali.

Intendiamo ora ribaltare il problema, ossia rispondere in qualche modo ad una domanda quale:

(°) una certa moneta è buona?

A livello intuitivo si può rispondere nel modo seguente: lanciamo la moneta un certo numero di volte e controlliamo se il numero di teste uscite è approssimativamente uguale a quello delle croci. Questa strategia, pur nella sua plausibilità, è troppo vaga per poter essere utilmente impiegata: occorre precisare meglio cosa si intende per «approssimativamente uguale».

Prima di affrontare la questione è importante sottolineare un aspetto. Se lanciando 10 volte una moneta otteniamo 10 teste è ragionevole sospettare che la moneta non sia buona; tuttavia non lo potremo affermare “con certezza”. L’evento in questione non è impossibile con una moneta buona; ha solo una probabilità molto bassa di verificarsi: $\left(\frac{1}{2}\right)^{10} \cong 0{,}000976$ (e, come si è sottolineato a proposito dei numeri in ritardo sulle ruote del lotto, talvolta si verificano eventi a bassa probabilità).

In altre parole, bisogna tener presente che la risposta alla (°) non sarà mai «sì» o «no» al di là di ogni dubbio. Così, ad esempio, se su 1.000 persone scelte a caso in una popolazione di 100.000, più di 500 sono mancine, non potremo essere certi che la probabilità che una persona sia mancina è maggiore di1/2: può darsi che il caso abbia fatto concentrare nel campione un numero di persone mancine molto superiore alla frequenza nella popolazione. In sostanza, domande come la (°) non ammettono una risposta definitiva.

L’idea allora è la seguente: fissiamo un elevato valore della probabilità, detto *livello di sicurezza*, ad esempio 0,95, e studiamo come si comporta una moneta buona nel 95% dei casi. Se la nostra moneta non segue questo comportamento, potremo rifiutare che essa sia buona, con la certezza di non prendere una decisione errata almeno nel 95% dei casi.

Poniamo allora questo problema: data una moneta buona ed eseguendo, ad esempio, 100 1anci, determinare 1’intervallo intorno a 50 nel quale, con probabilità 0,95, si trova il numero di teste ottenute.

Si tratta, in sostanza, di un problema inverso rispetto a quelli richiamati all’inizio del paragrafo, nei quali era noto l’intervallo e si doveva determinare la probabilità. Qui è nota la probabilità e si deve determinare l’intervallo. Volendo determinare k in modo che X = «numero di teste ottenute in 100 lanci» soddisfi la condizione:

$$p(50 - k \leq X \leq 50 + k) = 0{,}95$$

sfruttando l'approssimazione normale alla distribuzione binomiale ($m = np = 50$: $\sigma = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 5$), si può determinare k in modo che:

$$p\left(-\frac{k}{5} \le \frac{X - 50}{5} \le \frac{k}{5}\right) = 0{,}95$$

ossia:

$$p\left(-\frac{k}{5} \le X^* \le \frac{k}{5}\right) = 0{,}95$$

Con la notazione introdotta nel quinto capitolo si ha:

$$p\left(-\frac{k}{5} \le X^* \le \frac{k}{5}\right) = 2N\left(\frac{k}{5}\right)$$

Da $2N\left(\frac{k}{5}\right) = 0{,}95$ segue $N\left(\frac{k}{5}\right) = 0{,}475$.

Nella Tavola alla fine del capitolo quinto si legge che il valore $N(z) = 0{,}475$ si ottiene per $z = 1{,}96$. In definitiva $\frac{k}{5} = 1{,}96$, da cui $k = 9{,}8$.

Ciò significa che, lanciando 100 volte una moneta buona, la probabilità di ottenere testa un numero di volte compreso tra 50 – 9,8 e 50 + 9,8, ossia tra 41 e 59 (estremi inclusi), è 0,95.
Quindi, se lanciando una moneta qualsiasi 100 volte si ottiene testa un numero di volte inferiore a 41 o superiore a 59, si hanno due possibilità alternative:

a) la moneta non è buona
b) la moneta è buona e si è verificato un evento con probabilità 0,05.

Ne segue che, *con probabilità* 0,95, *la moneta non è buona*.

In conclusione, possiamo adottare la seguente strategia: lanciamo la moneta 100 volte; se otteniamo testa un numero di volte inferiore a 41 o superiore a 59 rifiutiamo l'ipotesi che la moneta sia buona. Siamo sicuri di non sbagliare nel 95% dei casi.

I calcoli possono essere ripetuti scegliendo un diverso livello di sicurezza, ad esempio 0,99. In tal caso si trova $k \cong 12{,}9$. Se in 100 lanci si ottiene testa un numero di volte minore di 38 o maggiore di 62 la moneta non è buona con probabilità 0,99. Il maggior grado di certezza che vogliamo raggiungere per rifiutare che la moneta sia buona ha ristretto l'intervallo delle possibilità.

Osservazione. Il procedimento finora esposto è significativo nel senso che consente di rifiutare l'ipotesi che la moneta sia buona se si ottiene testa un numero di volte al di fuori dell'intervallo determinato, con un'elevata probabilità di non prendere una decisione errata. Si osservi, tuttavia, che se si ottiene testa un numero di volte compreso nell'intervallo non è lecito dedurre che, con elevata probabilità, la moneta sia buona. Supponiamo che nei 100 lanci si sia ottenuto testa 55 volte. Allora non possiamo rifiutare, ai livelli di sicurezza che abbiamo introdotto, che la moneta sia buona. Ma non possiamo sostenere che con elevata probabilità $p = 0{,}5$. Il nostro risultato concorda altrettanto bene con il valore $p = 0{,}53$ o con $p = 0{,}55$ e cosi via. Questo aspetto del problema sarà approfondito nel §4.

Consideriamo ora il caso di un dado. Il problema è più complesso essendovi 6 possibili esiti. Si può tuttavia procedere come nel caso della moneta se prendiamo in esame l'uscita di una sola faccia, ad esempio quella con il numero 6.

Lanciamo 360 volte il dado. Se è buono, la variabile casuale X = «numero dei 6 usciti» ha una distribuzione binomiale con E(X) = $np = 360 \cdot \frac{1}{6} = 60$ e $\sigma = \sqrt{360 \cdot \frac{1}{6} \cdot \frac{5}{6}} \cong 7{,}07$.

La variabile $X^* = \frac{X - 60}{7{,}07}$ ha distribuzione approssimativamente normale.

Come in precedenza, affinché $p(60 - k \le X \le 60 + k) = 0{,}95$, si determina k in modo che $p\left(-\frac{k}{7{,}07} \le X^* \le \frac{k}{7{,}07}\right) = 0{,}95$, e si ottiene $\frac{k}{7{,}07} = 1{,}96$ da cui, in defintiva, $k \cong 13{,}86$.

Quindi: con probabilità 0,95 il numero di volte in cui esce 6, lanciando 360 volte un dado buono, è compreso tra 60 – 13,86 e 60 + 13,86, ossia tra 47 e 73, estremi inclusi.

La strategia è allora la seguente: *se*, *lanciando un qualsiasi dado* 360 *volte*, *otteniamo il* 6 *un numero di volte inferiore a* 47 *o superiore a* 73, *al* 95% *il dado non è buono*.

In generale, per trovare l'intervallo nel quale cadono *con probabilità* 0,95 i valori di una variabile casuale X con distribuzione normale di valore medio m e deviazione standard σ, si può sfruttare la seguente formula:

$$m - 1{,}96\sigma \le X \le m + 1{,}96\,\sigma \qquad (1)$$

mentre, con *probabilità* 0,99, l'intervallo è:

$$m - 2{,}58\sigma \le X \le m + 2{,}58\,\sigma \qquad (2)$$

Esempio. Su 1.000 nascite si sono osservati 535 maschi. Al livello di sicurezza 0,95 si può sostenere l'equiprobabilità delle nascite maschili e femminili? E al livello 0,99?

Se le nascite maschili e femminili fossero equiprobabili, il numero dei maschi avrebbe valore medio 500 e deviazione standard $\sigma = \sqrt{1.000 \cdot \frac{1}{2} \cdot \frac{1}{2}} \cong 15{,}8$.

Al livello di sicurezza 0,95 (detto X il numero dei maschi) dalla (1) si ottiene:

$$500 - 1{,}96 \cdot 15{,}8 \le X \le 500 + 1{,}96 \cdot 15{,}8$$

ossia $470 \le X \le 530$.

Poiché il numero osservato di 535 maschi supera 530, si può rifiutare l'ipotesi di equiprobabilità.

Al livello 0,99, risultando, per la (2), $460 \le X \le 540$ l'ipotesi di equiprobabilità non può essere rifiutata.

In definitiva, il valore osservato di 535 è sufficientemente lontano da 500 per consentirci di rifiutare l'ipotesi al livello 0,95. Tuttavia non è così lontano da 500 da garantirci di poter rifiutare l'ipotesi al livello 0,99.

§2 Il campionamento

Spesso si vogliono ottenere informazioni relative ad un dato fenomeno ed è materialmente impossibile, oppure troppo lungo e costoso, procurarsi i dati in modo completo ed esauriente. Uno dei metodi più sfruttati per stimare le grandezze incognite è il ricorso ad un *campione*.

Prima di occuparci, nel prossimo paragrafo, del problema della stima di un parametro incognito, esponiamo brevemente qualche caratteristica del campionamento.

Siano dati, ad esempio, i seguenti numeri:

1, 3, 6, 8, 14

La loro media aritmetica è $\overline{X} = 6{,}4$; la mediana è 6 e la varianza (quadrato dello scarto quadratico medio) è $s^2 = 20{,}24$.

Consideriamo tutti i possibili campioni costituiti con 3 dei 5 numeri dati e di ciascuno di essi consideriamo la media aritmetica e la mediana.

Si ottiene la seguente tabella:

Campione	X_M = media del campione	X_{Me} = mediana del campione
1, 3, 6	10/3	3
1, 3, 8	4	3
1, 3, 14	6	3
1, 6, 8	5	6
1, 6, 14	7	6
1, 8, 14	23/3	8
3, 6, 8	17/3	6
3, 6, 14	23/3	6
3, 8, 14	25/3	8
6, 8, 14	28/3	8

I possibili campioni sono 10. Se immaginiamo di aver estratto a sorte un campione, la media del campione è una variabile casuale X_M i cui possibili valori sono riportati nella seconda colonna; ciascuno di essi ha probabilità 1/10.

I1 valore medio di X_M risulta:

$$E(X_M) = \frac{10}{3} \cdot \frac{1}{10} + 4 \cdot \frac{1}{10} + 6 \cdot \frac{1}{10} + \ldots + \frac{25}{3} \cdot \frac{1}{10} + \frac{28}{3} \cdot \frac{1}{10} = 6,4$$

Come si vede, e come si potrebbe dimostrare in generale, si ha:

$$\boxed{E(X_M) = \overline{X}}$$

vale a dire *il valore medio di X_M è proprio la media aritmetica dell'intera popolazione*.

Consideriamo ora le mediane.
La mediana del campione è un'altra variabile casuale X_{Me} il cui valore medio risulta:

$$E(X_{Me}) = 3 \cdot \frac{1}{10} + 3 \cdot \frac{1}{10} + 3 \cdot \frac{1}{10} + 6 \cdot \frac{1}{10} \ldots + 8 \cdot \frac{1}{10} + 8 \cdot \frac{1}{10} = 5,7$$

Questa volta $E(X_{Me})$ non coincide con la mediana della popolazione originaria dei 5 numeri.

Un'analoga circostanza si sarebbe verificata se avessimo considerato la variabile casuale X_v = «varianza del campione»: il valore medio di X_v non coincide con la varianza s^2 dell'intera popolazione.

In definitiva, alcuni parametri relativi ai campioni hanno un valore medio che coincide con quello della popolazione; altri parametri non hanno questa proprietà.

Presentiamo ora qualche risultato relativo al caso in cui il campionamento è eseguito con ripetizione, ossia con la possibilità che un oggetto venga scelto più volte.

In tal caso si dimostra che:

(a) $E(X_M) = \overline{X}$ (come nel caso senza ripetizione)

(b) $E(X_{Me}) = \overline{X}$, ossia il valore medio di X_{Me} *non* è, come ci si aspetterebbe, la mediana della popolazione, ma la media aritmetica della popolazione.

(c) $E(X_v) = \frac{n-1}{n} \cdot s^2$ (dove n è il numero di elementi del campione e s^2 la varianza della popolazione). Ossia il valore medio delle varianze dei campioni è $\frac{n-1}{n}$ volte la varianza della popolazione. Se n è abbastanza grande, allora $\frac{n-1}{n}$ è prossimo a 1 e approssimativamente $E(X_v) = s^2$.

In seguito ci occuperemo di casi in cui la popolazione è sufficientemente grande in modo che il campionamento con o senza ripetizione conduca a risultati approssimativamente uguali (se la popolazione è molto grande, la probabilità che un oggetto venga scelto due volte è trascurabile).

Si può dimostrare che, se la popolazione di partenza ha una distribuzione pressoché normale, anche le variabili casuali introdotte (X_M, X_{Me}, X_v) hanno distribuzione normale. Tuttavia, anche se la popolazione non ha distribuzione normale, le variabili in questione hanno ancora una distribuzione approssimativamente normale, purché n sia abbastanza grande ($n \geq 30$ per X_M e X_{Me}; $n \geq 100$ per X_v).

Per il seguito è opportuno rilevare che, come si potrebbe dimostrare rigorosamente, la variabile casuale X_M ha valore medio m e deviazione standard σ tali che:

$$m = E(X_M) = \overline{X}; \qquad \sigma = \frac{s}{\sqrt{n}} \qquad (3)$$

dove $\overline{X}$ e s sono rispettivamente la media aritmetica e lo scarto quadratico medio della popolazione.

Consideriamo infine i successivi esiti del lancio di una moneta, di un dado o di un qualsiasi esperimento di Bernoulli, in cui p è la probabilità di successo.

Una sequenza di n esiti può essere pensata come un campione estratto dall'insieme di tutte le sequenze possibili di n esiti. La frequenza F dei successi è una variabile casuale legata alla X = «numero dei successi» dalla formula:

$$F = \frac{X}{n}$$

ed ha la stessa distribuzione binomiale di X.

Si dimostra che il valore medio e la varianza di F sono date da:

$$E(F) = p; \quad var(F) = \frac{pq}{n} \quad \left(\text{per cui: } \sigma = \sqrt{\frac{pq}{n}}\right) \qquad (4)$$

§3 La stima di una quantità incognita

(A) Stima puntuale

Supponiamo di voler conoscere la statura media degli abitanti di una certa regione. Scartando la possibilità di misurare le altezze di tutte le persone della popolazione, si può ricorrere ad un campione. Si sceglie a caso un certo numero di persone e si calcola la loro altezza.

La media delle altezze ottenute non è altro che un valore della variabile casuale X_M introdotta nel paragrafo precedente.

Poiché il valore medio di X_M è proprio la media aritmetica $\overline{X}$ della popolazione, si dice che X_M è uno *stimatore corretto* di $\overline{X}$ e il valore misurato nel campione può essere assunto come «stima» di $\overline{X}$.

In generale, data una qualsiasi grandezza relativa ad una popolazione, un suo *stimatore corretto* è una qualsiasi variabile casuale il cui valore medio è 1a grandezza in questione.

Si è detto che, se il campionamento è con ripetizione, o se la popolazione è molto ampia, $E(X_{Me}) = \overline{X}$, per cui anche X_{Me} è uno stimatore corretto di $\overline{X}$.

Quando abbiamo più stimatori corretti della stessa grandezza, sorge spontanea la domanda di quale sia migliore o, come si dice, quale sia il più efficiente. Se si ricorda il significato della deviazione standard, è naturale assumere come stimatore più efficiente quello con deviazione standard minore, in quanto i valori assunti saranno più «raccolti» intorno al valore medio (ossia alla grandezza incognita).

Nel caso in esame si dimostra che:

$$\sigma(X_M) < \sigma(X_{Me})$$

per cui X_M è uno stimatore più efficiente di X_{Me}; anzi, si dimostra che X_M è lo stimatore più efficiente di X.

Se, invece, volessimo stimare la mediana delle altezze della popolazione della regione, non potremmo ricorrere a X_{Me} (ossia alla mediana del campione) quale stimatore, poiché $E(X_{Me})$ non è la mediana della popolazione.

Analogamente, per stimare la varianza o lo scarto quadratico medio delle altezze delle persone, non possiamo, a stretto rigore, ricorrere a X_v, in quanto il suo valore medio, come si è detto, è $E(X_v) = \frac{n-1}{n} \cdot s^2$ e non è uguale a s^2. X_v è, come si dice, uno *stimatore distorto* della varianza s^2 della popolazione.

Uno stimatore corretto per la varianza della popolazione è la variabile casuale $\frac{n}{n-1} X_v$ il cui valore medio è proprio s^2. Comunque, per campioni abbastanza grandi, l'errore che si commette stimando la varianza della popolazione con la varianza del campione può essere trascurato.

La variabile casuale F = «frequenza di successi» in un esperimento di Bernoulli è uno stimatore corretto della probabilità di successo. Infatti, come detto nel paragrafo precedente, $E(F) = p$.

Tra l'altro:

X_M come stimatore di $\overline{X}$

F come stimatore di p

hanno, come segue dalle formule del paragrafo precedente, deviazione standard che diminuisce al crescere di n: più numeroso è il campione, più efficiente è la stima ottenuta.

In tutti i casi, la variabile casuale assunta come stimatore fornisce un valore, detto stima puntuale, della grandezza incognita.

(B) Stima intervallare

Nel punto (A) precedente si è visto come stimare, in alcuni semplici casi, una grandezza incognita. È importante ottenere qualche ulteriore informazione circa la «bontà» della stima, ossia determinare un intervallo intorno al valore stimato nel quale cade con elevata probabilità il valore da stimare. Si parla allora di *stima intervallare*.

Ricolleghiamoci con quanto visto nel §1. Supponiamo di avere rifiutato l'ipotesi che una moneta sia buona. Si pone allora per essa il problema di stabilire la probabilità p che esca testa (avendo escluso, ad un certo livello di sicurezza, che sia $p = 1/2$).

Si lancia la moneta un certo numero di volte e si assume la frequenza relativa f dei successi (f è un valore della variabile casuale F = «frequenza dei successi» della quale abbiamo parlato in precedenza) come stima di p. Infatti F è uno stimatore corretto di p.

Cerchiamo ora di andare oltre e di stabilire quanto attendibile sia questa stima. Per le stesse ragioni per le quali non si può essere assolutamente certi che la moneta non sia buona (anche se abbiamo rifiutato l'ipotesi che lo sia), non potremo stabilire con certezza che p coincide con un suo valore stimato f. Possiamo tuttavia, in sintonia con la trattazione del §1, porre il problema in questa forma:

> se f è il valore stimato di p, qual è l'intervallo di centro f nel quale p cade con elevata probabilità, ad esempio 0,95?

Si tratta, in altre parole, di determinare k in modo che $p(f - k \leq p \leq f + k) = 0{,}95$.
Siamo in grado di risolvere questo problema per le variabili casuali come F che hanno distribuzione, anche solo approssimativamente, normale.

Per poter applicare la (1) del §1, occorre determinare:

$$m = E(F) = p \quad \text{e} \quad \sigma = \sigma(F) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p\cdot(1-p)}{n}}$$

ma, in questo caso, sia m che σ dipendono da p, che è incognito. Per i nostri scopi è sufficiente segnalare che si può dimostrare quanto segue: si può ricorrere alla (1) in cui m e σ sono «stimati» sostituendo p con f, purché n sia abbastanza grande (in pratica basta che sia superiore a 30).

***Esempio* 1**. Lanciando una moneta 150 volte, si è ottenuta testa 120 volte. Determinare, al livello di sicurezza 0,95, una stima intervallare della probabilità p di ottenere testa con quella moneta.

Con le notazioni precedenti, $f = \dfrac{120}{150} = 0{,}8$ è la stima puntuale di p; m è stimato con 0,8 e σ con $\sqrt{\dfrac{0{,}8\cdot 0{,}2}{150}} \cong 0{,}03266$. Per la (1) del §1, si ha:

$$0{,}8 - 1{,}96 \cdot 0{,}03266 \leq p \leq 0{,}8 + 1{,}96 \cdot 0{,}03266$$

ossia:

$$0{,}736 \leq p \leq 0{,}864$$

Si può essere certi al 95% che la probabilità di ottenere testa con quella moneta è compresa tra 0,736 e 0,864.
Se ripetiamo i conti al livello di sicurezza 0,99, ossia con la (2) del §1, si trova:

$$0{,}716 \leq p \leq 0{,}884$$

Come è ovvio, per essere più sicuri, dobbiamo accettare un intervallo più ampio di valori.

***Esempio* 2**. Stimare, al livello di sicurezza 0,95, quante sono le persone mancine di una popolazione di 100.000 individui, sapendo che in un campione di 200 unità si sono osservati 60 mancini.
In questo caso p è la probabilità che una persona scelta a caso sia mancina, ossia:

$$p = \frac{\text{numero totale mancini}}{100.000}$$

Si ha $f = \dfrac{60}{200} = 0{,}3$ come stima di p e di m; σ è stimato con $\sqrt{\dfrac{0{,}3\cdot 0{,}7}{200}} \cong 0{,}0324$.

Per la (1) del §1 si ha: $\quad 0{,}3 - 1{,}96 \cdot 0{,}0324 \leq p \leq 0{,}3 + 1{,}96 \cdot 0{,}0324$
ossia: $\quad 0{,}2365 \leq p \leq 0{,}3635$

Al livello di sicurezza 0,95 il numero delle persone mancine nella popolazione è compreso tra 23.650 e 36.350.

***Esempio* 3**. *Stima del valore medio*. Stimare, al livello di sicurezza 0,95, la durata media delle lampadine prodotte da una ditta, sapendo che, in un campione di 625 lampadine, si è osservata una durata media di 800 *h* con scarto quadratico medio di 100 *h*.

Essendo, per le (3) del §2, $E(X_M) = \overline{X}$; $\sigma = \frac{s}{\sqrt{n}}$, per poter applicare la (1) del §1, dobbiamo stimare $\overline{X}$ e s, ossia la durata media e lo scarto quadratico medio della popolazione delle lampadine. Avendo a disposizione un campione abbastanza grande (n = 625), possiamo approssimare $\overline{X}$ e s con i valori 800 *h* e 100 *h* rilevati nel campione, per cui m = 800 *h* e $\sigma = \frac{100\,h}{\sqrt{625}} = 4\,h$. Dalla (1) del §1 si ha:

$$800 - 4 \cdot 1{,}96 \le X_M \le 800 + 4 \cdot 1{,}96 \quad \text{ossia: } 792{,}16 \le X_M \le 807{,}84$$

Quindi, al livello di sicurezza del 95%, la vita media delle lampadine prodotte dalla ditta sarà compresa tra 729,16 *h* e 807,84 *h*.

§4 Verifica di ipotesi: considerazioni generali

Nel §1 abbiamo visto come rifiutare l'ipotesi che una moneta o un dado sia buono con un elevato livello di sicurezza e si è posta l'attenzione sul fatto che non rifiutare un'ipotesi ad un certo livello non significa accettarla allo stesso livello: se, al livello di sicurezza 0,95, non si è respinta l'ipotesi che una moneta sia buona ciò non significa che, con probabilità 0,95, la moneta è buona, ossia con p = 0,5. Prima di presentare qualche altro test intendiamo soffermarci a chiarire questo aspetto.

Uno dei compiti che lo statistico si prefigge è quello di verificare o rifiutare una certa ipotesi; ad esempio: è buono il dado? È buona la moneta? Vi è connessione tra altezza e peso delle persone? Vi è differenza tra uomini e donne nei confronti di una certa prova? È efficace un farmaco? Due campioni di pezzi sono stati prodotti dalla stessa macchina?
A seconda del tipo di domanda sono stati elaborati test appropriati.
Per i nostri scopi è sufficiente considerare solo i più semplici, poiché dal loro esame emergono già le caratteristiche salienti dei procedimenti impiegati nella statistica induttiva.

Chiamiamo *ipotesi nulla*, e la indichiamo con H_0, l'ipotesi che si intende verificare o rifiutare. Nel §1 l'ipotesi nulla era: «la moneta è buona». Nel test statistico, l'ipotesi nulla H_0 viene messa a confronto con un'altra ipotesi H_1, detta *ipotesi alternativa*. Non necessariamente l'ipotesi alternativa è tutto ciò che non rientra in H_0; ad esempio, l'ipotesi che la moneta sia buona (p = 0,5) può essere messa in alternativa a p = 0,6, oppure a p = 0,7, oppure a p < 0,3, e così via.`

Un test è congegnato in modo da consentire una scelta tra l'ipotesi nulla e l'ipotesi alternativa. Tale scelta, in ogni caso, non è eseguibile con assoluta certezza, in quanto vi è sempre un margine di errore.

È opportuno distinguere due tipi di errore:

(a) *errore di primo tipo* è quello che si commette accettando H_1 nel caso in cui sia vera H_0, ossia è la possibilità di respingere H_0 per errore.

(b) *errore di secondo tipo* è quello che si commette accettando H_0 nel caso in cui sia vera H_1.

Nel §1 abbiamo posto l'accento sull'errore del primo tipo: in genere accettavamo che una moneta non fosse buona con una probabilità di errore di 0,05 (ossia 1 – 0,95). È questo l'esatto significato della nostra affermazione di pag. 2: nel caso di 100 lanci, se otteniamo testa un numero di volte inferiore a 41 o superiore a 59, la moneta non è buona con probabilità 0,95. Ora, se si ottiene testa, sempre su 100 lanci, un numero di volte compreso tra 41 e 59 (estremi inclusi) e si accetta che la moneta è buona, si va incontro ad un possibile errore del secondo tipo. La misura dell'errore del secondo tipo dipende da come è formulata l'ipotesi alternativa. In genere, diminuendo la possibilità di errore di primo tipo, aumenta il pericolo di commetterne uno di secondo tipo: passando dal livello di sicurezza 0,95 al livello 0,99 siamo più sicuri di non dichiarare truccata una moneta buona, ma corriamo un maggiore rischio di dichiarare buona una moneta truccata.

È chiaro che un test è tanto migliore quanto più piccole sono entrambe le possibilità di errore. Generalmente si fissa l'errore del primo tipo (di solito 0,05 o 0,0l) e si conduce il test in modo da minimizzare l'errore di secondo tipo. Anzitutto va scelta con oculatezza l'ipotesi alternativa. Ad esempio, se l'ipotesi nulla è H_0 = «la moneta è buona», e si eseguono l00 lanci, il test proposto nel §1 ha un piccolo errore di primo tipo, ma un elevato valore dell'errore di secondo tipo se l'ipotesi alternativa è H_1 = «la moneta non è buona». Tuttavia, se si assume come ipotesi alternativa, ad esempio, H_1 = «p = 0,6», si può ridurre sensibilmente la possibilità di errore del secondo tipo. In parole più semplici, se l'ipotesi che la moneta sia buona è posta in alternativa all'ipotesi che la moneta sia truccata in modo sensibile, il test assicura di poter decidere con elevata sicurezza quale delle ipotesi è vera per una data moneta.

Si può dimostrare che, all'aumentare degli elementi del campione o dei lanci effettuati, per un fissato valore dell'errore di primo tipo, si può diminuire a piacere la probabilità di errore di secondo tipo. L'ipotesi H_0 = «la moneta è buona» può essere posta in alternativa anche all'ipotesi H_1 = «la moneta non è buona» con una piccola probabilità dei due tipi di errore, pur di eseguire un numero di lanci abbastanza elevato.

§5 Il test χ^2 (chi quadro)

(A) Test di significatività

Riprendiamo in esame il caso del lancio del dado. Supponiamo di avere lanciato 360 volte un dado e di avere ottenuto i seguenti risultati:

1	2	3	4	5	6
45	68	70	55	64	58

e di voler verificare l'ipotesi H_0 = «il dado è buono».

Se è vera l'ipotesi, il valore atteso di ciascun esito è 60, per cui si tratta di confrontare la tabella realmente ottenuta con quella dei valori attesi:

1	2	3	4	5	6
60	60	60	60	60	60

e di valutare se le discordanze non sono troppo accentuate.

Una misura delle discordanze tra le due tabelle è la seguente:

$$\chi^2 = \frac{(45-60)^2}{60} + \frac{(68-60)^2}{60} + \frac{(70-60)^2}{60} + \frac{(55-60)^2}{60} + \frac{(64-60)^2}{60} + \frac{(58-60)^2}{60} =$$

$$= \frac{225+64+100+25+16+4}{60} = \frac{434}{60} \cong 7,23$$

Ci chiediamo ora se tale valore ci consente di rifiutare l'ipotesi che il dado sia buono al livello di sicurezza del 95%. Prima di procedere generalizziamo il discorso.
Se sono dati n possibili esiti $e_1, e_2,..., e_n$ e sono state osservate le frequenze $f_1, f_2,..., f_n$, mentre le frequenze teoriche previste in base ad un'ipotesi H_0 sono $p_1, p_2,..., p_n$; poniamo:

$$\chi^2 = \frac{(f_1 - p_1)^2}{p_1} + \frac{(f_2 - p_2)^2}{p_2} + \ldots + \frac{(f_n - p_n)^2}{p_n}$$

Questo numero può essere pensato come un valore di una variabile casuale (in diversi esperimenti si ottengono valori $f_1, f_2,..., f_n$ diversi). Tale variabile casuale ha valore nullo se le frequenze osservate coincidono con quelle teoriche. Se l'ipotesi H_0 formulata è corretta, i valori di χ^2 saranno «abbastanza piccoli». Per determinare quanto «piccoli» devono essere, occorrerebbe conoscere la distribuzione di probabilità di tale variabile casuale. Ai nostri fini è sufficiente ricorrere alla *Tavola riportata alla fine del capitolo*, relativa ai casi del livello di sicurezza 0,95 e 0,99 (e che può essere applicata purché le frequenze teoriche, in pratica i contenuti delle tabelle, siano superiori a 5).

Il valore h indica i gradi di libertà e viene determinato a seconda del tipo di applicazione del test χ^2. Nei casi come quello in esame h è il numero n delle frequenze osservate meno uno:

$$h = n - 1$$

Nella Tavola, per $h = 5$ si legge $\chi^2_{0,95} = 11,1$. Dato che il valore calcolato di 7,23 è inferiore a 11,1, al livello di sicurezza 0,95 l'ipotesi che il dado sia buono non è rifiutabile. Poiché per $h = 5$ si legge $\chi^2_{0,99} = 15,1$, l'ipotesi non è rifiutabile al livello di sicurezza 0,99 (evidentemente i valori di $\chi^2_{0,99}$ sono superiori a quelli di $\chi^2_{0,95}$; per essere più sicuri di non rifiutare l'ipotesi quando è vera, dobbiamo essere disposti ad accettare delle discordanze più ampie tra i valori osservati e quelli teorici).

Esempio. In una scuola la media delle valutazioni generali degli alunni ha avuto gli esiti riportati nella tabella che segue.
Per ogni 100 alunni si sono avuti:

ottimo	buono	sufficiente	insufficiente	gravemente insufficiente
8	20	46	16	10

Un nuovo insegnante ha dato le seguenti valutazioni a 100 alunni:

ottimo	buono	sufficiente	insufficiente	gravemente insufficiente
12	25	51	7	5

Stabilire se il nuovo insegnante adotta un metodo di valutazione che si discosta significativamente dalla media degli altri insegnanti.

L'ipotesi nulla H_0 è che i voti attribuiti dal nuovo insegnante non si discostino da quelli attribuiti dagli altri. Si ha:

$$\chi^2 = \frac{(12-8)^2}{8} + \frac{(25-20)^2}{20} + \frac{(51-46)^2}{46} + \frac{(7-16)^2}{16} + \frac{(5-10)^2}{10} =$$
$$= \frac{16}{8} + \frac{25}{20} + \frac{25}{46} + \frac{81}{16} + \frac{25}{10} \cong 11,36$$

Poiché i gradi di libertà sono 5 – 1 = 4 e $\chi^2 > \chi^2_{0,95}$ (11,36 > 9,49), si può rifiutare l'ipotesi nulla al livello di sicurezza 0,95, ossia il nuovo professore è di manica più larga; tuttavia, poiché $\chi^2 < \chi^2_{0,99}$, al livello di sicurezza 0,99 non si può rifiutare l'ipotesi H_0 che egli si comporti come gli altri e che le discordanze osservate siano puramente casuali.

(B) Tavole di contingenza

Supponiamo di voler studiare l'efficacia di un nuovo farmaco sui sofferenti di una certa malattia. A tal fine consideriamo 200 malati: a 100 di essi (1° gruppo) somministriamo il farmaco, agli altri 100 (2° gruppo) non diamo nessuna cura. Si osserva quanto segue:

	guariti	non guariti	totale
1° gruppo	80	20	100
2° gruppo	70	30	100
totale	150	50	200

Facciamo l'ipotesi H_0 che il farmaco non abbia alcun effetto. In tal caso i 150 guariti e i 50 non guariti sarebbero equamente ripartiti nei due gruppi e le frequenze sarebbero:

	guariti	non guariti	totale
1° gruppo	75	25	100
2° gruppo	75	25	100
totale	150	50	200

Si ha:

$$\chi^2 = \frac{(80-75)^2}{75} + \frac{(20-25)^2}{25} + \frac{(70-75)^2}{75} + \frac{(30-25)^2}{25} \cong 2{,}67$$

Con tabelle di questo tipo il numero h dei gradi di libertà si calcola moltiplicando il numero di righe meno uno per il numero delle colonne meno uno. In questo caso si ha: $h = (2-1)\cdot(2-1) = 1$. Per $h = 1$, $\chi^2_{0,95} = 3{,}84 > 2{,}67$. Quindi al livello di sicurezza 0,95 non si può rifiutare l'ipotesi che il farmaco non abbia alcun effetto. Poiché, in generale, $\chi^2_{0,99} > \chi^2_{0,95}$, l'ipotesi che il farmaco non abbia alcun effetto non si può ovviamente rifiutare al livello di sicurezza 0,99.

(C) Bontà dell'adattamento

Il test χ^2 può essere adoperato per valutare la bontà dell'adattamento di una distribuzione teorica ad una empirica. Molte volte si è detto, ad esempio, che le altezze delle persone, o i loro pesi, o le durate delle lampadine, o le dimensioni dei pezzi prodotti da una macchina hanno una distribuzione normale. Come si accerta un fatto del genere?

Consideriamo la seguente tabella di altezze di 100 ragazzi:

classi di altezza (in cm)	numero dei ragazzi
meno di 140	8
140-150	10
150-160	17
160-170	34
170-180	19
oltre 180	12

Assumendo 135 cm e 185 cm come altezze delle persone delle classi estreme, si trova che la media aritmetica delle altezze è $\overline{X}$ = 163,20 cm e lo scarto quadratico medio è s = 13,88.
Consideriamo le frequenze teoriche che si avrebbero se la distribuzione fosse normale con valore medio m = 163,20 e deviazione standard σ = 13,88.
Ricorriamo alla Tavola dei valori della distribuzione normale alla fine del Capitolo quinto per calcolare le frequenze teoriche per 100 ragazzi.

(1) $$p(X \le 140) = p\left(X^* \le \frac{140-163{,}2}{13{,}88}\right) = p(X^* \le -1{,}67) = 0{,}5 - N(1{,}67) \cong 0{,}05$$

(2) $$p(140 \le X \le 15) = p\left(\frac{140-163{,}2}{13{,}88} \le X^* \le \frac{150-163{,}2}{13{,}88}\right) = p(-1{,}67 \le X^* \le -0{,}95) =$$
$$N(1{,}67) - N(0{,}95) \cong 0{,}12$$

Analogamente si trova:

(3) $p(150 \le X \le 160) \cong 0{,}24$
(4) $p(160 \le X \le 170) \cong 0{,}28$
(5) $p(170 \le X \le 180) \cong 0{,}2$
(6) $p(X \ge 180) \cong 0{,}11$

Il numero h dei gradi di libertà è 6 – 1 – 2 = 3 (occorre togliere 2 in quanto nel calcolo delle frequenze teoriche si sono sfruttate due quantità incognite m e σ stimate con $\overline{X}$ e s).

Per h = 3, $\chi^2 < \chi^2_{0,95}$ = 7,81. Quindi non si può rifiutare, al livello di sicurezza 0,95, l'ipotesi che le altezze abbiano una distribuzione normale: il test conferma l'ipotesi.

TAVOLA PER IL TEST χ^2

Valori di $\chi^2_{0,95}$

h	$\chi^2_{0,95}$	h	$\chi^2_{0,95}$	h	$\chi^2_{0,95}$	h	$\chi^2_{0,95}$
1	3,84	7	14,1	13	22,4	19	30,1
2	5,99	8	15,5	14	23,7	20	31,4
3	7,81	9	16,9	15	25,0	21	32,7
4	9,49	10	18,3	16	26,3	22	33,9
5	11,1	11	19,7	17	27,6	23	35,2
6	12,6	12	21,0	18	28,9	24	36,4

Valori di $\chi^2_{0,99}$

h	$\chi^2_{0,99}$	h	$\chi^2_{0,99}$	h	$\chi^2_{0,99}$	h	$\chi^2_{0,99}$
1	6,63	7	18,5	13	27,7	19	36,2
2	9,21	8	20,1	14	29,1	20	37,6
3	11,3	9	21,7	15	30,6	21	38,9
4	13,3	10	23,2	16	32,0	22	40,3
5	15,1	11	24,7	17	33,4	23	41,6
6	16,8	12	26,2	18	34,8	24	43,0