

6 Applicazioni

6.1 Un problema diretto: la produzione di barre

Una macchina produce barre di acciaio a sezione circolare la cui lunghezza ottimale dovrebbe essere di 5 metri ed il diametro della sezione di 4 centimetri. Le barre effettivamente prodotte, che si suppongono tra loro indipendenti, hanno una lunghezza aleatoria con distribuzione normale di media $\mu_1 = 5$ m e scarto standard $\sigma_1 = 4$ cm. Il diametro della sezione è una variabile aleatoria, indipendente dalla precedente, e con distribuzione normale di media $\mu_2 = 4$ cm e scarto standard $\sigma_2 = 0,8$ cm. Una generica barra prodotta può essere direttamente venduta senza modifiche se la sua lunghezza è compresa tra 4,95 m e 5,05 m e la sua sezione tra 2,8 cm e 5,2 cm.

- a) Qual è la probabilità di poter vendere senza modifiche una generica barra prodotta?^a
- b) Su 55.000 barre prodotte, quante ci si aspetta di poterne vendere senza modifiche?

^aLa questione è tratta dal problema n.3 della sessione ordinaria PNI, Esame di Stato 1998

- a) Sia L la v.a. normale "lunghezza della barra" e sia D la v.a. normale "sezione della barra". La barra può essere venduta senza modifiche se (e solo se) vale:

$$4,95 \text{ cm} \leq L \leq 5,05 \text{ cm} \quad \text{e} \quad 2,8 \text{ cm} \leq D \leq 5,2 \text{ cm}$$

Il quesito richiede di calcolare la probabilità che si verifichino entrambe le condizioni. Valutiamo innanzitutto la probabilità

$$\mathbf{P(4,95 \leq L \leq 5,05)}$$

1. Standardizziamo la v.a. L

$$\begin{aligned} P(4,95 \leq L \leq 5,05) &= P\left(\frac{4,95 - 5}{0,04} \leq Z \leq \frac{5,05 - 5}{0,04}\right) \\ &= P(-1,25 \leq Z \leq 1,25) \end{aligned}$$

2. Per la simmetria del grafico della densità normale si ha

$$\begin{aligned} P(-1,25 \leq Z \leq 1,25) &= 2P(0 \leq Z \leq 1,25) \\ &= 2(P(Z \leq 1,25) - P(Z \leq 0)) \end{aligned}$$

3. E, ricorrendo alle tavole, otteniamo

$$2(P(Z \leq 1,25) - P(Z \leq 0)) \simeq 2(0,894 - 0,5) = 0,788$$

Poi, in modo esattamente analogo, otteniamo

$$\mathbf{P(2,8 \leq D \leq 5,2) \simeq 0,866}$$

Pertanto la probabilità di poter mettere in vendita la barra senza modifiche è²⁷

$$\begin{aligned} P(4,95 \leq L \leq 5,05 \text{ e } 2,8 \leq D \leq 5,2) &= P(4,95 \leq L \leq 5,05) \cdot P(2,8 \leq D \leq 5,2) \\ &\simeq 0,788 \cdot 0,866 \simeq \boxed{0,68} \end{aligned}$$

b) Su 50.000 barre ci si aspetta di poterne vendere senza modifiche, circa

$$55.000 \cdot 0,68 = \boxed{37.400}$$

Infatti possiamo interpretare il valore di probabilità ottenuto nel punto a) come un valore a cui si approssima la frequenza relativa dell'evento in esame, su un "grande" numero di prove, con probabilità "grande" (forma intuitiva della Legge dei grandi numeri).

Osservazione. In alternativa, possiamo pensare di modellizzare la situazione mediante la distribuzione binomiale, caratterizzata da $n = 55.000$ prove, ciascuna con probabilità $p = 0,68$ di successo. Il valore atteso della corrispondente variabile aleatoria è proprio np e, per l'interpretazione statistica del valore atteso, np è "la" approssimazione richiesta.

²⁷Assumiamo che la v.a. L e la v.a. D siano indipendenti. In tale ipotesi la legge della moltiplicazione afferma che l'intersezione dei due eventi in questione è il prodotto delle loro singole probabilità di realizzarsi.

6.2 Un problema inverso: il periodo di funzionamento di un prodotto

La vita media di un tipo di dispositivo prodotto da una certa ditta ha distribuzione normale con media 61 mesi e deviazione standard 6 mesi. La ditta intende vendere il dispositivo con una garanzia: se smette di funzionare entro T mesi, verrà sostituito con uno nuovo.

Se la ditta vuole contenere entro il 2% la probabilità di sostituzione del prodotto, quale periodo di garanzia T può proporre?

Indichiamo con X la variabile aleatoria che rappresenta la vita media del dispositivo. Per ipotesi essa ha distribuzione normale con parametri $\mu = 61$ mesi e $\sigma = 6$ mesi.

1. Operiamo prima sulla v.a. normale standard Z , in modo da poter ricorrere alle tavole della distribuzione di probabilità. In termini della v.a. Z , la richiesta del quesito si traduce nel chiedere *per quale valore $k < 0$ di Z si ha*

$$P(Z \leq k) = 0,02$$

Però le tavole forniscono valori approssimati di probabilità del tipo $P(Z \leq k)$ solo per $k > 0$. Pertanto:

- cercheremo sulle tavole una approssimazione del valore k^+ tale che

$$P(Z \leq k^+) = 0,98$$

- il valore $k < 0$ richiesto sarà

$$k = -k^+$$

Ora, per $k^+ = 2,05$ le tavole forniscono il valore di probabilità 0,97982. Dunque $k = -2,05$ e possiamo dire che

$$P(Z \leq -2,05) \simeq 0,02$$

2. Per concludere basta esprimere il risultato ottenuto in termini della v.a. X : al valore $z = -2,05$ corrisponde il valore

$$x = \mu + \sigma z = 61 - 6 \cdot 2,05 = 48,7 \text{ mesi}$$

Ciò comporta che

$$P(X \leq 48,7) \simeq 0,02$$

Possiamo così concludere che il dispositivo smetterà di funzionare entro il tempo

$$T = 48 \text{ mesi}$$

con una probabilità minore del 2%. Perciò sono 48 i mesi di garanzia che la ditta può proporre stando nei limiti desiderati.

6.3 Altezze di una popolazione: aspetti quantitativi

Dall'elaborazione dei dati raccolti alla visita di leva dei nati nell'anno 1970, la statura risultò distribuita normalmente con media $\mu \simeq 173,94$ cm e deviazione standard $\sigma \simeq 8,63$ cm. Fino al 2014, per risultare abili all'arruolamento, era necessario avere una statura di almeno 165 cm.

Scelto a caso un giovane tra quelli appartenenti a tale classe di leva, qual è la probabilità che la sua statura fosse di almeno 165 cm?

Soluzione: la probabilità è circa 0,85

6.4 La curva normale: l'ordine nel caos -lettura-

Leggi il testo²⁸ di Mlodinov da pag.179 a pag.184 e schematizza gli aspetti matematici relativi alla curva normale che il testo mette in evidenza, soffermandoti sugli elementi nuovi che emergono. Ne riportiamo di seguito solo alcuni passi significativi²⁹.

La distribuzione normale descrive il modo in cui molti fenomeni variano intorno a un valore centrale che ne rappresenta l'esito più probabile; nel suo *Essai philosophique sur les probabilités*, Laplace sosteneva che questa nuova matematica poteva essere usata per valutare l'attendibilità di testimonianze giuridiche, per prevedere i tassi di matrimonio e calcolare i premi assicurativi. Ma nell'ultima edizione di quell'opera, Laplace era già ultrasessantenne, quindi toccò a un uomo più

Studi statistici di questo tipo erano già stati condotti in precedenza, ma Quételet fece qualcosa di più con quei dati: non si limitò a calcolare le medie, ma analizzò il modo in cui i dati si discostavano dalla media. Ovunque guardasse, Quételet trovava la distribuzione normale: nella propensione al crimine, al matrimonio e al suicidio, nell'altezza degli indiani americani, nella circonferenza toracica dei soldati scozzesi (si imbatté in un campione di 5738 circonferenze toraciche in una vecchia copia dell'«Edinburgh Medical and Surgical Journal»). Nell'altezza di 100.000 giovani francesi chiamati alla leva trovò significativa una deviazione dalla distribuzione normale: se si tracciava un grafico per mettere in relazione il numero dei coscritti e la loro altezza, la curva a campana risultava distorta: ce n'erano troppo pochi appena sopra il metro e 56, e troppi appena al di sotto di quell'altezza. Quételet ipotizzò che la differenza – circa 2000 «uomini bassi» in più – fosse dovuta a frode, o magari a qualche favore concesso, dato che gli uomini sotto il metro e 56 venivano riformati.

²⁸La lettura che proponiamo è tratta da *La passeggiata dell'ubriaco* di L. Mlodinov [34, pag.179-184].

²⁹Ciò è dovuto a ragioni di spazio, ma è opportuno che il testo indicato venga letto per intero.

esempio meno noto è quello di Justin Wolfers, economista della Wharton School, che ha trovato indizi di frode sportiva nei risultati di circa 70.000 partite di basket del college.¹⁸

I point spread sono fissati dai bookmaker, ma in realtà sono determinati dalla massa degli scommettitori, perché i bookmaker li calcolano in modo da bilanciare la domanda. (I bookmaker guadagnano sulle commissioni, e cercano di far scommettere la stessa cifra su entrambe le squadre in modo da trarne un profitto comunque vada a finire l'incontro). Per misurare con che efficacia gli scommettitori valutano due squadre, gli economisti usano un numero chiamato «errore di previsione», che è la differenza tra il margine di vittoria della squadra favorita e il point spread determinato dal mercato. Forse non sorprenderà che l'errore di previsione, essendo una tipologia di errore, segua la distribuzione normale. Wolfers ha calcolato che la media è 0, il che significa che i point spread non tendono né a sovrastimare né a sottostimare le squadre, e la deviazione standard è 10,9 punti, il che vuol dire che circa 2/3 delle volte il point spread si trova entro 10,9 punti dal margine di vittoria. (In uno studio sugli incontri professionali di football si è scoperto un risultato simile, con una media di 0 e una deviazione standard di 13,9 punti.)¹⁹

Quando Wolfers ha esaminato il sottoinsieme di giochi che riguardavano squadre molto favorite, ha scoperto qualcosa di sorprendente: erano troppo poche le partite in cui la super-favorita vinceva con un margine poco superiore al point spread, ed erano davvero troppe le partite in cui la favorita vinceva di poco meno dello spread. Tornava in scena l'anomalia di Quételet. Anche Wolfers, come Quételet e Poincaré, è giunto alla conclusione che doveva trattarsi di frode. La sua