

TAB. 4.3. *Calcolo delle frequenze teoriche normali in intervalli di classe di una distribuzione con $n = 138$, media = 552,11 e $s = 79,32$*

Classi	f	Limiti reali di classe	Valori z dei limiti di classe	Proporzioni di area tra media e limiti reali	Proporzioni di area tra media e i limiti reali	Frequenze teoriche normali di classe $\times 138$
		$+\infty$	$+\infty$	0,5000		
750-779	1	749,5	2,49	0,4936	0,0064	0,88
720-749	2	719,5	2,11	0,4826	0,0110	1,52
690-719	3	689,5	1,73	0,4582	0,0244	3,37
660-689	8	659,5	1,35	0,4115	0,0467	6,44
630-659	12	629,5	0,98	0,3365	0,0750	10,35
600-629	13	599,5	0,60	0,2257	0,1108	15,29
570-599	17	569,5	0,22	0,0871	0,1386	19,13
540-569	18	539,5	-0,16	0,0636	0,1507	20,80
510-539	18	509,5	-0,54	0,2054	0,1418	19,57
480-509	17	479,5	-0,92	0,3212	0,1158	15,98
450-479	16	449,5	-1,29	0,4015	0,0803	11,08
420-449	10	419,5	-1,67	0,4525	0,0510	7,04
390-419	2	389,5	-2,05	0,4798	0,0273	3,77
360-389	1	$-\infty$	$-\infty$	0,5000	0,0202	2,79
Σ 138					1,000	138,01

classe superiore da 749,5 a $+\infty$ nella scala dei punteggi. I *valori z* corrispondenti ai limiti reali di ogni classe, riportati nella quarta colonna della tabella 4.3, vengono calcolati sottraendo il valore di ogni limite reale dalla media e dividendo il risultato per la deviazione standard. Così il valore z di 599,9 è $(599,9 - 552,11)/79,32 = 0,60$. Successivamente, si calcolano le proporzioni di area tra *media* e i *punti z* ; queste vengono ottenute dalla tavola delle aree della curva normale. A questo punto è possibile calcolare le proporzioni di area tra i *punti z* corrispondenti ai vari limiti reali della distribuzione; questo passaggio richiede il calcolo delle differenze tra le proporzioni precedentemente calcolate. Così il valore della sesta colonna, 0,0064, è dato dalla differenza tra i due primi valori della quinta colonna: $0,5000 - 0,4936 = 0,0064$. Le frequenze teoriche normali, ultima colonna della tabella 4.3, si ottengono moltiplicando per n (cioè per il numero dei punteggi, che nel nostro caso è 138) le proporzioni di area tra i limiti reali. Pertanto, i prodotti disposti nell'ultima colonna della tabella 4.3 sono le *frequenze attese* o *teoriche normali* distribuite in classi, secondo una distribuzione normale con $n = 138$, *media* = 552,11 e $s = 79,32$. Se il lettore traccia una curva normale e indica i *valori z* dei limiti reali di classe, troverà facile seguire il tipo di calcolo indicato nella tabella 4.3.

6.2. Adattamento di una curva normale a una distribuzione di frequenza

Una distribuzione di frequenza, i cui punteggi siano di fatto distribuiti normalmente, può venire adattata ad una curva normale. Il procedimento richiede, in primo luogo, il calcolo delle frequenze delle classi, come è illustrato nella tabella 4.3. Qui di seguito, però, diamo alcune spiegazioni che possono aiutare a capire meglio tale adattamento.

La tabella 4.3 è sistemata in modo tale da facilitare il calcolo delle proporzioni di area o le frequenze relative comprese tra i limiti reali delle rispettive classi. Una volta determinate tali aree, si moltiplicano per 138 (il numero dei casi del nostro esempio, quindi numero che varia di volta in volta) allo scopo di ottenere le frequenze teoriche assolute normali. L'ampiezza dell'intervallo esatto della classe inferiore parte da $-\infty$ e arriva a 389,5; mentre quella dell'intervallo esatto della

TAB. 9.11. *Frequenze teoriche e frequenze osservate di un campione di frequenze di una distribuzione normale*

f_o	F_t	$f_o - F_t$	$\frac{(f_o - F_t)^2}{F_t}$
6 $\begin{cases} 1 \\ 2 \\ 3 \end{cases}$	5,77 $\begin{cases} 0,88 \\ 1,52 \\ 3,37 \end{cases}$	0,23	0,01
8	6,44	1,56	0,38
12	10,35	1,65	0,26
13	15,29	-2,29	0,34
17	19,13	-2,13	0,24
18	20,80	-2,80	0,38
18	19,57	-1,57	0,13
17	15,98	1,02	0,07
16	11,08	4,92	2,18
10	7,04	2,96	1,24
3 $\begin{cases} 2 \\ 1 \end{cases}$	6,56 $\begin{cases} 3,77 \\ 2,79 \end{cases}$	-3,56	1,93
Σ 138	138,01	-0,01	7,16

2.6. Controllo della forma della popolazione di una distribuzione di frequenze

Il chi-quadrato è molto utile anche per verificare l'ipotesi se un campione di osservazioni sia stato tratto da una popolazione con distribuzione normale. La terminologia inglese a questo proposito usa l'espressione: «goodness-of-fit».

Per illustrare questo procedimento ci riferiamo alla distribuzione delle frequenze della tabella 4.3, con lo scopo di verificare se l'accordo tra frequenze teoriche e frequenze osservate sia sufficientemente alto da permettere di sostenere l'ipotesi nulla che la popolazione da cui proviene il campione è distribuita in modo normale. Per comodità riportiamo le frequenze osservate (f_o) e le frequenze teoriche (F_t) nella tabella 9.11.

Le caratteristiche della distribuzione di frequenza della tabella 9.11 sono: $n = 138$, $\bar{X} = 552,11$ e $s = 79,32$. L'ipotesi nulla è l'assunzione della normalità della distribuzione della popolazione, per cui frequenze teoriche e frequenze osservate dovrebbero corrispondere.

Prima di calcolare il χ^2 , è opportuno raggruppare le prime tre classi della distribuzione e le ultime due. Questa operazione va sempre fatta quando le frequenze osservate sono molto piccole (ad esempio inferiori a 5), perché le oscillazioni casuali di tali frequenze determinano bruschi sbalzi nel valore del chi-quadrato.

Dopo aver raggruppato tali frequenze, il numero delle classi si restringe e diventa 11, ma solo 8 ($11 - 3$) di esse sono indipendenti. Per determinare le frequenze attese, si devono pareggiare le frequenze teoriche e le frequenze attese relativamente a tre costanti: n , $\bar{X}$ ed s . Il procedimento risulta, quindi, espresso da tre restrizioni:

$$[9.13] \quad \Sigma F_t = \Sigma f_o; \quad \Sigma F_t X = \Sigma f_o X; \quad \Sigma F_t X^2 = \Sigma f_o X^2$$

Il valore di χ^2 è 7,16, con $V = 11 - 3 = 8$.

Osservando la tavola del chi-quadrato si trova che il valore calcolato χ^2_{cal} di 7,16 ha nella riga $\nu = 8$ una probabilità P di circa 0,50. Nel campionamento da una popolazione normale, sotto determinate condizioni, una discrepanza come quella osservata tra f_o ed F_t può accadere circa il 50% delle volte solo per causa delle fluttuazioni di campionamento. Non vi è quindi motivo di dubitare sulla assunzione della normalità della popolazione da cui è stato tratto il campione.

È importante sottolineare che, mediante il test del chi-quadrato, l'assunzione della normalità non è provata; si afferma solo che può ritenersi tale.

Il limite maggiore che presenta il χ^2 nella funzione di verificare la normalità di una popolazione, partendo dalle frequenze di un campione, è dovuto alla sua inadeguatezza a tener conto dei segni algebrici degli scarti. Una analisi di queste differenze nella tabella 9.11 fa emergere che le frequenze osservate sono minori delle frequenze teoriche nelle classi centrali della distribuzione e maggiori nelle altre. Questo dato suggerisce che la forma della popolazione può qualche volta essere più schiacciata della normale, ma il test del chi-quadrato non è adatto a rilevarlo.

V=gradi di libertà

Poichè $p=0,50>0,05$

si rifiuta l'ipotesi
alternativa e si

accetta l'ipotesi nulla
ossia non esistono
ragioni per escludere
che nella
popolazione.....