

Alcuni contestuali aspetti storico-statistici della gara di maratona maschile e femminile

Dinamiche e caratterizzazioni analitico-comparative delle due liste mondiali "all-time"

Otello Donzelli

Collaboratore del Centro Studi e Ricerche FIDAL per il Settore di Statistica Applicata

1) INTRODUZIONE

Recentemente è stata trattata in questa sede un'analisi di tipo inferenziale su alcuni aspetti stocastici della lista mondiale "all-time" della gara maschile di maratona. Come a suo tempo si è potuto constatare, diversi interessanti aspetti sono emersi a seguito dell'indagine condotta sui dati storici della lista in questione. Si ricorderà anche che l'analisi era stata incentrata sull'esame di tre primarie caratteristiche statistiche del fenomeno, e cioè media, deviazione-standard e varianza delle età degli atleti che ne facevano parte, con e senza l'applicazione del criterio della loro suddivisione per grandi aree geografiche di appartenenza.

Poiché la gara di maratona in questi ultimi anni ha polarizzato crescenti interessi anche nell'ambito internazionale del settore femminile, è stata senz'altro ritenuta produttiva la programmazione di una nuova analisi della gara che, seppure con modalità investigative differenti adottate allo scopo di evitare situazioni analitiche poco attraenti perché ripetitive, comprendesse comunque punti di contatto con la precedente per poter poi avere la possibilità di operare utili comparazioni di certi specifici parametri distintivi dei due settori.

In effetti, questa indagine statistica può considerarsi, sotto certi aspetti, un logico e naturale proseguimento dell'altra anche perché, con l'occasione, si presenta il destro di poter allargare, contestualmente nei due settori di attività agonistica, il campo di azione dello studio del fenomeno sportivo in ordine alle molteplici alternative che la metodologia di ricerca è in grado di offrire proprio nelle due fattispecie prese in esame.

L'indagine comparativa è completata da una particolare verifica di significatività dei coefficienti di correlazione

lineare calcolati per il settore maschile e femminile. Infine, non sarà inutile sottolineare che le tecniche statistiche adottate nell'indagine rispondono ai più classici canoni richiesti per una loro ottimale applicazione. È stato fatto riferimento, in proposito, anche ad interpretazioni di numerosi assimilabili fenomeni, caratteristici di altri campi delle scienze sperimentali, che risultano frequentemente proposti ai più qualificati livelli bibliografici della materia.

2) La funzione log-normale rappresentativa dei due fenomeni sportivi

2.1 Elementi descrittivi del metodo dei momenti dall'origine adottato per la definizione dei parametri della funzione

In considerazione del fatto che la rappresentazione grafica delle distribuzioni di frequenza di entrambi i fenomeni sportivi presi in esame è palesemente attribuibile ad un medesimo tipo di curva campanulare, precisamente con coda destra più allungata della sinistra, si ritiene molto confacente alla natura dei fenomeni stessi assumere la *funzione log-normale* quale loro interprete analitica. La caratteristica appena descritta della curva funzionale, pertanto, comporta per entrambe le due fattispecie la definizione di "*funzione log-normale con asimmetria positiva*". La scrittura di detta funzione (il tipo di asimmetria interferisce naturalmente sui relativi parametri per ora ignoti) è la seguente:

$$f(x) = \frac{N \delta}{\sqrt{2\pi} (x - \theta)} \exp \left\{ -1/2 [\lambda + \delta \log_e (x - \theta)]^2 \right\}$$

La curva è stata rappresentata graficamente per punti nella Tav. n. 1 e nella Tav. n. 2. Nella prima è stato adottato per il fenomeno reale il criterio della sua rappresentazione mediante *istogramma* (con classi di età pari ad 1 anno), mentre nella seconda è stato seguito il criterio di rappresentazione mediante *spezzata*, dilatando deliberatamente il passo grafico della variabile indipendente per meglio evidenziare la lieve asimmetria positiva del fenomeno a livello maschile (quella a livello femminile è già piuttosto evidente con la scala adottata nella Tav. n. 1).

Intanto è necessario premettere che la funzione log-normale gode di ragguardevoli proprietà formali che giustificano abbondantemente il suo impiego a scopi descrittivi nei vari campi di investigazione empirica. In effetti, tale funzione trova congruo riscontro in molteplici fenomeni sociali e naturali per cui il suo impiego è pressoché irrinunciabile in talune ricerche, specialmente di natura

scientifico (1). Comunque, nell'analisi statistico-sportiva esistono numerosi esempi di una sua pratica ed efficiente applicazione. Anche nello specifico settore della medicina sportiva, e questo oggi non rappresenta davvero un fatto sporadico, la sua applicazione trova sicuramente impieghi molto produttivi in tutti quei fenomeni bio-medici di ricerca e controllo che rappresentino particolari andamenti crescenti e decrescenti a tasso variabile. Quanto appena detto a significare che le modificazioni, del resto solo formali, che sorgono nell'applicazione di tale funzione rispetto ai canoni analitici della fin troppo nota funzione di Gauss, permettono in sostanza che di questa sia praticamente considerata come una variante interpretativa (leggi: capacità di ponderazione analitica del grado di asimmetria della sua curva rappresentativa).

Per l'analisi dei due fenomeni sportivi presi in considerazione è stato ritenuto confacente al caso l'uso del metodo dei *momenti*. Secondo la definizione del *Vianelli* "si dice momento di origine μ e di grado r della frequenza assoluta y_i il prodotto della frequenza per l' r -esima potenza dello scarto di x_i da μ , ossia: $(x_i - \mu)^r y_i$. Ciò premesso, chiamasi momento di origine μ e di grado r - o di origine r - della distribuzione statistica, la somma delle r -esime potenze degli scarti $(x_i - \mu)$, ponderate con pesi dati dalle corrispondenti frequenze".

Quanto appena enunciato può scriversi:

$$m_{\mu, r} = \frac{\sum_{i=1}^n (x_i - \mu)^r y_i}{\sum_{i=1}^n y_i}$$

dove μ è un valore qualsiasi ed r è l'ordine del momento. La formula in questione per $\mu = 0$ diviene quindi l'espressione dei momenti dall'origine, cioè:

$$m_r = \frac{\sum_{i=1}^n x_i^r y_i}{\sum_{i=1}^n y_i}$$

A questo punto si può procedere per il calcolo dei parametri $N, \theta, \delta, \lambda$ della funzione log-normale. Per tale calcolo si stabilisce quindi l'uguaglianza tra momenti empirici e momenti teorici. Considerando i primi due momenti dall'origine e i momenti centrali secondo e terzo si instaura il sistema:

(1) La funzione log-normale è conosciuta anche come funzione di McAlister, dal nome del matematico che ne ha scoperto e codificato le sue caratteristiche peculiari. Con l'occasione, non sarà inutile ricordare che una variabile si distribuisce in modo log-normale quando il logaritmo della variabile è distribuito normalmente, ovvero, in altri termini, una variabile casuale si distribuisce in modo log-normale quando ad essa corrisponde la funzione di densità.

$$\left\{ \begin{aligned} \frac{1}{N} \sum_{i=1}^n y_i &= \frac{1}{N} \int_{\theta}^{\infty} \frac{N \delta}{\sqrt{2\pi} (x - \theta)} \exp \{-1/2 [\lambda + \delta \log_e (x - \theta)]^2\} dx \\ \frac{1}{N} \sum_{i=1}^n x_i y_i &= \frac{1}{N} \int_{\theta}^{\infty} \frac{N \delta}{\sqrt{2\pi} (x - \theta)} \exp \{-1/2 [\lambda + \delta \log_e (x - \theta)]^2\} dx \\ \frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2 y_i &= \frac{1}{N} \int_{\theta}^{\infty} \frac{N \delta}{\sqrt{2\pi} (x - \theta)} \exp \{-1/2 [\lambda + \delta \log_e (x - \theta)]^2\} dx \\ \frac{1}{N} \sum_{i=1}^n (x_i - \mu)^3 y_i &= \frac{1}{N} \int_{\theta}^{\infty} \frac{N \delta}{\sqrt{2\pi} (x - \theta)} \exp \{-1/2 [\lambda + \delta \log_e (x - \theta)]^2\} dx \end{aligned} \right.$$

2.2 Calcolo dei parametri della funzione applicata al settore maschile

Per il calcolo dei già accennati momenti empirici si procede secondo lo sviluppo della Tabella n. 1 la quale, tra l'altro, comprende i dati necessari per le successive elaborazioni.

A pag. n. 81 del volume edito a cura della I.A.A.F. "Marathon Handbook 1992", autore il dott. Ottavio Castellini, si legge che le 1292 performances della lista mondiale della gara sono relative a n. 492 performers dai quali, però, occorre eliminare n. 3 performers di cui non è stata purtroppo possibile conoscere l'età di nascita. Per i restanti 489 atleti la distribuzione delle loro età al momento del conseguimento delle performances nella lista "all-time" in questione compare nella colonna A della predetta Tav. n. 1. L'età è considerata, per motivi pratici di calcolo, come poi si avrà modo di constatare, in funzione del valore centrale della classe di età con ampiezza pari ad 1 anno per cui, ad esempio, il valore 20 anni è relativo al valore centrale della classe di età compresa tra 19 anni e 1/2 e 20 anni e 1/2. Intanto questa soluzione permette la rappresentazione del fenomeno mediante il classico istogramma contestualmente alla parallela rappresentazione grafica me-

Tabella N. 1

DATI STATISTICI DEL SETTORE MASCHILE DELLA GARA DI MARATONA PER LA DEFINIZIONE DEI PARAMETRI DELLA FUNZIONE LOG-NORMALE, DEGLI INDICI DI ASIMMETRIA E DI ACCOSTAMENTO.

A	B	C	D	E	F	G	H	I	L	M	N
20	3	9	-11	-33	363	-3993	43923	0,083954	1	2	4
21	0	0	-10	0	0	0	0	2,79690	3	-3	9
22	7	49	-9	-63	567	-5103	45927	7,19917	7	0	0
23	13	169	-8	-104	832	-6656	53248	14,91679	15	-2	4
24	34	1166	-7	-238	1666	-11662	81634	25,71512	26	8	64
25	37	1369	-6	-222	13332	-7992	47952	37,88673	38	-1	1
26	44	1936	-5	-220	1100	-5500	27500	48,77096	49	-5	25
27	46	2116	-4	-184	736	-2944	11776	55,86870	56	-10	100
28	70	4900	-3	-210	630	-1890	5670	57,83169	57	13	169
29	47	2209	-2	-94	188	-376	752	54,79797	55	-8	64
30	50	2500	-1	-50	50	-50	50	48,05360	48	2	4
31	48	2304	0	0	0	0	0	39,36550	39	9	81
32	28	784	1	28	28	28	28	30,36865	30	-2	4
33	24	576	2	48	96	192	384	22,21614	22	2	4
34	14	196	3	42	126	378	1134	15,50454	16	-2	4
35	5	25	4	20	80	320	1280	10,37704	10	-5	25
36	6	36	5	30	150	750	3750	6,69120	7	-1	1
37	8	64	6	48	288	1728	10368	4,17349	4	4	16
38	2	4	7	14	98	686	4802	2,52698	3	-1	1
39	0	0	8	0	0	0	0	1,48995	2	-2	4
40	1	1	9	9	81	729	6561	0,85787	1	0	0
41	1	1	10	10	100	1000	10000	0,48353	0	1	1
42	1	1	11	11	121	1331	14641	0,26737	0	1	1
	489	20405		-1158	8632	-39024	371380	488,99948	489	84	582
								arrot. 0,00052			
								489,00000			

- A) x'_i = variabile indipendente = valore centrale della classe $x_i - x_{i+1}$
- B) y_i = variabile dipendente = frequenza
- C) y_i^2 = quadrato della variabile dipendente
- D) x''_i = variabile indipendente dall'origine arbitraria di: 31 anni di età dell'atleta = 0
- E) $x''_i y$ = prodotto D B
- F) $(x''_i)^2 y_i$ = prodotto D² B
- G) $(x''_i)^3 y_i$ = prodotto D³ B
- H) $(x''_i)^4 y_i$ = prodotto D⁴ B
- I) y^*_i = valore teorico di calcolo della variabile dipendente della funzione log-normale
- L) y^{**}_i = valore teorico unitario di arrotondamento di y^*_i
- M) $y_i - y^{**}_i$ = differenza fra frequenza effettiva e teorica unitaria
- N) $(y_i - y^{**}_i)^2$ = quadrato della differenza fra frequenza effettiva e teorica unitaria

Nota: il campo di variazione della classe di età è pari ad 1 anno per cui i valori iniziali e finali di classe corrispondono rispettivamente a 1/2 anno in meno e 1/2 anno in più di x'_i

dianete "spezzata" riferita ad anni di età interi; in secondo luogo, il valore 1 della classe permetterà, nell'applicazione delle varie formule di calcolo, una pratica loro semplificazione.

Facendo riferimento ai dati della Tav. n. 1, la media, per la formula:

$$\mu = \bar{x} = a + b \frac{\sum_{i=1}^n x_i'' y_i}{N}, \text{ nella quale:}$$

a = valore centrale della classe di origine 0
b = ampiezza della classe di età = 1, si avrà:

$$\mu = 31 + 1 \left(\frac{-1158}{489} \right) = 31 - 2,36810 = 28,63190.$$

Il momento centrale secondo, ovvero il valore di varianza, è dato da:

$$m'_2 = \sigma^2 = b^2 \left[\frac{\sum_{i=1}^n x_i''^2 y_i}{N} - \left(\frac{\sum_{i=1}^n x_i'' y_i}{N} \right)^2 \right] =$$

$$= 1^2 \left[\frac{8632}{489} - \left(\frac{-1158}{489} \right)^2 \right] = 17,65235 - (2,36810)^2 =$$

$$= 12,04445, \text{ per cui si avrà anche: } \sigma = \sqrt{12,04445} = 3,47051.$$

Il momento centrale terzo è dato da:

$$m'_3 = \sigma^3 = b^3 \left[\frac{\sum_{i=1}^n x_i''^3 y_i}{N} - 3 \frac{\sum_{i=1}^n x_i''^2 y_i}{N} \cdot \frac{\sum_{i=1}^n x_i'' y_i}{N} + \right.$$

$$\left. + 2 \left(\frac{\sum_{i=1}^n x_i'' y_i}{N} \right)^3 \right] = 1^3 \left[\frac{-39024}{489} - 3 \frac{8632}{489} \cdot \frac{-1158}{489} + \right.$$

$$\left. + 2 \left(\frac{-1158}{489} \right)^3 \right] = -79,80368 + 125,40759 - 26,56012 =$$

$$= 19,04379.$$

Con l'occasione, si può prendere atto che γ_1 = indice di asimmetria risulta:

$$\gamma_1 = m'_3 / \sigma^3 = 19,04379 / (3,47051)^3 = 19,04379 / 41,80035 = + 0,45559, \text{ ossia esistenza asimmetrica positiva, ciò che conferma la congruità di adozione del modello log-normale in luogo di quello normale.}$$

Per il calcolo del momento t.mo della distribuzione log-normale il procedimento è il seguente:

$$m_t(\theta)^* = \int_{\omega}^{\infty} \frac{N \delta (x - \theta)^t}{\sqrt{2\pi} (x - \theta)} \exp \{-1/2 [\lambda + \delta \log_e (x - \theta)]^2\} dx.$$

Ponendo per brevità di elaborazione:

1) $(x - \theta) = v$

2) $e^{1/2 \delta^2} = \omega$

3) $e^{-\lambda \delta} = \rho,$

l'equazione poc'anzi citata può essere così riscritta:

$$m_t(\theta)^* = \int_{\omega}^{\infty} \frac{N \delta v^t}{\sqrt{2\pi} v} \exp \{-1/2 [\lambda + \delta \log_e v]^2\} dv =$$

$$= e^{t^2/2\delta^2} \cdot \omega \delta = \omega^t \rho^t.$$

Pertanto, in conseguenza di quanto appena detto i momenti centrali secondo e terzo risultano:

$$m'_2^* = m_2^*(\theta) - [m_1^*(\theta)]^2 = \omega^4 \rho^2 - \omega^2 \rho^2 = \omega^2 \rho^2 (\omega^2 - 1);$$

$$m'_3^* = m_3^*(\theta) - 3 m_2^*(\theta) m_1^*(\theta) + 2 [m_1^*(\theta)]^3 =$$

$$= \omega^9 \rho^3 - 3 \omega^4 \rho^2 \omega \rho + 2 \omega^3 \rho^3 = \omega^3 \rho^3 (\omega^2 - 1)^2 (\omega^2 + 2).$$

Il sistema che instaura le uguaglianze tra i momenti teorici ed empirici si trasforma, quindi, in:

(momento t.mo)	}	N = 489
(momento primo dall'origine)		$\omega \rho + \theta = 28,63190$
(momento centrale secondo)		$\omega^2 \rho^2 (\omega^2 - 1) = 12,04445$
(momento centrale terzo)		$\omega^3 \rho^3 (\omega^2 - 1)^2 (\omega^2 + 2) = 19,04379$

Dividendo l'ultima equazione membro a membro per la penultima elevata a 3/2 si perviene all'equazione:

$$\sqrt{(\omega^2 - 1)} (\omega^2 + 2) = \frac{19,04379}{(12,04445)^{3/2}} = \frac{19,04379}{41,80040} = 0,45559.$$

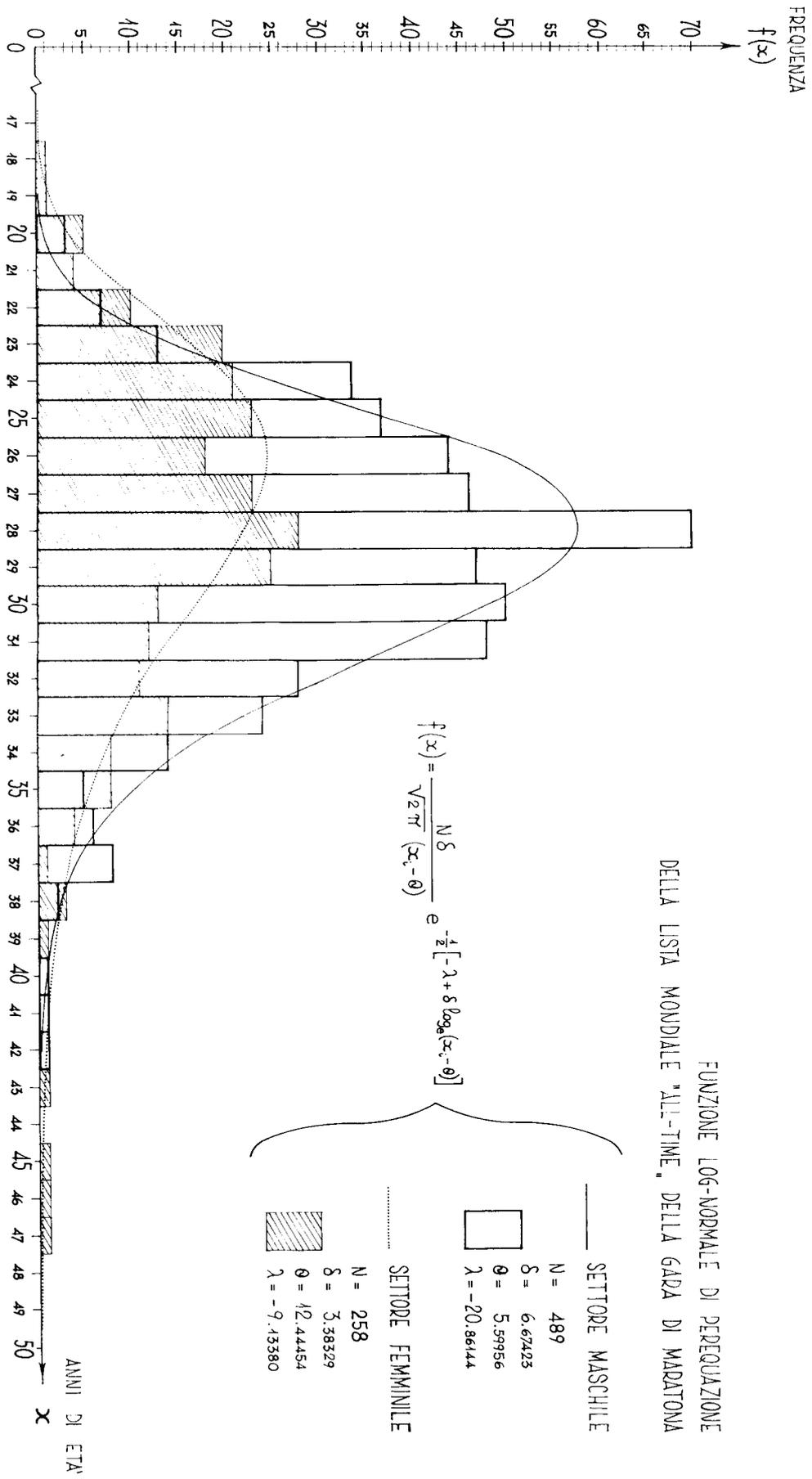


Tavola N. 1

GRAFICO ORIGINALE DELL'AUTORE

Ponendo $u^2 = \omega^2 - 1$ si giunge all'equazione cubica:
 $u^3 + 3u - 0,45559 = 0$.

Infatti, con i relativi passaggi si ha:

$$\sqrt{u^2} (\omega^2 + 2) = 0,45559,$$

$$u [(u^2 + 1) + 2] = 0,45559,$$

$u^3 + 3u - 0,45559 = 0$, la cui unica radice in soluzione reale è 0,150723, non competendo al problema le altre due complesse coniugate. Poiché è stato posto

$u^2 = \omega^2 - 1$, si ha anche: $\omega = \sqrt{u^2 + 1} = \sqrt{0,022717 + 1} = \sqrt{1,022717} = 1,01129$. E ancora, avendo posto $\omega = e^{1/2\delta}$, si ha che:

$$\delta = \frac{1}{\sqrt{2 \log_e \omega}} = \frac{1}{\sqrt{2 \log_2 1,01129}} = \frac{1}{\sqrt{0,02245}} = \frac{1}{0,14983} = 6,67423.$$

Dalla terza equazione del sistema in ordine al momento centrale secondo si ha:

$$\rho = \sqrt{\frac{12,04445}{(1,01129)^2 [(1,01129)^2 - 1]}} = \sqrt{\frac{12,04445}{0,02322}} = \sqrt{518,71016} = 22,77521.$$

Avendo in precedenza posto $\rho = e^{-\lambda\delta}$ sarà anche:

$$\lambda = -\delta \log_e \rho = -6,67423 \log_e 22,77521 = [(-6,67423)(3,12567)] = -20,86144.$$

Infine, dalla seconda equazione dell'ultimo sistema si ricava:

$$\theta = 28,63190 - \omega\rho = 28,63190 - [(1,01129)(22,77521)] = 28,63190 - 23,03234 = 5,59956.$$

Le operazioni fin qui eseguite ci permettono ora di definire l'equazione della funzione log-normale interpolatrice in grado di rappresentare in modo congruo il fenomeno sportivo preso in esame, e cioè:

$$f(x) = \frac{[(489)(6,67423)]}{\sqrt{2\pi} (x - 5,59956)} \exp \left\{ -1/2 [-20,86144 + 6,67423 \log_e (x - 5,59956)]^2 \right\}.$$

Nella Tav. n. 1 è possibile prendere atto che l'andamento grafico della curva log-normale assunta a livello analitico rispecchia con rimarchevole fedeltà quello analogo ricavato mediante l'uso di istogramma. Per la precisione, rimane da dire che le frequenze teoriche così codificate andrebbero calcolate mediante integrazione della funzione rappresentativa, poc'anzi definita, per ciascuna classe. A questo punto si può anche aggiungere che con soddisfacente approssimazione le frequenze teoriche possono essere ricavate mediante moltiplicazione delle ordinate della curva log-normale calcolate al centro di ciascuna classe (qui è opportuno ricordare il vantaggio già accennato in precedenza circa la scelta del valore di classe pari ad 1 anno e quello centrale di conseguenza commisurato in anni interi in funzione degli specifici estremi di classe) per le ampiezze delle corrispondenti classi così intese.

Una prima idea del grado di bontà della capacità di rappresentazione analitica della funzione adottata è data dalla constatazione dell'entità degli scarti tra frequenze empiriche e teoriche, lasciando in un secondo momento all'applicazione delle formule degli appropriati indici di accostamento il compito della relativa misura. In relazione all'accennato criterio di valutazione dell'aspetto teorico, la Tav. n. 1 riporta, nelle colonne M ed N, i dati che esprimono un primo e caratterizzante aspetto del fenomeno. Intanto, si può dire che, ad eccezione della marcata escursione esistente tra il valore teorico e quello empirico riferita agli atleti di 28 anni di età, gli scarti residui risultanti in sostanza dalla contrapposizione tra il valore centrale delle classi ed il valore dell'ordinata teorica corrispondente non dimostrano nel complesso discrasie tali che possano far ritenere non appropriata la scelta del tipo specifico di funzione adottato per una ottimale rappresentazione del fenomeno. A questo proposito c'è da notare che la netta differenza, peraltro unica per tutta la distribuzione, tra il valore teorico ed empirico che si verifica per gli atleti di 28 anni di età, sta a significare solo che il valore (modale) presenta carattere di intensità piuttosto accentuata rispetto a quelli relativi ad età immediatamente contigue. In altri termini pratici, per l'età modale di 28 anni esiste una molto alta probabilità di invarianza in funzione dello scorrere del tempo. In effetti, si deve tener presente che i contigui valori inferiori 25, 26, 27 e superiori 29, 30 e 31, ed a maggior ragione gli altri meno prossimi a 28, si attestano rispettivamente a 37, 44, 46, 47, 50 e 48 frequenze. A questo si deve aggiungere il fatto che la lista all-time in questione rappresenta un *excursus* storico di ben 27 anni. Per la precisione, gli atleti che aprono e chiudono con le migliori prestazioni delle due annate che comprendono il periodo in questione sono: Ibrahim Hussein (KEN) nato il 3 giugno 1958, che ha corso la maratona di Boston del 20 Aprile 1992 in 2:08:14, 15° della lista all-time e Morio Shigematsu (JPN) nato il 21 giugno 1940, che ha corso la maratona di Chiswick del 12 Giugno 1965, 713° della li-

sta. Non sarà inutile sottolineare che quanto appena detto rende chiara anche a livello intuitivo l'improbabilità di una variazione dell'età di massima frequenza, quanto meno a breve scadenza.

2.3 Calcolo dei parametri della funzione applicata al settore femminile

Ripetendo la procedura di calcolo dei parametri del settore maschile, per quello femminile (i relativi dati statistici sono contenuti nella Tabella n. 2) si avrà:

$$\mu = 32 + 1 \left(\frac{-1035}{258} \right) = 32 - 4,01628 = 27,98372,$$

tenendo presente che il già citato volume da cui sono stati tratti i dati statistici anche per il settore femminile fornisce n. 885 performances per n. 259 atlete, diventate poi 258 in quanto per un'atleta non si è potuto entrare in possesso dei relativi dati anagrafici (2).

Il momento centrale secondo, ovvero il valore di varianza, è dato da:

$$m'_2 = \sigma^2 = b^2 \frac{\sum_{i=1}^n x_i''^2 y_i}{N} - \left(\frac{\sum_{i=1}^n x_i'' y_i}{N} \right)^2 = 1^2 \frac{9839}{258} +$$

$$- \left(\frac{-1035}{258} \right)^2 = 38,13566 - 16,09316 = 22,04250,$$

per cui si avrà anche una deviazione-standard

$$\sigma = \sqrt{22,04250} = 4,69494$$

Il momento centrale terzo è dato da:

$$m'_3 = \frac{-60159}{258} - 3 \frac{9839}{258} \cdot \frac{-1035}{258} + 2 \left(\frac{-1035}{258} \right)^3 =$$

$$= -233,17442 + 458,95847 - 129,11953 = 96,66452.$$

Analogamente a quanto si è visto per il settore maschile, si può prendere atto con l'occasione che il parametro $\gamma_1 =$ indice di asimmetria risulta:

(2) Per il dettaglio della lista all-time del settore interessato è consigliabile la consultazione diretta del già descritto volume "Marathon Handbook 1992", autore il dott. Ottavio Castellini, che tra l'altro contiene numerosi altri tipi di classificazione sotto l'aspetto tecnico-storico.

$\gamma_1 = m'_3 / \sigma^3 = 96,66452 / (4,69494)^3 = 96,66452 / 103,48831 = +0,91474$, ossia esistenza di asimmetria positiva, ciò che conferma anche in questo caso, anzi con maggiore incidenza rispetto al settore maschile (praticamente in misura doppia poiché $0,91474 / 0,5559 = 2,00781$), la congruità di adozione del modello log-normale in luogo di quello normale.

Per il calcolo del momento t.mo della distribuzione log-normale il sistema che instaura le uguaglianze tra momenti teorici ed empirici è il seguente:

$$\begin{cases} \text{(momento t.mo)} & \left\{ \begin{array}{l} N = 258 \\ \omega\rho + \theta = 27,98372 \\ \omega^2 \rho^2 (\omega^2 - 1) = 22,04250 \\ \omega^3 \rho^3 (\omega^2 - 1)^2 (\omega^2 + 2) = 96,66452 \end{array} \right. \end{cases}$$

Dividendo l'ultima equazione membro a membro per la penultima elevata a 3/2 si perviene all'equazione:

$$\sqrt{(\omega^2 - 1)} (\omega^2 + 2) = \frac{96,66452}{(22,04250)^{3/2}} = \frac{96,66452}{103,48831} = 0,93406.$$

Ponendo $u^2 = \omega^2 - 1$ si perviene alla equazione cubica: $u^3 + 3u - 0,93406 = 0$.

Infatti, con i relativi passaggi si ha:

$$\sqrt{u^2} (\omega^2 + 2) = 0,93406,$$

$$u (u^2 + 1) + 2 = 0,93406,$$

$u^3 + 3u - 0,93406 = 0$, la cui unica radice in soluzione reale è 0,302158, non competendo al problema le altre due complesse coniugate.

Avendo posto $u = \omega^2 - 1$, si ha anche: $\omega = \sqrt{u^2 + 1} = \sqrt{0,91299 + 1} = \sqrt{1,091299} = 1,04465$. Considerato che $e^{1/82}$ si ha allora

$$\delta = \frac{1}{\sqrt{2 \log_e \omega}} = \frac{1}{\sqrt{2 \log_e 1,04465}} =$$

$$= \frac{1}{\sqrt{0,08736}} = \frac{1}{0,29557} = 3,38329.$$

Dalla terza equazione del sistema in ordine al momento centrale si ha:

$$\rho = \sqrt{\frac{22,04250}{(1,04465)^2 [(1,04465)^2 - 1]}} = \sqrt{221,26581} =$$

$$= 14,87501.$$

Tabella N. 2

DATI STATISTICI DEL SETTORE FEMMINILE DELLA GARA DI MARATONA PER LA DEFINIZIONE DEI PARAMETRI DELLA FUNZIONE LOG-NORMALE, DEGLI INDICI DI ASIMMETRIA E DI ACCOSTAMENTO.

A	B	C	D	E	F	G	H	I	L	M	N
18	1	1	-14	-14	196	-2744	38416	0,24423	0	1	1
19	1	1	-13	-13	169	-2197	28561	1,14322	1	0	0
20	5	25	-12	-60	720	-8640	103680	3,15345	3	2	4
21	4	16	-11	-44	484	-5324	58564	7,08542	7	-3	9
22	10	100	-10	-100	1000	-10000	100000	11,91068	12	-2	4
23	20	400	-9	-180	1620	-14580	131220	16,86803	17	3	9
24	21	441	-8	-168	1344	-10752	86016	20,97712	21	0	0
25	23	529	-7	-161	1127	-7889	55223	23,59219	24	-1	1
26	18	324	-6	-108	648	-3888	23328	24,51725	25	-7	49
27	23	529	-5	-115	575	-2875	14375	23,92410	24	-1	1
28	28	784	-4	-112	448	-1792	7168	22,19101	22	6	36
29	25	625	-3	-75	225	-675	2025	19,75220	20	5	25
30	13	169	-2	-26	52	-104	208	16,99760	17	-4	16
31	12	144	-1	-12	12	-12	12	14,22557	14	-2	4
32	11	121	0	0	0	0	0	11,63415	12	-1	1
33	14	196	1	14	14	14	14	9,33399	9	6	36
34	8	64	2	16	32	64	128	7,36967	7	1	1
35	8	64	3	24	72	216	648	5,74139	6	2	4
36	4	16	4	16	64	256	1024	4,42307	5	-1	1
37	1	1	5	5	25	125	625	3,37569	4	-3	9
38	3	9	6	18	108	648	3888	2,55623	3	0	0
39	1	1	7	7	49	343	2401	1,92310	2	-1	1
40	0	0	8	0	0	0	0	1,43897	1	-1	1
41	0	0	9	0	0	0	0	1,07190	1	-1	1
42	0	0	10	0	0	0	0	0,79555	1	-1	1
43	1	1	11	11	121	1331	14641	0,58870	0	1	1
44	0	0	12	0	0	0	0	0,43460	0	0	0
45	1	1	13	13	169	2197	28561	0,32024	0	1	1
46	1	1	14	14	196	2744	38416	0,23564	0	1	1
47	1	1	15	15	225	3375	50625	0,17322	0	1	1
	258			-1035	9839	-60159	789767	257,99818	258	58	219
								arrot. 0,00182			
								258,00000			

- A) x'_i = variabile indipendente = valore centrale della classe $x_i - x_{i+1}$
- B) y_i = variabile dipendente = frequenza
- C) y_i^2 = quadrato della variabile dipendente
- D) x''_i = variabile indipendente dall'origine arbitraria di: 31 anni di età dell'atleta = 0
- E) $x''_i y$ = prodotto D B
- F) $(x''_i)^2 y_i$ = prodotto D² B
- G) $(x''_i)^3 y_i$ = prodotto D³ B
- H) $(x''_i)^4 y_i$ = prodotto D⁴ B
- I) y^*_i = valore teorico di calcolo della variabile dipendente della funzione log-normale
- L) y^{**}_i = valore teorico unitario di arrotondamento di y^*_i
- M) $y_i - y^{**}_i$ = differenza fra frequenza effettiva e teorica unitaria
- N) $(y_i - y^{**}_i)^2$ = quadrato della differenza fra frequenza effettiva e teorica unitaria

Nota: il campo di variazione della classe di età è pari ad 1 anno per cui i valori iniziali e finali di classe corrispondono rispettivamente a 1/2 anno in meno e 1/2 anno in più di x'_i

Avendo in precedenza posto $\rho = e^{-\lambda\delta}$ sarà quindi:

$$\lambda = -\delta \log_e \rho = -3,38329 \log_e 14,87501 = -3,38329 \cdot 2,69968 = -9,13380.$$

Infine, dalla seconda equazione dell'ultimo sistema si ricava:

$$\theta = 27,98372 - \omega\rho = 27,98372 - [(1,04465)(14,87501)] = 27,98372 - 15,53918 = 12,44454.$$

Analogamente a quanto ottenuto nelle applicazioni per il settore maschile, anche per quello femminile le operazioni fin qui eseguite ci permettono di definire l'equazione della funzione log-normale interpolatrice in grado di rappresentare con apprezzabile fedeltà il fenomeno sportivo ora preso in considerazione, e cioè:

$$f(x) = \frac{[(258)(3,38329)]}{\sqrt{2\pi}(x - 12,44454)} \exp \{-1/2 [-9,13380 + 3,38329 \log_e (x - 12,44454)]^2\}.$$

Anche per il settore femminile si può riscontrare una buonissima aderenza dei valori teorici della curva log-normale rispetto a quelli sperimentali. Una particolarità che la distingue da quella del corrispondente settore maschile, a parte ovviamente l'andamento dei valori assoluti delle frequenze, è rappresentata dalla diversità dei tassi (minori) di incremento e decremento dei valori della variabile dipendente rispetto all'avanzare delle età delle atlete, fatto che comporta di conseguenza un più "disteso" andamento campanulare della curva. Gli scarti tra valori teorici ed effettivi sembrano appartenere, anche in questo caso, ad un apprezzabile livello di adattamento della funzione al fenomeno sportivo in esame. La misura di tale livello, comunque, sarà calcolata nel successivo capitolo insieme a quella relativa al settore maschile, cosa che permetterà di effettuare diretti e sicuramente interessanti riscontri tra le due diverse realtà sportive.

2.4 Considerazioni sugli aspetti analitici principali caratterizzanti i due fenomeni

Per quanto riguarda i principali aspetti che diversificano le due realtà sportive, tra i tanti possibili se ne possono elencare almeno quattro di una certa importanza, e precisamente:

1) il campo di variazione degli anni di età relativi al settore femminile (18/47) è molto più ampio di quello del settore maschile (20/42 anni), nonostante il totale delle sue frequenze risulti notevolmente inferiore al corrispondente dell'altro;

2) il valore modale del settore maschile si discosta nettamente da quello degli immediati contigui (27 e 29 anni), al contrario di quanto si verifica nel settore femminile dove per un arco di ben 7 anni (da 23 a 29 compresi) si registra una certa omogeneità di valori in quanto per un totale parziale di 158 frequenze si ha uno scarto massimo pari a sole 10 frequenze. Si noti che nel settore maschile si è in presenza di uno scarto pari a ben 24 frequenze per un totale parziale quasi equivalente (163 frequenze);

3) nel settore maschile si verifica una perfetta coincidenza dell'età relativa sia al valore modale teorico che a quello effettivo, al contrario di quanto avviene nel settore femminile dove al valore modale teorico di 26 anni di età delle atlete si contrappone un corrispondente effettivo di 28. Poiché le frequenze effettive relative alle atlete di 26 anni di età risultano inferiori a quelle corrispondenti a 23, 24, 25, 27, 28 e 29 anni, non può essere esclusa a priori, in quanto il campione estratto esprime una sua precisa ed attendibile rappresentatività proprio in virtù delle modalità di casuale estrazione dei dati statistici, la considerazione che a quella specifica età le atlete possano trovarsi, per tanti validi motivi, in un transitorio regime di flessione di rendimento qualitativo, verosimilmente anche collegabile con quanto sarà indicato subito appresso;

4) le frequenze effettive delle età fino a 24 anni del settore femminile sopravanzavano nettamente quelle del corrispondente arco di età del settore maschile, con ciò confermando anche nell'ambito dello sport quanto suggerito dai classici canoni bio-medico-fisiologici che evidenziano nella donna particolari condizioni che favoriscono un anticipato livello della sua maturazione psico-fisica rispetto all'uomo. Si osservi inoltre, in proposito, che un eventuale campione del settore femminile estratto con pari numerosità di quello maschile potrebbe realisticamente aumentare la misura dello stato di fatto riscontrato. Lieve che sia, c'è anche da rimarcare che tale situazione si ripresenta con le medesime caratteristiche a partire dai 37/38 anni di età, tenendo anche presente che le frequenze cessano del tutto a 42 anni di età nel settore maschile, mentre per quello femminile è possibile riscontrare frequenze anche fino a 47 anni di età. Infine, in quest'ultimo caso, a differenza dell'altro opposto, si può dire che obiettivamente è piuttosto difficile scindere interpretazioni reali del fenomeno da quelle formulate a qualsivoglia livello teorico.

Quanto detto finora circa un'osservazione di massima dei rispettivi andamenti tecnico-storici è inquadrabile nell'ambito eminente statistico. Naturalmente il discorso si complica alquanto a livello strettamente tecnico poiché non esistono evidentemente in pratica parametri di valutazione esclusivamente tecnica che possano agire da tramite per una efficiente comparazione oggettiva dei due settori. Nella interpretazione del fatto puramente agonistico, quand'anche in presenza delle medesime caratteristiche

tecnico-storiche del tipo di gara, occorre necessariamente richiamarsi ad elementi in grado di discriminare solo in senso oggettivo, con tutte le difficoltà che comportano casi del genere. In effetti, a questo punto, e ciò vale per tutto quello che può esserle a lui demandato, lo statistico ha chiaramente terminato il suo compito, naturalmente spettando al tecnico specializzato di puntualizzare sulla materia dalla propria ottica di osservazione, traendo eventuali specifiche conclusioni sulla scorta dell'indagine statistica condotta a livello di analisi esclusivamente matematica.

3) Morfologia delle distribuzioni di frequenza

3.1 Misura delle asimmetrie

Ora che sono stati ricavati i parametri delle curve funzionali dei due fenomeni sportivi si tratta di vedere, in primis, in quale misura le relative variabili statistiche, che le hanno conformate, si interconnettono intorno ai propri valori mediani (28 anni di età per il settore maschile e 26 per quello femminile) in relazione ad una eventuale simmetria dei rami delle curve ante e post-mediana della distribuzione. Questo a significare che si potrà parlare, qualora in presenza di una perfetta sovrapposizione dei due rami in questione, praticamente mediante l'esito di un ribaltamento intorno alla ordinata del punto mediano di una sorta di legge che governa l'andamento del primo semi-ramo del fenomeno con la medesima caratteristica ed intensità, ma in senso inverso, del secondo. Per la precisione, c'è da rilevare anche che non necessariamente la simmetria deve essere legata alla forma unimodale della distribuzione (come ad esempio risulta nel nostro caso). In effetti, questo vale per qualsiasi altra forma di curva purché sia rispettato il vincolo della perfetta sovrapposizione circa il ribaltamento già descritto di uno dei due rami sull'altro. Si potrà allora definire, in caso di mancata perfetta sovrapposizione, come asimmetria positiva quella che prevede una curva (od anche istogramma, ove nel caso) che evidenzia prolungamento dalla parte degli scarti positivi rispetto al punto mediano, mentre si dirà asimmetria negativa nel caso opposto. Si osservi in proposito, sempre nella Tav. 1, come il tratto ascendente della curva del settore maschile, nella fattispecie nell'arco di età che va da 20 a 28 anni, presenti, seppure non troppo marcatamente, un sistema di tassi di crescita di intensità superiore a quello corrispondente di decrescita da 28 a 42 anni. Di contro, quanto appena detto risulta più evidente nel settore femminile la cui caratteristica di sviluppo del fenomeno si manifesta molto chiaramente anche nel relativo grafico. Tutto questo può essere interpretato in entrambi i settori, naturalmente in misura diversa, come maggiore affluenza (densità) da parte degli atleti estratti a situarsi in uno specifico arco di età della lista all-time, cioè quello relativo alle frequenze crescenti. Per la misura delle asimmetrie, delle quali sono state de-

scritte le caratteristiche in forma generica e piuttosto grezza, ci si gioverà ora di due indici specifici molto rispondenti dal punto di vista strettamente statistico. Essi sono:

- 1) l'indice di asimmetria: γ_1 ;
- 2) l'indice di curtosi γ_2 .

Avendo già chiarito poc'anzi le caratteristiche principali dell'indice di asimmetria in ordine allo studio morfometrico della distribuzione, resta da dire qualcosa sull'altro, il quale caratterizza un altro aspetto della medesima materia di studio. Come il precedente, anche l'indice di curtosi è indirizzato alla misura di un particolare aspetto o forma della curva. Precisamente, esso misura il grado di accostamento del poligono o curva di frequenza osservata alla curva di Gauss, per il semplice motivo che non tutte le curve simmetriche e di forma campanulare sono normali. Questo allo scopo di prendere atto analiticamente di quanto differisce la distribuzione dai caratteri tipici della curva normale, tenendo presente che questa fornisce un indice di curtosi $\gamma_2 = 3$. In conseguenza di ciò, per $\gamma_2 > 3$ la curva in esame sarà detta *leptocurtica* (od anche ipernormale o iperbinomiale), mentre per $\gamma_2 < 3$ sarà detta *platicurtica* (o anche iponormale o ipobinomiale), essendo per $\gamma_2 = 3$ la curva normale *mesocurtica*.

Il calcolo di tali indici, considerati i numerosi dati statistici, si presenta alquanto laborioso nel caso di impiego diretto delle espressioni di definizione. In casi del genere è molto opportuno ricorrere al metodo dei momenti, per l'appunto già descritto ed adottato in precedenza per altro scopo (calcolo dei parametri delle curve log-normali). Per il settore maschile, come si ricorderà, era stato subito calcolato di passaggio l'indice di asimmetria γ_1 . Lo sviluppo delle operazioni per ricavare l'indice è stato esposto nella sua interezza a proposito del calcolo del momento centrale terzo relativamente al settore maschile. Per il settore femminile il procedimento è risultato notevolmente abbreviato per ovvi motivi di analogia di calcolo. Comunque, si ripetono, per comodità, i rispettivi valori che risultavano:

- 1) Settore maschile $\gamma_1 = 19,04379 / 41,80035 = +0,45559$.
- 2) Settore femminile $\gamma_1 = 96,66452 / 103,48831 = +0,91474$ dai quali, come è stato già detto, per il settore femminile si evidenzia una asimmetria positiva praticamente doppia dell'altra, da attribuire a motivazioni già abbondantemente discussi.

Vediamo ora, però, l'altro aspetto di disnormalità. L'indice di curtosi, nella sua formulazione generalizzata (3), è dato da:

$$\gamma_2 = \frac{\sum_{i=1}^n (x_i - \mu)^4 y_i}{N \sigma^4} - 3.$$

(3) Il coefficiente di eccesso o di curtosi (kurtosis) di Pearson così formulato è in diretta relazione al valore 3 della curva normale.

Come accennato in precedenza, si ricorre per praticità di calcolo al metodo dei momenti, avendo già a disposizione diversi parametri calcolati per altri motivi e che possono essere sfruttati giovandosi di alcune relazioni che subito si indicano successivamente alla formula che è:

$\gamma_2 = \frac{m'_4}{\sigma^2}$, tra l'altro notevolmente più compatta della precedente.

In questo caso si ha:

m'_4 = momento centrale quarto;

σ = deviazione standard.

Il momento centrale quarto in termini di momenti dall'origine è dato da:

$m'_4 = m_4 - 4 m_1 m_3 + 6 m_1^2 m_2 - 3 m_1^4$, dove il secondo membro dell'uguaglianza è dato dallo sviluppo di m'_4 in termini di momenti dall'origine in relazione alla formula:

$$m'_4 = \frac{\sum_{i=1}^n (x_i - \mu) y_i}{\sum_{i=1}^n y_i}$$

Poiché i momenti dall'origine sono dati da:

$$1) m_1 = \frac{\sum_{i=1}^n x_i y_i}{y_i} = \frac{-1158}{489} = -2,36810$$

$$2) m_2 = \frac{\sum_{i=1}^n x_i^2 y_i}{y_i} = \frac{8632}{489} = 17,65235$$

$$3) m_3 = \frac{\sum_{i=1}^n x_i^3 y_i}{y_i} = \frac{-39024}{489} = -79,80368$$

$$4) m_4 = \frac{\sum_{i=1}^n x_i^4 y_i}{y_i} = \frac{371380}{489} = 759,46830$$

(i valori del numeratore e del denominatore delle frazioni compaiono nelle colonne interessate della Tav. n. 1), si avrà:

$$m'_4 = 759,46830 - 4 (-79,80368) (-2,36810) + 6 (17,65235) (-2,36810)^2 - 3 (-2,36810)^4 = 503,14580.$$

Essendo $\sigma^4 = (3,47051)^4 = 145,06853$, l'indice di curtosi per il settore maschile sarà dato da:

$$\gamma_2 = \frac{503,14580}{145,06853} = 3,46833,$$

per cui, con riferimento al valore 3 della curva normale si avrà: $\gamma_2 - 3 = 3,46833 - 3 = +0,46833$. Poiché si è in presenza di differenza positiva la curva è leptocurtica.

Analogamente al settore maschile, per quello femminile, sempre in relazione ai propri momenti dall'origine si ha:

$$1) m_1 = \frac{-1035}{258} = 4,01163$$

$$2) m_2 = \frac{9839}{258} = 38,13566$$

$$3) m_3 = \frac{-60159}{258} = 233,17442$$

$$4) m_4 = \frac{789767}{258} = 3061,11240.$$

Poiché il momento centrale quarto in termini di momenti dall'origine è dato dalla nota relazione: $m'_4 = m_4 - 4 m_1 m_3 + 6 m_1^2 m_2 - 3 m_1^4$, della quale ci si è già giovato in precedenza per il settore maschile, risulterà:

$$m'_4 = 3061,11240 - 4 (-233,17442) (-4,01163) + 6 (38,13566) (-4,01163)^2 - 3 (-4,01163)^4 = 2224,84693.$$

Essendo $\sigma^4 = (4,69494)^4 = 485,87011$, l'indice di curtosi per il settore femminile sarà dato da:

$$\gamma_2 = \frac{2224,84693}{485,87011} = 4,57909, \text{ per cui, con riferimento al}$$

valore 3 della curva normale, si avrà $\gamma_2 - 3 = 4,57909 - 3 = +1,57909$.

Poiché si è in presenza di differenza positiva la curva, anche in questo caso è leptocurtica.

3.3 Confronto dei dati di calcolo

I valori di asimmetria e di curtosi delle curve rappresentative dei due fenomeni, e precisamente:

	Asimmetria	Curtosi
Settore maschile	+0,45559	+3,46833
Settore femminile	+0,91474	+4,57909

ci informano che, pur avendo le due curve valori positivi di asimmetria e di curtosi, queste si differenziano per intensità di carattere, e cioè asimmetria e curtosi del settore femminile rispettivamente più che doppia e più del 32% rispetto al settore maschile. Ciò che riguarda il valore di curtosi è da mettere in relazione al fatto che questa risulta molto sensibile all'altezza, alla lunghezza delle "code" e, seppure in misura inferiore, alla "gibbosità" della curva stessa negli intorni del suo massimo. Per avere una conferma della esattezza del calcolo numerico ci si può peraltro riferire alla disuguaglianza di Wilkins, il quale, in relazione agli indici di asimmetria e curtosi, ha dimostrato che $\gamma_2 \geq \gamma_1^2 + 1$. Si può facilmente constatare che per entrambi i settori indagati tale disuguaglianza è rispettata.

Infatti risulta:

a) Settore maschile
 $3,46833 > (0,45559)^2 + 1;$

b) Settore femminile
 $4,57909 > (0,91474)^2 + 1.$

Per quanto riguarda la procedura di riferimento al valore 3 della curva normale c'è da dire, per comprenderne appieno il significato, che un corretto riconoscimento della forma di una distribuzione statistica può essere fatto solo mediante un'analisi numerica poiché la semplice rappresentazione grafica nel piano cartesiano di una curva può non essere in grado di manifestare le sue effettive caratteristiche morfometriche (circa, appunto, l'aspetto campanulare, simmetrico e asintotico dell'asse delle x) a causa di una possibile forma illusoria della rappresentazione visiva in relazione ai moduli adottati per le scale coordinate del diagramma.

Infine, sempre a proposito della procedura di riferimento alla curva di Gauss, si tenga presente che $\gamma_1 = 0$ per qualsiasi curva di tipo simmetrico (comprese anche le non gaussiane), così come per curve diverse da quella gaussiana può bene risultare anche $\gamma_2 = 3$.

4) Significatività della differenza fra i coefficienti di correlazione lineare delle distribuzioni campionarie di frequenza dei due settori di gara

4.1 Elementi descrittivi del procedimento

Esaminato nel suo complesso il fenomeno sportivo sotto gli aspetti già descritti in precedenza, il discorso può essere convenientemente concluso con una valutazione a livello di correlazione lineare. I dati statistici iniziali relativi ai due settori interessati alla nuova analisi compaiono nelle colonne A e B delle Tabelle n. 1 e n. 2. Soprassedendo

sulla procedura di calcolo delle due equazioni di regressione e dei relativi coefficienti di correlazione lineare in quanto materia già ampiamente trattata in diverse altre precedenti occasioni, sarà sufficiente, in proposito, prendere atto della definizione delle equazioni e dei valori dei coefficienti di correlazione che risultano:

1) Settore maschile

Equazione di regressione lineare:

$$Y = 56,73320 - 1,14429 x.$$

Coefficiente di regressione lineare:

$$r = -0,36386.$$

1) Settore femminile

Equazione di regressione lineare:

$$Y = 25,43203 - 0,51791 x;$$

$$r = -0,50701.$$

Come si può facilmente osservare, i valori dei coefficienti di correlazione evidenziano un differente grado di connessione dei dati statistici dei due aspetti settoriali del fenomeno. Peraltro, si può anche notare che al variare per entrambi i settori dei valori della medesima variabile indipendente *età dell'atleta*, corrisponde un diverso incremento della variabile dipendente *n. di frequenze riscontrate* nella propria lista all-time (netta differenza di valore dei coefficienti angolari delle equazioni).

Il problema di tale tipo di indagine può essere allora posto in questi termini: poiché tutti i fattori (o quanto meno la stragrande maggioranza) che intervengono nell'andamento del fenomeno possono essere inquadrati in un medesimo contesto in relazione al valore tecnico delle prestazioni atletiche (vedi: età dell'atleta, nazionalità, provenienza da specifica grande area geografica, caratteristiche antropometriche, scuola, consolidate tradizioni sportive nella specialità, ecc.) si tratta di verificare, con metodi statistici incontestabili, se la differenza esistente fra i due differenti coefficienti di correlazione può o meno essere giustificata nell'ambito di una ammissibile oscillazione di natura campionaria, e stabilire, conseguentemente, che non esiste diversità di incidenza dei fattori in questione.

In altri termini, si tratta praticamente di stabilire se le molteplici forze interagenti in ognuno dei due singoli settori presentino, nel loro insieme, caratteristiche simili per qualità (denominazione dei vari fattori possibili) e quantità (misura della loro incidenza). Nel caso di una verifica nel senso opposto, si dovrà convenire che la differenza fra i due coefficienti di correlazione è da imputare a cause e fattori sistematici propri ed autonomi per ognuno dei singoli settori. In tale evenienza, potrebbe essere opportuno, da parte dei tecnici specializzati e degli studiosi della materia, procedere in una precisa direzione di ricerca sull'interpretazione e sulla misura dell'incidenza del

complesso dei fattori presi in considerazione, ovvero, se del caso, procedere nell'esame di ulteriori parametri di valutazione del fenomeno per discriminare i sicuri e vantaggiosi contenuti innovativi (4).

4.2 Verifica della significatività mediante l'uso della trasformata Z di Fisher

Vediamo ora quello che risulta dalle procedure analitiche indirizzate alla verifica della significatività di cui si è parlato e che rappresenta senz'altro uno degli elementi più qualificanti della natura del problema.

In sostanza, si deve provare l'ipotesi che i due campioni estratti del settore maschile e femminile provengono da due popolazioni in cui i coefficienti di ρ_1 e ρ_2 sono uguali fra loro. Ciò permette di risolvere, appunto, il problema della valutazione della significatività della differenza tra i due coefficienti di correlazioni campionati r_1 e r_2 relativi a due campioni di ampiezza n_1 e n_2 (nella fattispecie: n_1 = numerosità del campione del settore maschile = n. 489 unità; n_2 = numerosità del campione del settore femminile = n. 258 unità), cosa che può anche comportare una loro eventuale differenza, ma entro certi precisi limiti di oscillazione campionaria non in contraddizione con l'ipotesi formulata.

La verifica dell'ipotesi è da porre in diretta relazione ad uno specifico teorema sulle distribuzioni di probabilità, e precisamente quello che enuncia: "Se le variabili casuali x_i sono indipendenti fra loro e tutte distribuite normalmente, con medie μ_i e varianze σ^2 ($i = 1, 2, 3, \dots, n$), la

variabile casuale $Y = \sum_{i=1}^n a_i x_i$, in cui le a_i sono delle costanti, è distribuita normalmente, con media $\sum_{i=1}^n a_i \mu_i$ e

varianza $\sum_{i=1}^n a_i^2 \sigma^2$ ". Come corollario si ha: "Se le varia-

bili casuali x_i sono tutte indipendenti tra loro ed hanno tutte distribuzione normale con stessa media μ e stessa

varianza σ^2 , la variabile casuale $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ è normale, con media μ e varianza σ^2/n .

Se la dimostrazione del teorema è piuttosto laboriosa, e pertanto se ne omette il procedimento perché non strettamente necessario, quella del poc'anzi citato corollario è invece immediata, qualora nel teorema in questione si sostituisca alla a_i il valore costante $1/n$, tenendo presente che $\mu_i = \mu$ e che $\sigma_i^2 = \sigma^2$.

Quanto appena detto è di importanza fondamentale poiché riproduce la condizione di un campione bernouilliano (5) di variabili casuali normali, indicandone la distribuzione campionaria della media. Il calcolo per la verifica dell'ipotesi formulata si giova dalla trasformata Z di r_1 e r_2 , sulla scorta del teorema appena illustrato.

I valori della trasformata Z di Fisher del coefficiente di correlazione lineare r sono proposti nella Tabella n. 3 secondo la formula:

$$Z = 1/2 \log_e \frac{1+r}{1-r}$$

Pertanto, in relazione a quanto finora

detto, si ha che la variabile casuale: $Z_1 - Z_2$ è normale con media $E(Z_1 - Z_2) = E(Z_1) - E(Z_2) =$

$$Z = 1/2 \log_e \frac{1+r_1}{1-r_1} - 1/2 \log_e \frac{1+r_2}{1-r_2}$$

e varianza:

$\text{Var}(Z_1 - Z_2) = 1/(n_1 - 3) + 1/(n_2 - 3)$, dove E deve intendersi come *expectation of* che nella terminologia inglese assume il significato di valore medio della variabile casuale in senso probabilistico. Se è vera l'ipotesi nulla $H_0 = \rho_1 = \rho_2$ si ha $E(Z_1 - Z_2) = 0$, per cui la variabile casuale:

$$\frac{(Z_1 - Z_2)}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}}$$

è normale standardizzata.

(5) L'aggettivo è riferito al matematico svizzero Jacques Bernoulli (1654-1705), autore della prima organica opera sulla teoria del calcolo delle probabilità intitolata "Arte delle congetture" (orig. "Ars conjectandi", postuma, 1715), che contiene la sua celebre proposizione (nota come teorema di Bernoulli): "Con il crescere del numero delle prove, è da attendersi con probabilità sempre maggiore che un evento si verifichi con una frequenza effettiva sempre più vicina a quella che porterebbe a prevedere la sua probabilità teorica".

(4) Un interessante traguardo di ricerca in questo senso potrebbe essere offerto, ad esempio, da una applicazione della cosiddetta *analisi fattoriale* sulla base di un congruo numero, meglio se molto elevato, di variabili, anche originariamente eterogenee fra loro, per poter successivamente ricavare un organico limitato numero di nuove variabili in grado di definire le caratteristiche del fenomeno in modo molto compatto ed al tempo stesso altrettanto efficiente. In tale ottica di osservazione queste nuove variabili, anzi più precisamente raggruppamenti di variabili "esplicative", non costituiscono grandezze specifiche da porre, come è consuetudine, in reciproca relazione diretta secondo le classiche elaborazioni di regressione e correlazione, bensì variabili di volta in volta tutte dipendenti e tutte indipendenti. Il pregio maggiore di tale procedura di analisi (multivariata) è quindi dato, secondo il concetto espresso da I. Adelman e C. Taft-Morris che hanno approfondito problematiche di questo genere, da una specifica applicazione di massima potenza ed efficienza dell'analisi di regressione di variabili *osservate* (molte) su fattori *non osservati* (pochi). In sostanza, per quanto può riguardare il fenomeno sportivo in questione, l'analisi potrebbe essere indirizzata per "generalizzare", e "classificare" alcuni fermi ed inoppugnabili criteri di base per la determinazione dei loro fattori comuni, ma ancora ignoti, giovandosi di particolari e sofisticate tecniche statistiche (e qui se ne potrebbe facilmente apprezzare la raffinata radice di interpretazione anche nel fenomeno da prendere in esame), le quali, tra l'altro, rappresentano lo strumento elettivo delle indagini statistiche connotate da preminenti contenuti di psicologia sperimentale - e questi non sembrano ragionevolmente essere pochi nella fattispecie che ci riguarda - che comportano presenza di tipiche implicazioni proprie delle scienze psicosometriche.

A questo punto si tratterà di verificare l'ipotesi nulla mediante valutazione di quanto la variabile casuale in questione si discosta dalla media con valore zero dalla distribuzione normale.

Riepilogando sulla fattispecie che ci riguarda: su due campioni di n. 489 unità estratte casualmente per il settore maschile e n. 258 unità per quello femminile da una lista all-time conformata su un medesimo tipo di classificazione tecnico-storica si vuole provare l'ipotesi che il tipo di settore (praticamente il sesso con tutti i parametri tecnici di valutazione ad esso attinenti) non influisce nella relazione tra età e distribuzione delle sue frequenze nella specifica qualità ed interpretazione tecnico-storica del tipo di lista, comune per entrambi i settori.

Per quanto già esposto, il procedimento di calcolo è il seguente:

1) Valori delle trasformate di Fischer (6)

a) Settore maschile

$$r = -0,36386$$

(6) Poiché la Tabella n. 3 dei valori (al quarto decimale) delle trasformate Z di Fisher prevede valori di riferimento dei coefficienti di correlazione solo fino al secondo decimale, è stato ritenuto opportuno procedere al calcolo ex novo, considerando per entrambe le variabili valori fino al quinto decimale, secondo l'usuale criterio di arrotondamento sempre seguito in questa sede

Tabella N. 3

VALORI DELLA TRASFORMATA Z DI FISHER DEL COEFFICIENTE DI CORRELAZIONE R.

$$Z = 1/2 \log_e \frac{1 + r}{1 - r}$$

r	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	r
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0601	0,0701	0,0802	0,0902	0,0
0,1	0,1003	0,1104	0,1206	0,1307	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923	0,1
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2621	0,2769	0,2877	0,2986	0,2
0,3	0,3095	0,3205	0,3316	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118	0,3
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361	0,4
0,5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777	0,5
0,6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480	0,6
0,7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,020	1,045	1,071	0,7
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422	0,8
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647	0,9

$$Z = 1/2 \log_e \frac{1 + (-0,36386)}{1 - (-0,36386)} = 1/2 \log_e \frac{0,63614}{0,36386} =$$

$$= 1/2 (\log_e 0,63614 - \log_e 1,36386) =$$

$$= 1/2 (-0,45234 - 0,31032) = 1/2 (-0,76266) = -0,38133.$$

b) settore femminile

$$r = -0,50701$$

$$Z = 1/2 \log_e \frac{1 + (-0,50701)}{1 - (-0,50701)} = 1/2 \log_e \frac{0,49299}{1,50701} =$$

$$= 1/2 (\log_e 0,49299 - \log_e 1,50701) =$$

$$= 1/2 (-0,70727 - 0,41013) = 1/2 (-1,11740) = -0,55870$$

2) Applicazione della formula

$$v.c = \frac{(Z_1 - Z_2)}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}} = \frac{(-0,38133 - 0,55870)}{\sqrt{1/(489 - 3) + 1/(258 - 3)}} =$$

$$= \frac{-0,94003}{\sqrt{0,00206 + 0,00392}} = \frac{-0,94003}{0,07733} = -12,15608 \quad (7)$$

Con riferimento ad una ipotesi alternativa $H_1: \rho_1 \neq \rho_2$ per un livello di significatività $\alpha = 0,05$, deve essere quindi rifiutata l'ipotesi nulla $H_0: \rho_1 = \rho_2$ della non influenza del tipo di settore sulla relazione fra età e numero di frequenza nell'ambito della lista in questione in quanto il valore $-12,15608$ risulta esterno all'intervallo di accettazione che per il predetto livello di significatività va da $-1,96$ a $+1,96$.

Si dovrà ammettere, di conseguenza, che non è manifestamente falsa l'ipotesi che per ognuno dei due tipi di settore di gara esista una differente "legge sportiva" che conforma la propria lista all-time, sebbene dal punto di vista della correlazione fra le età e le loro distribuzioni di frequenza appaia logico presumere, nel complesso, una certa coincidenza, per entrambi i settori, di fattori essenziali ed interagenti per il raggiungimento da parte degli atleti di un così ambito traguardo sportivo.

Verosimilmente il sesso dell'atleta, a prescindere naturalmente dalla differenza tecnica esistente fra l'una e l'altra lista (perché ininfluente su questo tipo di discorso), attribuisce un ruolo determinante nella diversa distribuzione della frequenza per età, sulle cui caratteristiche analitiche è stato detto in precedenza.

(7) I valori Z_1 e Z_2 della formula possono essere calcolati anche mediante l'uso di una tavola di tangenti iperboliche in quanto:

$$1/2 \log_e \frac{1+r}{1-r} = \text{arch. tanh. } r.$$

Sarebbe senz'altro auspicabile, in tal senso, una ricerca extra-statistica proprio perché in massima parte le metodologie di base della preparazione atletica e le interpretazioni agonistiche della gara stessa sembrano procedere, dopotutto, lungo direttive ed itinerari paralleli.

5) Determinazione dell'indice quadratico di accostamento tra i valori empirici e teorici delle funzioni log-normali

5.1 Elementi descrittivi e calcolo della sua misura

Quale misura che definisca il grado di accostamento tra i valori empirici del fenomeno e quelli teorici delle funzioni rappresentative può essere considerata appropriata quella relativa ad una funzione del tipo:

$$h = (y_1 - y_1^*), (y_2 - y_2^*), (y_3 - y_3^*), \dots (y_n - y_n^*),$$

dove:

h = indice di accostamento;

y = valore effettivo;

y^* = valore teorico della funzione log-normale tale che possa soddisfare le seguenti due condizioni:

1) attribuzione della stessa importanza a due scarti uguali, ma di segno opposto;

2) aumento del valore numerico direttamente proporzionale al crescere degli scarti.

È necessario premettere che gli indici assoluti di accostamento si ottengono eseguendo la media aritmetica semplice o quadratica degli scarti in valore assoluto:

$$a) I_{a1} = \frac{\sum_{i=1}^n |y_i - y_i^*|}{n}$$

$$b) I_{a2} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i^*)^2}{n}},$$

in relazione rispettivamente alla media semplice aritmetica e quadratica.

I valori numerici di tali indici, però, sono in diretta dipendenza delle grandezze delle y_i , per cui occorre rapportarli, al fine di poter effettuare corretti confronti, alla media aritmetica delle frequenze empiriche.

Le due formule diventano allora rispettivamente:

$$a) I'_{a1} = \frac{I_{a1}}{\sum_{i=1}^n y_i/n} = \frac{\sum_{i=1}^n |x_i y_i^*|}{\sum_{i=1}^n y_i/n}$$

$$b) I_{a2}'' = \frac{I_{a2}}{\sum_{i=1}^n y_i/n} = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - y_i^*)^2/n}{\sum_{i=1}^n y_i/n}}}{\sum_{i=1}^n y_i/n}$$

Poiché per la fattispecie che ci riguarda risulta indubbiamente più rispondente alla natura del fenomeno l'applicazione del secondo (perché più sensibile agli scarti elevati), il confronto sarà stabilito in sua funzione. Anzi per meglio conformare l'indice ad un criterio unico di interpretazione dello specifico rapporto, al denominatore sarà molto opportunamente considerata la media quadratica in luogo di quella aritmetica, per cui l'indice, in definitiva sarà da:

$$a) IQ = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - y_i^*)^2/n}{\sum_{i=1}^n y_i^2/n}}}{\sqrt{\frac{\sum_{i=1}^n (y_i - y_i^*)^2}{\sum_{i=1}^n y_i^2}}}$$

I valori degli indici quadratici di accostamento risulteranno quindi:

$$a) IQ (m) = \sqrt{\frac{582}{20405}} = \sqrt{0,02852} = 0,16888 =$$

= **16,888 %** della corrispondente media quadratica per il settore maschile.

$$b) IQ (f) = \sqrt{\frac{219}{4564}} = \sqrt{0,047984} = 0,21905 =$$

= **21,905 %** della corrispondente media quadratica per il settore femminile.

Anche non essendo tali valori la misura di un ineccepibile adattamento delle relative funzioni ai singoli fenomeni settoriali, pur tuttavia è possibile rilevare, nella sostanza, che questi ne rispecchiano fedelmente il carattere fondamentale, che è poi quello che più interessa in tale situazione, e cioè la capacità di poter stabilire una corretta comparazione fra gli esiti delle due singole analisi.

Il rapporto fra le due misure, precisamente $16,888/21,905 = 0,77096$, ci informa praticamente che il tipo di funzione adottato presenta un adattamento migliore per quanto riguarda il settore maschile, quantificabile appunto nella misura di $1 - 0,77096 = 0,22904$, cioè pari al 22,904% di una migliore aderenza al fenomeno nei confronti dell'altro analogo del settore femminile.

6) Conclusione

Dallo studio delle due distribuzioni si è visto, in sostanza, che pur avendo i due settori un andamento storico ricon-

ducibile ad un medesimo tipo di funzione rappresentativa, tuttavia le caratteristiche specifiche di ognuno non appaiono tali da poter permettere l'attribuzione ad una stessa spiegazione il loro sviluppo, nonostante sia ragionevole supporre che la gara di maratona, in relazione al preciso significato storico della dinamica della sua evoluzione, preveda il dominio di un insieme di ben precise norme di comportamento di "vita atletica" non troppo dissimile fra l'uno e l'altro sesso di atleta, nel senso che quanto riguarda sotto ogni aspetto le metodiche di allenamento del periodo pre-gara e l'interpretazione agonistica del tipo di competizione possono ritenersi nel complesso rispettate in egual misura per entrambe le categorie di atleti (8).

Riesaminiamo brevemente i principali elementi statistico-tecnici dei due settori:

1) l'età media dell'atleta risulta di anni 28,63190 (pari ad anni 28 e 8 mesi) per il settore maschile, mentre se ne registra una di anni 27,98372 (pari ad anni 28) per quello femminile, per cui si può constatare un anticipo medio di ingresso in lista di ben 8 mesi a favore del settore femminile;

2) il valore modale (vale a dire l'età più frequente) risulta pari a 28 anni per entrambi i settori, per cui la "capacità" di ingresso in lista in funzione dell'età equipara i due sessi, se così si vuole intendere il grado di maturazione atletica più frequente per ottenere un traguardo storico-sportivo di tale importanza;

3) la deviazione-standard risulta di anni 3,47051 (pari a 3 anni e 1/2 circa) per il settore maschile, mentre il settore femminile registra un valore di anni 4,69494 (pari a 4 anni e 8 mesi), con una differenza pari a ben 1 anno e 2 mesi, cosa che ci informa subito sulla presenza di una maggiore dispersione di età nel settore femminile;

4) la singolare diversità di compattezza fra i due settori, risultando negli intorni del valore modale molto più omo-

(8) Nell'eventualità di non poter dare scontata, nemmeno in linea generica, tale asserzione, si renderebbe indispensabile, al fine di una corretta formulazione di ipotesi alternativa, appurare dal punto di vista quantitativo il senso di tutto ciò che possa riguardare almeno la diversità dei tipi e delle modalità dei carichi di lavoro (e relative specifiche attinenze) inerenti all'uno e l'altro sesso d'atleta. Un problema di questo genere, però, dovrebbe necessariamente essere spostato su altri termini propositivi, che naturalmente esulano in gran parte dalla materia trattata in questa sede. In effetti, l'analisi fatta prende in considerazione il fenomeno solo sotto l'aspetto numerico della sua classificazione in funzione di un medesimo criterio storico-sportivo scelto ed applicato con identico procedimento ad entrambi i settori, soprassedendo deliberatamente su implicazioni che investono problematiche di troppo spinta natura extra-statistica. Se del caso, è comunque sempre possibile, anche a così diverso livello di analisi, il raggiungimento di più articolati traguardi di ricerca mediante l'uso di ulteriori strumenti statistici suggeriti da una più o meno raffinata metodologia di analisi.

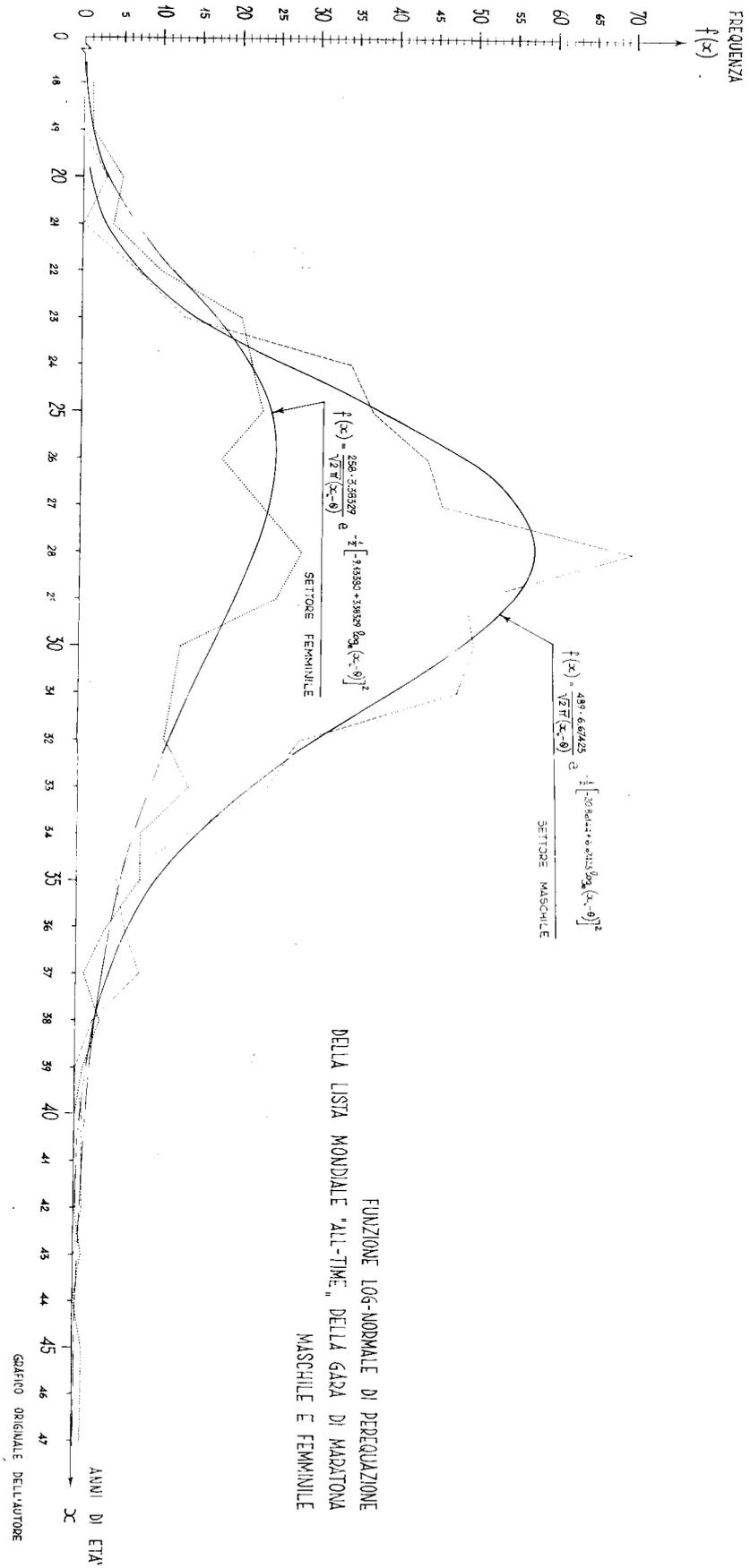


Tavola N. 2

genea quella del fenomeno a livello femminile (la massima frequenza assoluta del settore maschile appare nettamente isolata rispetto alle sue immediatamente contigue), attribuisce verosimilmente un minor grado di "maturazione" di ingresso in lista agli atleti di età inferiore a 28 anni ed una loro corrispettiva scarsa tendenza a rimanervi dopo l'età in questione, la quale in questa ottica di osservazione funge chiaramente da discriminante (9);

5) riscontro di superiorità netta delle frequenze assolute (confermata abbondantemente dalle teoriche) del settore femminile per atleti fino a 24 anni, nonostante esista un altrettanto sfavorevole rapporto (n. 258 unità rilevate contro 489) fra la numerosità del campione relativo alle atlete e l'altro;

6) riscontro di superiore precocità e longevità a favore del settore femminile (18 anni contro 20 e 47 contro 42 rispettivamente), con la precisione che nulla vieta ad un riscontro più marcato a livello di universo del fenomeno;

7) valori superiori di asimmetria e curtosi da parte del settore femminile che, giusto in termini strettamente statistici, misurano in forma quantitativa, un insieme delle precedenti considerazioni (si può notare, in proposito, come le differenze degli indici siano piuttosto sensibili);

8) netta differenza esistente a livello delle rispettive correlazioni ($r = -0,50701$ per il settore femminile in contrapposizione ad $r = -0,36386$ per quello maschile), che indica, più di quanto non sembri, uno squilibrio di connessione fra età e frequenze all'interno del fenomeno a livello maschile.

Infine, a conferma delle diverse caratteristiche dei due fenomeni a livello della specifica correlazione, si può insistere sull'importanza del valore della trasformata Z di Fischer (-12,15608), che risulta esterno ai valori-soglia di significatività della variabile casuale del test che indica esistenza o meno di influenza del sesso dell'atleta nei confronti del rapporto età/frequenza all'interno del proprio settore. In effetti, per poter avallare un'affermazione contraria alle risultanze del test in questione, il valore avrebbe dovuto ricadere all'interno del campo di variazione, per $\alpha = 0,05$ di significatività, compreso tra -1,96 e +1,96.

Comunque, tutte le considerazioni fatte in ordine alla natura del fenomeno nel suo complesso, in concomitanza con le numerose discordanze riscontrate nelle varie argomentazioni affrontate (talune anche notevoli, come si è vi-

sto), suggeriscono di esaminare l'insieme dei fattori che hanno conformato le due liste all-time sotto certi precisi punti di vista di studio del fenomeno, punti di vista che rimangono naturalmente di competenza del tecnico qualificato della materia, non potendo lo statistico offrire in proposito contributi diversi da quelli scaturiti dall'indagine campionaria così come in precedenza è stata condotta.

*Indirizzo dell'Autore:
Prof. Otello Donzelli
Via Umberto Saba, 26
00144 Roma*

(9) Negli intorni del valore modale del settore femminile la dinamica della lista all-time chiaramente non esprime caratteristiche analoghe, il che fa presumere, con molta ragionevolezza, che per le atlete interagiscono con maggiore incidenza certi precisi fattori, eventualmente da individuare e "pesare" con altre appropriate procedure di analisi.