Problematiche di statistica inferenziale su alcuni aspetti stocastici di una lista mondiale "All-time" di atletica leggera

La gara maschile di maratona in un campionamento casuale a passo sistematico

Otello Donzelli
Collaboratore del Centro Studi e Ricerche della F.I.D.AL. per il Settore di Statistica Applicata

Sebbene ogni appassionato di atletica leggera conosca molto bene il significato ed i vantaggi pratici che derivano dalla consultazione di una qualsiasi lista *all-time* nel campo delle varie discipline agonistiche, non è inutile rimarcare, con l'occasione, come sia opinione diffusa che il tipo di lista in questione risponda, in massima parte, al solo scopo della conoscenza di una notizia, particolarità o curiosità "storica" nell'ambito della disciplina sportiva oggetto di osservazione.
Per la verità, la capacità specifica della lista in questione quale elemento informatore del fenomeno si spinge ben oltre questi limiti se è vero, come in effetti è, che essa può rappresentare per lo statistico od anche semplicemente per un qualsiasi operatore sportivo, uno dei più efficaci supporti operativi per poter procedere in modo congruo all'interno del fenomeno sportivo indagato, cioè in termini strettamente analitici. Tra l'altro, risulterà anche ovvio come la descrizione e la precisione delle caratteristiche sistematico-strutturali del fenomeno trattato siano in diretta connessione da una parte in riferimento al grado di ampiezza delle unità statistiche rilevate, e dall'altra in relazione al numero delle diverse modalità espresse da ogni unità stessa.
E' noto anche come liste *all-time* di vario genere, così intese, vengano proposte in pubblicazioni sportive specializzate quali annuari, almanacchi, riviste, periodici, monografie ecc., ma per forza di cose con limiti ben precisi circa la qualità dell'informazione ed il numero delle unità statistiche rilevate, per cui una eventuale classificazione, poniamo, di 20, 30 o 50 unità in questione può facilmente non riuscire ad adempiere alla precisa funzione di individuare i caratteri rappresentativi primari del fenomeno in senso rigidamente statistico, con il limite, pertanto, di soddisfare una semplice e superficiale richiesta di informazione sulla situazione cronologica o tecnica.
Sarà quindi chiaro, allora, che l'accuratezza e l'ampiezza del numero di rilevazioni statistiche quali risultano nel volume già citato nel sommario, rappresentino l'elemento-cardine irrinunciabile per lo sviluppo di un'analisi a livello di affidabilità dell'importante "universo finito" del fenomeno indagato.
L'analisi è stata mirata all'approfondimento di alcuni notevoli, e nella fattispecie anche interessanti, aspetti stocastici direttamente collegabili ai due seguenti quesiti:
1) "Sulla scorta del corpo di dati statistici a disposizione, esiste una dimostrazione analitico-matematica del fatto che la distribuzione per classi di valore tecnico delle 1292 performances, rispetto alle fasce di età degli atleti in cui queste sono state conseguite, non sia dovuta al caso, ovvero che tale distribuzione sia da attribuire a ben precise cause sistematico-strutturali di natura esclusivamente sportiva?".
2) "Esiste, parimenti, una dimostrazione parallela nel senso di cui sopra circa il fatto che il grado di compattezza (o in antitesi: di dispersione) dell'età degli autori delle 1292 performances, suddivise per categorie di atleti appartenenti a grandi aree geografico-sportive, non sia dovuto al caso, ovvero che tale grado di compatezza (o in antitesi: di dispersione) sia da attribuire anch'esso ad altrettanto ben precise cause del tipo già descritto?".
La risposta è non solo affermativa, ma anche di due diverse modalità per ognuno dei quesiti proposti: una prima modalità è da porre in relazione all'esame di tutte indistintamente le 1292 prestazioni atletiche della lista, con lo svantaggio, però, della dilatazione dell'indagine con conseguenze operative facilmente immaginabili, agli effetti dei calcoli, circa la gravosa analisi di tutti i dati statistici a disposizione; una seconda modalità, ed è quella adottata in questa sede, in relazione alla possibilità di impiego di rapide e precise tecniche di indagine campionaria. Vediamo di questa seconda modalità, appunto, gli elementi caratterizzanti l'itinerario di ricerca. In primo luogo, la numerosità ed il tipo del campione, fattori delegati a determinare l'attendibilità dell'indagine.
Per quanto riguarda la numerosità del campione, il problema consiste nella estrazione, inderogabilmente casuale, di un certo numero di unità statistiche elementari in grado di rappresentare il fenomeno come entità-campione della complessità delle sue 1292 osservazioni sperimentali, le quali, nella fattispecie costituita, formano il cosiddetto "universo finito". La dimensione del campione dovrà rispondere a certi determinati requisiti, in quanto dipendente del grado di precisione con cui si vuole condurre l'indagine.
Comunque, per poter operare entro ben definiti margini di affidabilità, sarebbe senz'altro auspicabile che la distribuzione di frequenza delle età degli atleti al momento del conseguimento delle performances fosse riconducibile, o quanto meno compatibile, alla curva di Gauss. In caso affermativo, si potrà ragionevolmente supporre che il fenomeno assuma, nell'intera popolazione di dati, un carattere di *normalità*, e che, pertanto, la distribuzione dei dati campionari possa essere perequata correttamente mediante la curva gaussiana, con tutti i vantaggi della sua standardizzazione.
Il controllo di normalità dei dati mediante ispezione diretta della specifica tabella, però, non sempre si presenta agevole per questo tipo di verifica, per cui è allora conveniente affidarsi all'esito della trasposizione grafica dei dati delle variabili nell'apposita carta anamorfica di probabilità, la quale presenta le spaziature delle ascisse secondo le comuni scale, mentre le ordinate si svolgono in ragione di una scala di probabilità espressa in %, con valori (cumulati) che iniziano da 0,01 e terminano a 99,99 (lo specifico diagramma in seguito proposto renderà chiaro la non rappresentabilità grafica dei valori 0 e 100, in perfetta sintonia con l'andamento asintotico delle *code* della curva normale).
Considerando nella citata carta anamorfica le intensità in corrispondenza di determinati valori della variabile indipendente x, se la distribuzione integrale crescente (cioè la funzione di ripartizione) di una presunta curva normale si sistemerà nel grafico più o meno lungo una ideale retta signifi-

cherà che il fenomeno è riconducibile alla dinamica gaussiana. La Tabella n. 1 e la Tavola n. 1 ci confermano, appunto, che il fenomeno sportivo, può essere ragionevolmente assimilato alla codificazione della curva normale, essendo trascurabili, nel complesso, le deviazioni dei dati dalla retta perequatrice. Possiamo, perciò, affidarci alla formula specifica che in proposito suggerisce il numero di unità statistiche da considerare e che dovranno essere estratte con procedura casuale. La formula in questione è data da:

$$\frac{L}{2} = \lambda \frac{s}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$, dove nella nostra fattispecie:

$\frac{L}{2} = 0.5$ = 6 mesi di età (in più o in meno della vera media) = semicampo di variazione del parametro in oggetto, cioè dell'età media degli atleti (il cui valore effettivo dell'intera lista *all-time* qui risulta ininfluente), i quali concorrono a formare le 1292 performances;

λ = 1.645 = valore di ascissa dello scarto quadratico medio della curva normale standardizzata (per un intervallo di confidenza pari al 90%);

s = 3.29 = scarto quadratico medio ricavato da indagine-pilota;

N = 1292 = numerosità dell'*universo finito;*

n = numerosità incognita del campione da estrarre casualmente dall'*universo finito.*

Si avrà quindi:

$$0.5 = 1.645 \frac{3.29}{\sqrt{n}} \sqrt{\frac{1292 - n}{1292 - 1}},$$

$$0.5 = \frac{5.41205}{\sqrt{n}} \sqrt{\frac{1292 - n}{1291}},$$

$$0,5 \sqrt{n} = 5,41205 \sqrt{\frac{1292 - n}{1291}},$$

$$0.25\ n = \frac{29.29028\ (1292 - n)}{1291},$$

$$322.75\ n = 29.29028\ (1292 - n),$$

$$352.0428\ n = 37843.04176,$$

$$n = \frac{37843.04176}{352.0428} = 107.49557$$ arrotondato a 108 per necessario vincolo ad eccesso.

Per l'operazione di estrazione casuale è stato fatto riferimento alla formula del cosiddetto *"passo sistematico"*, vale a dire:

$$\frac{N}{n} = k$$ dove:

N = numero di osservazione dell'*universo finito;*

n = dimensione numerica cercata del campione;

k = *"passo sistematico"* da applicare ai numeri d'ordine delle prestazioni atletiche della lista *all-time* secondo la dimensione numerica del campione appena citata,

per cui si avrà:

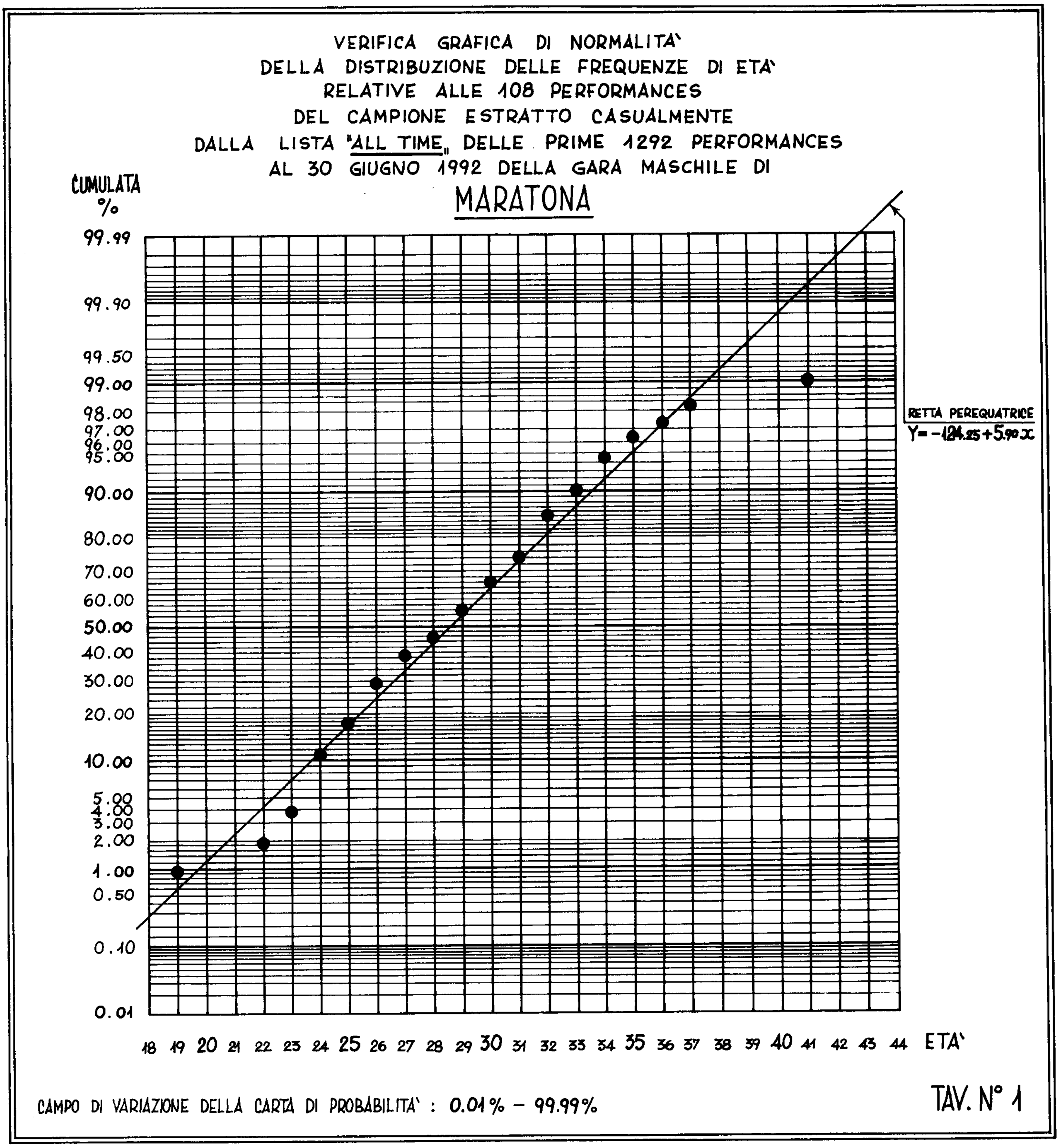
$$\frac{1296}{108} = 11.96$$ arrot. a 12

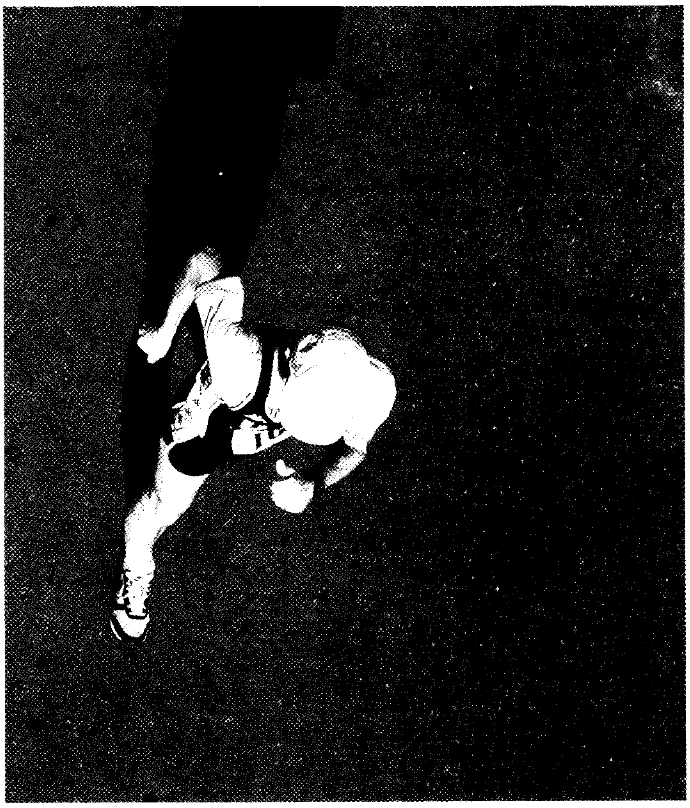
Per l'inizio della sequela è stato opportunamente considerato l'atleta al primo posto della lista in questione al 30 Giugno 1992, cioè il keniano Dinsamo Belayne che ha corso la maratona di Rotterdam nel 1988 nel tempo record di 2:06:50.
L'elenco completo delle estrazioni risulta in dettaglio nella Tabella n. 2. Non sarà inutile sottolineare che diversi atleti compaiono più volte in relazione ad altrettanto diverse estrazioni, le quali identificano le diverse performances secondo il tempo e l'età dell'atleta *al momento* del loro conseguimento. Ciò proprio in ossequio alla "filosofia sportiva" della lista *all-time*, che gratifica l'atleta nella menzione di ognuna delle proprie eventuali prestazioni al vertice storico della specialità.
Consideriamo, ora, il problema la cui soluzione è legata alla risposta che riguarda il primo quesito. L'analisi dovrà essere indirizzata, per quanto già detto, alla indicazione della misura in cui le frequenze osservate (per ogni combinazione di classe di età e fascia di tempi tecnici) differiscono dalle fre-

TABELLA N. 1

DISTRIBUZIONE INTEGRALE DI FREQUENZA DELLE ETÀ DEGLI ATLETI RELATIVE ALLE 108 PERFORMANCES DEL CAMPIONE CASUALE ESTRATTO DALLA LISTA *ALL TIME* MONDIALE AL 30 GIUGNO 1992 DELLA GARA MASCHILE DI MARATONA

N.	ETÀ	FREQUENZA ASSOLUTA	FREQUENZA CUMULATA ASSOLUTA	FREQUENZA CUMULATA %
1	19	1	1	0,92
2	20	-	-	
3	21	-	-	
4	22	1	2	1,85
5	23	2	4	3,70
6	24	8	12	11,11
7	25	7	19	17,59
8	26	12	31	28,70
9	27	9	40	37,04
10	28	9	49	45,37
11	29	11	60	55,55
12	30	9	69	63,89
13	31	12	81	75,00
14	32	11	92	85,18
15	33	5	97	89,81
16	34	5	102	94,44
17	35	2	104	96,30
18	36	1	105	97,22
19	37	1	106	98,15
20	38	-	-	
21	39	-	-	
22	40	-	-	
23	41	1	107	99,07
24	42	1	108	100,00


VERIFICA GRAFICA DI NORMALITA'
DELLA DISTRIBUZIONE DELLE FREQUENZE DI ETA'
RELATIVE ALLE 108 PERFORMANCES
DEL CAMPIONE ESTRATTO CASUALMENTE
DALLA LISTA "ALL TIME" DELLE PRIME 1292 PERFORMANCES
AL 30 GIUGNO 1992 DELLA GARA MASCHILE DI
MARATONA
CUMULATA %
99.99
99.90
99.50
99.00
98.00
97.00
96.00
95.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
5.00
4.00
3.00
2.00
1.00
0.50
0.10
0.01
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
ETA'
RETTA PEREQUATRICE
Y= -124.25+5.90x
CAMPO DI VARIAZIONE DELLA CARTA DI PROBABILITA' : 0.01% - 99.99%
TAV. N° 1



quenze che ci si potrebbe aspettare in assenza di legami associativi e gli esiti che definiscono le righe e le colonne di una tabella a doppia entrata che caratterizzi le varie combinazioni di modalità delle due variabili codificate *età* e *tempo tecnico*. Tutto ciò che è possibile con l'ausilio di una particolare statistica chiamata χ^2 (chi quadro) (1), la quale riveste un ruolo di fondamentale importanza nel campo dell'indagine squisitamente probabilistica.
Peraltro, considerando che l'analisi adottata è di tipo campionario, ci si deve innanzitutto domandare se un determinato legame associativo fra i dati non sia dovuto al caso, nel senso di aver estratto proprio *quel* campione della popolazione e non magari *un altro* di caratteristiche potenzialmente differenti. Comunque, il fenomeno da noi posto sotto osservazione con l'estrazione casuale dei dati, di cui abbiamo già visto le modalità di effettuazione, è stato inquadrato secondo il combinato disposto della Tabella n. 3, a doppia entrata, nella quale i dati statistici relativi alle 25 caselle risultanti della combinazione delle 5 modalità di riga per le 5 modalità di colonna dei 2 caratteri forniscono l'insieme dei parametri la cui elaborazione darà la risposta al quesito secondo la proposizione tecnica della tabella stessa.
In base all'ipotesi di indipendenza dei fenomeni in questione è possibile, in effetti, associare alla frequenza reale di osservazione un'altra, teorica, che sarà determinata, appunto, ai fini della verifica di una eventuale relazione significativa.
Poiché la distribuzione di χ^2 dipende anch'essa dal numero delle unità che formano il campione, e quindi dai *gradi di libertà* (2), l'analisi sarà condotta in funzione dei valori tabellari g.l., e cioè:

g.l. = (r - 1) (c - 1), formula nella quale

r = numero di righe

c = numero delle colonne

della tabella a doppia entrata (detta anche *di contingenza*).
Pertanto sarà:

g.l. = (5 - 1) (5 - 1),

cioè g.l. = 16.

---

(1) La prima trattazione di tale statistica risale al 1875 ad opera del tedesco F.R. Helmert, autore di alcuni studi sulla distribuzione della somma dei quadrati degli scarti standardizzati dei campioni di una popolazione. La descrizione del concetto di *costante campionaria* con la sua formulazione matematica (1892) è però dovuta all'italiano Paolo Pizzetti (Parma, 1860 - Pisa, 1918), all'epoca professore di Geodesia nell'Università di Genova e dal 1900 professore di Geodesia e Meccanica Celeste in quella di Pisa.

(2) Il concetto di *gradi di libertà* nel campo della statistica è stato introdotto dall'inglese R.A. Fisher (East Finchley, 1890 - Adelaide, 1962), che lo ha proposto in analogia all'interpretazione dei gradi di libertà in un sistema dinamico. Vediamo un elementare esempio pratico di questo concetto.
Se viene imposto un vincolo tale che : a + b + c + d = 100, esiste chiaramente *libertà* di fissare 3 dei 4 parametri dell'equazione, in quanto il quarto è determinato in modo univoco; in effetti se poniamo, ad esempio, a = 10, b = 25, c = 30, d necessariamente dovrà risultare 35. Poiché in questa fattispecie esiste libertà di fissare 3 dei 4 parametri dell'equazione, si dirà che questa usufruisce di 3 gradi di libertà. Se però si impone ai quattro termini un ulteriore vincolo, ad esempio: $a^2 + b^2 + c^2 + d^2 = 2842$, non sarà più possibile fissare liberamente 3 dei 4 parametri, bensì solo 2, perché il terzo, se diverso da quello effettivo (ergo: liberamente scelto), non potrà permettere la verifica dell'equazione. In effetti, se a = 10, b = 25, i valori c, d possono essere univocamente determinati, per la verifica dell'equazione, solo con la soluzione contestuale di c + d = 65 e $c^2 + d^2 = 2117$, da cui, risolvendo in d, per esempio, si avrà:
$(65 - d)^2 + d^2 = 2117$, cioè: $4225 + d^2 - 130\,d + d^2 = 2117$.
Quindi sarà: $2d^2 - 130d + 2108 = 0$, per cui alle radici di questa equazione corrisponderanno i valori dei 2 residui parametri. Infatti:

$$d = \frac{130 \pm \sqrt{16900-16864}}{4} = \begin{cases} 31 \\ 34 \end{cases}$$, i cui valori verificano ancora,

sempre con i medesimi 4 termini, l'uguaglianza a 100, e cioè:
10 + 25 + 31 + 34 = 100. Sarà chiaro, allora, che la definizione dei gradi di libertà caratterizzerà l'analisi di un fenomeno in funzione della *dimensione* del campione e della *numerosità* dei parametri in gioco.

TABELLA N. 2

DATI TECNICI RELATIVI AL CAMPIONE CASUALE DI N. 108 PERFORMANCES ESTRATTE DALLE 1292 CLASSIFICATE NELLA LISTA *ALL-TIME* MONDIALE AL 30 GIUGNO 1992 DELLA GARA DI MARATONA DEL SETTORE MASCHILE

a Numero d'ordine
b Numero di estrazione casuale con passo sistematico = 12
c Cognome dell'atleta
d Nome dell'atleta
e Nazionalità
f Anno di nascita
g Tempo tecnico conseguito
h Località di conseguimento della performance
i Anno di conseguimento della performance
l anni di età dell'atleta al momento del conseguimento della performance

a	b	c	d	e	f	g	h	i	l
1	1	DINSAMO	Belayne	KEN	1957	2:06:50	Rotterdam	1988	31
2	13	SALAH	Ahmed	DJI	1956	2:07:07	Rotterdam	1988	32
3	25	SEKO	Toshihiko	JPN	1956	2:08:27	Chicago	1986	30
4	37	DINSAMO	Belayne	KEN	1957	2:08:39	Rotterdam	1989	32
5	49	MORISHITA	Koichi	JPN	1967	2:08:53	Beppu	1991	34
6	61	LOPES	Carlos	POR	1947	2:09:06	Chicago	1984	37
7	73	PETER	Jorg	GDR	1955	2:09:14	Berlin Gr.	1984	29
8	85	MIMURA	Toru	JPN	1962	2:09:23	Beppu	1991	29
9	97	SALAZAR	Alberto	USA	1958	2:09:29	New York	1982	24
10	109	BEARDSLEY	Dick	USA	1956	2:09:36.2	Duluth	1981	25
11	121	JÖRGENSEN	Henrick	DEN	1961	2:09:43	London	1985	24
12	133	CIERPINSKI	Waldemar	GDR	1950	2:09:54.9	Montreal	1976	26
13	145	SMET	Mark	BEL	1951	2:10:00	Berchem	1979	28
14	157	PETERSEN	Pat	USA	1959	2:10:04	London	1989	30
15	169	BETTIOL	Salvatore	ITA	1961	2:10:08	New York	1989	28
16	181	JONES	Hugh	GBR	1955	2:10:11	London	1987	32
17	193	KVERNMO	Geir	NOR	1955	2:10:17	London	1987	32
18	205	WELLS	Jeff	USA	1954	2:10:20	Eugene	1979	25
19	217	TERZI	Mehmet	TUR	1955	2:10:25	London	1987	32
20	229	PITAYO	Martin	MEX	1960	2:10:29	Chicago	1984	24

(segue Tabella N. 2)

21	241	SPEDDING	Charles	GBR	1952	2:10:32	London	1987	35
22	253	SEKO	Toshihiko	JPN	1956	2:10:35	Fukuoka	1984	28
23	265	NEDI	Dereje	ETH	1954	2:10:39	Tokyo	1983	29
24	277	PLAATJES	Mark	RSA	1962	2:10:41	Los Angeles	1988	26
25	289	WAKIIHURI	Douglas	KEN	1963	2:10:47	Seoul	1988	25
26	301	POLI	P. Giovanni	ITA	1957	2:10:49	Milano	1989	32
27	313	LODWICK	John	USA	1954	2:10:54	Eugene	1979	25
28	325	PARMENTIER	Armand	BEL	1954	2:10:57	Helsinki	1983	29
29	337	FROELICK	Marty	USA	1958	2:10:59	St. Paul	1987	29
30	349	CASTILLO	Maurilio	MEX	1962	2:11:01	St. Paul	1990	28
31	361	ABE	Fumiaki	JPN	1958	2:11:04	Otsu	1985	27
32	373	IKANGAA	Juma	TAN	1957	2:11:06	Hiroshima	1985	28
33	385	JONES	Hugh	GBR	1955	2:11:08	London	1988	33
34	397	NISHI	Masayuki	JPN	1964	2:11:10	Beijing	1986	22
35	409	DRAYTON	Jerome	CAN	1945	2:11:12.8	Fukuoka	1969	24
36	421	DZHUMANAZAROV	Satymkul	URS	1951	2:11:16	Moskow	1980	29
37	433	FOSTER	Jack	NZL	1932	2:11:18.6	Christchurch	1974	42
38	445	CANNON	David	GBR	1950	2:11:21.8	Montreal	1980	30
39	457	CHAUVELIER	Dominique	FRA	1956	2:11:24	Milano	1989	33
40	469	ENAIDANI	Ryoichi	JPN	1958	2:11:27	Beppu	1991	33
41	481	BEBLO	Leszek	POL	1966	2:11:28	London	1992	26
42	493	STEFFNY	Herbert	FGR	1955	2:11:30	Munchen	1989	34
43	505	LONG	Dave	GBR	1960	2:11:33	London	1988	28
44	517	CANNON	David	GBR	1950	2:11:35	Fukuoka	1980	30
45	529	STUART	Kenny	GBR	1957	2:11:36	Houston	1989	32
46	541	DAHL	Oyvind	NOR	1951	2:11:40	K. Marx St.	1980	29
47	553	SALZMANN	Ralf	FGR	1955	2:11:41	Stuttgart	1986	31
48	565	RYAN	Kevin	NZL	1949	2:11:43	Boston	1978	29
49	577	SALAZAR	Alberto	USA	1958	2:11:44	Niagara F.	1984	26
50	589	HALBERSTADT	John	RSA	1949	2:11:46	Chicago	1982	33
51	601	EDGE	Dave	CAN	1954	2:11:48	Fukuoka	1983	29
52	613	TEN KATE	Marty	HOL	1958	2:11:49	Rotterdam	1988	30
53	625	CINDOLO	Giuseppe	ITA	1945	2:11:50.6	Reggio Emil.	1976	31
54	637	HODGE	Robert	USA	1955	2:11:52	Fukuoka	1982	27
55	649	HOAG	Steve	USA	1947	2:11:54	Boston	1975	28
56	661	SUEYOSHI	Tomio	JPN	1961	2:11:54	Rotterdam	1991	30

(segue Tabella N. 2)

57	673	ZACHARIASSEN	Allan	DEN	1955	2:11:56	Rotterdam	1986	31
58	685	WAKIIHURI	Douglas	KEN	1963	2:11:57	Tokyo	1988	25
59	697	HUH	Eui Gu	KOR	1964	2:11:58	Seoul	1990	26
60	709	FOSTER	Kevin	GBR	1952	2:11:59	Tokyo	1991	41
61	721	EYESTONE	Ed	USA	1961	2:12:00	London	1990	29
62	773	CASTILLO	Maurillio	MEX	1962	2:12:02	London	1992	30
63	745	KAWAGUCHI	Koshiro	JPN	1954	2:12:04	Tokyo	1984	30
64	757	PETERSEN	Pat	USA	1959	2:12:06	New York	1983	24
65	769	KOTOV	Wladimir	URS	1958	2:12:07	Hamburg	1990	32
66	781	SOLOVEJ	Valerij	URS	1951	2:12:09	Fukuoka	1982	31
67	793	NORWOOD	Malcom	AUS	1964	2:12:10	St. Paul	1991	27
68	805	UNETANI	Yoshiaki	JPN	1944	2:12:12	Fukuoka	1970	26
69	817	NICOSIA	Salvatore	ITA	1963	2:12:13	Seoul	1987	24
70	829	MIMURA	Toru	JPN	1962	2:12:14	Cape Town	1988	26
71	841	KOZASU	Toru	JPN	1964	2:12:16	Tokyo	1990	26
72	853	HOWIENSON	Don	CAN	1954	2:12:18	Metairie	1980	26
73	865	DAHL	Oyvind	NOR	1951	2:12:19	London	1984	33
74	877	PFEFFER	Kirk	USA	1956	2:12:20	New York	1983	27
75	889	DÖRRENBACHER	Werner	FRG	1954	2:12:22	Essonne	1980	26
76	901	SLAA	Umbe	TAN	1951	2:12:23	Fukuoka	1985	34
77	913	EDGE	Dave	CAN	1954	2:12:24	Duluth	1985	31
78	925	REVEJN	Josè	BEL	1947	2:12:26	Amsterdam	1982	35
79	937	GOMEZ	Josè	MEX	1956	2:12:27	Rotterdam	1983	27
80	949	SPEDDING	Charles	GBR	1952	2:12:28	London	1988	36
81	961	SAKTAY	Nada	TAN	1959	2:12:30	Carpi	1991	32
82	973	SIRIANO	Gary	USA	1957	2:12:32	Eugene	1983	26
83	985	PFITZINGER	Pete	USA	1957	2:12:33	Montreal	1983	26
84	997	SZUCS	Csaba	HUN	1965	2:12:34	Paris	1992	27
85	1009	RENNER	Peter	NZL	1959	2:12:35	Sacramento	1990	31
86	1021	O'REILLY	Mike	IRL	1958	2:12:37	St. Paul	1989	31
87	1033	DAVIES-HALE	Paul	GBR	1962	2:12:38	Tokyo	1991	29
88	1045	PINTO	Antonio	POR	1966	2:12:39	Carpi	1991	25
89	1057	BORDIN	Gelindo	ITA	1959	2:12:40	Roma	1987	28
90	1069	EBBA	Hailu	ETH	1950	2:12:41	Eugene	1982	32
91	1081	HUSSEIN	Ibrahim	KEN	1958	2:12:41	Boston	1989	31
92	1093	DADI	Tesfaye	ETH	1969	2:12:42	Rotterdam	1992	23

(segue Tabella N. 2)

93	1105	JÖRGENSEN	Henrick	DEN	1961	2:12:44	St. Paul	1984	23
94	1117	NAGASHIMA	Hiroshi	JPN	1959	2:12:46	Beppu	1987	28
95	1129	POLI	P. Giovanni	ITA	1957	2:12:47	Honolulu	1988	31
96	1141	HIROYAMA	Tsutomu	JPN	1966	2:12:48	Beppu	1990	24
97	1153	DUBE	Nagash	ETH	1968	2:12:49	Beijing	1987	19
98	1165	BEARDSLEY	Dick	USA	1956	2:12:50	Houston	1981	25
99	1177	DURDEN	Benti	USA	1951	2:12:51	Eugene	1982	31
100	1189	RODGERS	Bill	USA	1947	2:12:51.3	Fukuoka	1978	31
101	1201	RYAN	Kevin	NZL	1949	2:12:53	New York	1983	34
102	1213	STUART	Kenny	GBR	1957	2:12:53	London	1989	32
103	1225	GARCIA CORALLES	Diego	ESP	1961	2:12:54	London	1991	30
104	1237	RAUNIG	Tom	USA	1959	2:12:55	Chicago	1983	24
105	1249	PETERSEN	Pat	USA	1959	2:12:56	London	1086	27
106	1261	DAHL	Oyvind	NOR	1951	2:12:57	London	1985	34
107	1273	BRACE	Steve	GBR	1961	2:12:58	London	1988	27
108	1285	ANDRIOPULOS	Spiros	GRE	1962	2:12:59	W. Berlin	1989	27

TABELLA N. 3

DISTRIBUZIONE PER CLASSI D'ETÀ E TEMPI TECNICI RELATIVI AGLI ATLETI AUTORI DELLE 108 PERFORMANCES CAMPIONATE DALLA LISTA *ALL-TIME* AL 30 GIUGNO 1992 DELLA GARA MASCHILE DI MARATONA

Tempi	da 2:06:50.00 a 2:08:03.80	da 2:08:03.81 a 2:09:17.60	da 2:09:17.61 a 2:10:31.40	da 2:10:31.41 a 2:11:45.20	da 2:11:45.21 a 2:12:59.00	Totali di riga
Età						
19-23	**0** (0.074) 0.074	**0** (0.185) 0.185	**0** (0.482) 0.482	**1** (1.074) 0.005	**3** (2.185) 0.304	**4** (4)
24-28	**0** (0.833) 0.833	**0** (2.083) 2.083	**8** (5.417) 1.232	**11** (12.083) 0.097	**26** (24.584) 0.081	**45** (45)
29-33	**2** (0.889) 1.388	**3** (2.222) 0.272	**5** (5.778) 0.105	**14** (12.889) 0.096	**24** (26.222) 0.188	**48** (48)
34-38	**0** (0.167) 0.167	**2** (0.417) 6.031	**0** (1.082) 1.082	**2** (2.417) 0.072	**5** (4.917) 0.001	**9** (9)
39-43	**0** (0.037) 0.037	**0** (0.093) 0.093	**0** (0.241) 0.241	**1** (0.537) 0.399	**1** (1.092) 0.008	**2** (2)
Totali di colonna	**2** (2)	**5** (5)	**13** (13)	**29** (29)	**59** (59)	**108**

L'analisi del fenomeno è condotta in tre fasi, e cioè:

1ª Fase
Formulazione dell'*ipotesi nulla*

2ª Fase
Calcolo del valore di χ^2

3ª Fase
Confronto del valore di cui alla precedente fase con il valore critico della statistica di χ^2.

Per quanto attiene alla prima fase, l'ipotesi nulla formulata riguarda l'esistenza o meno di una relazione, nelle 1292 performances, fra le categorie del loro valore tecnico e le classi d'età degli atleti che le hanno conseguite, per cui l'ipotesi nulla risulta: H_0= non esiste relazione significativa fra le categorie di valore tecnico e le classi di età degli atleti, mentre l'ipotesi alternativa è: H_1= esiste relazione significativa fra le categorie di valore tecnico e le classi di età. Per la seconda fase il calcolo di χ^2 viene effettuato mediante la misura di tutte le frequenze osservate da confrontare, mediante l'applicazione dell'apposita formula, con quelle teoriche, le quali, a loro volta, sono calcolate tenendo presente l'eventualità della loro indipendenza, secondo logica e risposta ai due seguenti tipi di quesiti:
a) "Qual'è la probabilità che una performance estratta a caso dalle 1292 della lista *all-time* abbia registrato, ad esempio, un tempo tecnico inferiore a 2:08:03.80?".
La risposta, facendo naturalmente riferimento al primo totale di colonna della Tabella n° 3, è 2/108.
b) "E qual'è la probabilità che una performance di cui al punto a) sia relativa ad un atleta di 19, 20, 21, 22 e 23 anni di età?".
La risposta, questa volta facendo riferimento al primo totale di riga della citata tabella, è 4/108.
Di conseguenza, la probabilità che una performance possegga contestualmente entrambi i requisiti è data dal prodotto delle due distinte probabilità, ovverossia:
(2/108) (4/108) = 0.000685871. La frequenza teorica, allora, sarà data dal prodotto di tale valore per 108 (*unità* del campione estratto), e cioè:
(0.000685871) (108) = 0.074, e così via. Come si può notare nella tabella, sotto il numero - in grassetto - delle osservazioni compare tra parentesi quello teorico, sul quale sarà basato il confronto. Analogamente alla prima casella presa ad esempio, e che riporta, appunto, valore osservato = 0 e valore teorico = 0.074, le restanti forniranno i valori teorici che saranno utilizzati nella formula per il calcolo dell'indice χ^2 destinato al confronto con il corrispondente del tabulato che, per quanto ci riguarda, sarà quello critico relativo a n.16 gradi di libertà ed allo 0.05 di significatività (95% di intervallo di fiducia). Vediamo, ora, come si procede per il calcolo del valore da confrontare con quello critico.

La formula che definisce tale valore è data da:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(f_0 - f_t)^2}{f_0} \right], \text{ nella quale}$$

f_0 = frequenza effettiva del fenomeno

f_t = frequenza teorica corrispondente.

Come si è visto in precedenza, il valore teorico che figura nella prima casella è 0.074, per cui il primo valore, che sommato a quelli delle restanti caselle darà la misura dell'χ^2 (riferito, cioè, al fenomeno), è dato da:

$$\frac{(0.000 - 0.074)^2}{0.074} = 0.074$$

Seguendo ordinatamente la serie dei valori entro le parentesi quadre della tabella si avrà allora:

1ª casella (0.000 - 0.074) / 0.074 = 0.074
2ª casella (0.000 - 0.185) / 0.185 = 0.185
3ª casella (0.000 - 0.482) / 0.482 = 0.482
4ª casella (1.000 - 1.074) / 1.074 = 0.005
....
....
....
....
....
22ª casella (0.000 - 0.093) / 0.093 = 0.093
23ª casella (1.000 - 0.241) / 0.241 = 0.241
24ª casella (1.000 - 0.537) / 0.537 = 0.399
25ª casella (1.000 - 1.092) / 1.092 = 0.008

la cui somma di valori propone un χ^2 = 15.557.
Dalla tavola della distribuzione dei valori χ^2, per l'occasione tratta dal volume: *Statistica per la ricerca sociale* di Hubert M. Blalock jr., Ediz. Il Mulino, Bologna (1984), pagg. 734, 735, 736, si prende atto che il valore critico di χ^2 per n. 16 gradi di libertà e per il 95%di intervallo di fiducia è pari a 26.30. Ciò significa che questo valore di χ^2 ha il 5% di probabilità di essere dovuto al caso. Si tenga quindi presente che il valore di 15.557 è molto più piccolo, motivo per cui tale valore ha una probabilità molto più grande di essere casuale. Pertanto, la verifica ci porta a non poter respingere l'ipotesi nulla, che si traduce nel fatto che *non è manifestatamente falsa* l'ipotesi che *non esiste* relazione significativa fra categorie di valore tecnico e classi di età degli atleti che le hanno conseguite. In termini sportivi questo significa che

agli effetti della preminenza storica assoluta nella lista *all-time* della specialità le performances non dipendono, come in apparenza può sembrare in senso tendenzialmente vincolante, da uno specifico ristretto arco di età dell'atleta.
Per comprendere in altri termini agonistici il significato di tale asserzione, si può fare riferimento, ad esempio, al palese fenomeno contrario della gara di velocità dei 100 metri piani, dove è invece manifesta una certa preponderanza strutturale, nella relativa lista *all-time*, di un livello di età degli atleti alquanto precoce, dovendosi registrare forte carenza di velocisti, poniamo, di 32/35 anni, nel senso che ciò, salvo rare eccezioni, è legato strutturalmente (cioè: non casualmente) all'appartenenza della quasi totalità dei velocisti, al vertice di tutti i tempi, a classi di età che non comprendono atleti di età alquanto avanzata (atleticamente parlando, s'intende).
Diciamo allora che, nel caso della specialità maschile della maratona, si è in presenza di una particolare specifica minoranza di atleti giovani e molto giovani, ma che tuttavia non è da attribuire affatto ad una complessiva inferiorità numerica nel senso di non predisposizione anagrafica alla gara, bensì esclusivamente al minor numero assoluto, a quelle specifiche età, di maratoneti in carriera ad alto livello. In effetti, i rapporti tecnici fra partecipazioni a questo tipo di gara ed ingressi in lista di atleti appartenenti alle varie classi di età permangono in massima parte su valori omogenei e stabili.
Quanto appena detto per ribadire che non esiste a priori motivo valido dal punto di vista statistico-sportivo per sostenere la tesi che il maratoneta giovane o molto giovane possieda minore capacità qualitativa per affrontare gare ad alti livelli agonistici rispetto ad un altro di età atleticamente più avanzata. Sotto questo punto di vista è assai sintomatico il fatto che un atleta di 19 anni, l'etiopico Dube Nagash, abbia corso nel 1987 una gara di maratona in 2:12:49 (3) a dimostrazione del fatto che ad atleti i quali praticamente iniziano la loro carriera sportiva (esistono, tra l'altro, diversi casi analoghi relativi ad atleti di 20, 21 e 22 anni che confermano questa tesi) non sono preclusi prestigiosi ingressi nella lista *all-time* della gara in questione. Non è dimostrabile, invero, la tesi inversa. In effetti, non si verifica nella lista presenza di atleti con "tardivo" inizio di carriera, dovendosi constatare che le rare prestazioni atletiche di soggetti ultratrentacinquenni appartengono tutte alle fasi finali di lunghe o addirittura lunghissime carriere agonistiche consumate, tra l'altro, anche nella pratica di altre specialità di fondo. Una ulteriore considerazione, di natura squisitamente statistica, già accennata in altri termini precedentemente, può essere fatta in proposito, e cioè: alla scarsità di ingressi in lista di maratoneti delle età atleticamente più precoci corrisponde una altrettanto scarsità di loro partecipanti alla gara, fatto, questo, che pur pregiudicando il particolare aspetto nel suo valore numerico assoluto non incide più di tanto in quello relativo, cioè in quello percentuale degli ingressi in lista, del resto molto più significativo per la descrizione del fenomeno in funzione delle età degli atleti al momento del conseguimento delle loro performances. Si può notare, oltretutto, che il complesso di queste caratteristiche anagrafiche si ripete frequentemente sia nelle primissime posizioni che in tutto l'arco storico-ordinale che comprende le 1292 performances della lista. Anzi, il valore più prossimo a χ^2 = 15.557, per n. 16 gradi di libertà, cioè 15.338, è relativo ad una probabilità pari al 50% di essere dovuto al caso, per cui è assolutamente fuori di dubbio che l'ipotesi nulla, della *non esistenza* di legami strutturali fra le classi di età e quelle di valore tecnico delle performances che hanno conseguito gli atleti, non appartenga alla natura del fenomeno al prescelto livello di significatività del 5%.
Un'analisi parallela darà, ora, anche la risposta al secondo quesito tecnico proposto all'inizio. In questo caso, per poter procedere è necessaria la disponibilità di una tabella di distribuzione delle singole età relative alle categorie di atleti appartenenti a specifiche grandi aree geografiche, le quali, pertanto, sono state opportunamente così classificate:

1) Europa continentale
2) Asia nord-orientale
3) Africa centro-meridionale
4) America settentrionale
5) America centrale
6) Oceania

tenendo presente che le nazionalità degli atleti fanno riferimento, naturalmente, alle relative situazioni politiche esistenti al momento del conseguimento delle prestazioni atletiche. La Tabella n. 4, ricavata sulla base dei dati statistici proposti nella Tabella n. 2, indica le singole età degli atleti secondo il nuovo criterio di distribuzione. Per quanto riguarda l'età in cifra intera non è stato fatto riferimento all'*età compiuta*, bensì alla differenza fra l'anno di conseguimento della performance e quello di nascita dell'atleta.
L'opportuna classificazione di questi dati consentirà, come poc'anzi detto, di dare una risposta anche al secondo quesito, il quale è legato alla verifica dell'ipotesi che le varianze campionarie separate delle età relative alle 6 grandi aree geografiche stimino lo stesso valore di varianza, cioè che possano essere considerate esito di campioni estratti da un *medesimo universo* (od anche da *universi con stessa varianza*). Deve essere verificata l'ipotesi, in sostanza, che:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma_u^2$$

dove:

(3) Questa è la performance dell'atleta che compare nel campione con estrazione casuale. Si tenga presente, comunque, che esiste anche una sua superiore prestazione tecnica, non estratta, conseguita a 23 anni di età a Tokyo nel 1991 con il tempo di 2:10:00

TABELLA N. 4

DISTRIBUZIONE DELLE ETÀ DI N. 108 ATLETI DEL CAMPIONE CON ESTRAZIONE CASUALE SECONDO L'APPARTENENZA A GRANDI AREE GEOGRAFICHE

1) *EUROPA CONTINENTALE*

BEL	28	29	35									
DEN	24	31	23									
ESP	30											
FRA	33											
FRG	34	31	26									
GBR	32	35	33	30	28	30	32	41	36	29	32	27
GDR	29	26										
GRE	27											
HOL	30											
HUN	27											
IRL	31											
ITA	28	32	31	24	28	31						
NOR	32	29	33	34								
POL	26											
POR	37	25										
TUR	32											
URS	29	32	31									

N. 46 rilevazioni (42.59% del totale del campione estratto)

$\bar{x}$ = 30.28
σ = 3.61271
σ^2 = 13.05169

2) *ASIA NORD-ORIENTALE*

JPN	30	34	29	28	27	22	33	30	30	26	26	26	28	24
KOR	26													

N. 15 rilevazioni (13.89% del totale del campione estratto)

$\bar{x}$ = 27.93
σ = 3.19523
σ^2 = 10.20952

(segue Tabella N. 4)

N. 5 rilevazioni (4.63% del totale del campione estratto)

$\bar{x}$ = 32.60

σ = 5.85662

σ^2 = 34.29999

σ_1^2 = varianza Europa continentale

σ_1^2 = varianza Asia nord-orientale

σ_1^2 = varianza Africa centro-meridionale

σ_1^2 = varianza America settentrionale

σ_1^2 = varianza America centrale

σ_1^2 = varianza Oceania

σ_u^2 = varianza *Universo dei dati*

(segue Tabella N. 4)

3) *AFRICA CENTRO-MERIDIONALE*

ETH	29	32	23	19	
DJI	32				
KEN	31	32	25	25	31
RSA	26	33			
TAN	28	34	32		

N. 15 rilevazioni (13.89% del totale del campione estratto)

$\bar{x}$ = 28.80
σ = 4.32930
σ^2 = 18.74286

4) *AMERICA SETTENTRIONALE*

CAN	24	29	26	31														
USA	24	25	30	25	25	29	26	27	28	29	24	27	26	26	25	31	31	24	27

N. 23 rilevazioni (21.30% del totale del campione estratto)

$\bar{x}$ = 26.91
σ = 2.39152
σ^2 = 5.71936

5) *AMERICA CENTRALE*

MEX	24	28	30	27

N. 4 rilevazioni (3.70% del totale del campione estratto)

$\bar{x}$ = 27.25
σ = 2.50
σ^2 = 6.25

6) Oceania

AUS	27			
NZL	42	29	31	34

La mancata verifica di tale uguaglianza implicherebbe, evidentemente, differenza strutturale del grado di compattezza (cioè: deviazioni-standard più o meno sensibili) delle età nelle varie grandi aree geografiche considerate, motivo per cui diventerebbe ragionevole presumere l'esistenza, all'interno di ognuna di esse, di fattori in grado di modificare gli equilibri delle frequenze di età, con tutte le conseguenze deducibili in relazione alle varie età di massima efficienza atletica.

In questo specifico caso, risulta più conveniente, perché più rispondente alla natura del fenomeno osservato, ricorrere alla cosiddetta "varianza combinata" (nella terminologia anglosassone: "pooled variance"). Il suo valore è dato dalla formula:

$$s^2 = \frac{\sum_{i=1}^{k} s_i^2 f_i}{\sum_{i=1}^{k} f_i}$$

, la quale costituisce una media ponderata delle s_i^2, usando come pesi $f_i = (n_i - 1)$.

Vediamo, ora, come si procede in pratica per l'applicazione della formula, la quale, considerando la varianza specifica appena descritta, darà direttamente il parametro di valutazione che occorre per dare risposta al quesito.

Il procedimento di verifica della uguaglianza, o meno, delle varianze di cui si è parlato, è dovuto all'inglese M.S. Barlett, il quale ha dimostrato che la quantità:

$$-\frac{1}{c}\sum_{i=1}^{k} f_i \ln \frac{s_i^2}{s^2}, \text{ dove } c = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{f_i} - \frac{1}{f}\right), \text{ e nella quale}$$

$f = \sum_{i=1}^{k} f_i$ si distribuisce all'incirca come χ^2 con (k - 1) gradi di libertà.

Se si tiene conto che: $\ln M = 2.302585 \log_{10} M$, il test suggerito dal Barlett può essere scritto nella forma:

$$\chi^2_{(k-1)} = -\frac{2.302585}{c} f_i \log \frac{s_i^2}{s^2}, \text{ cioè:}$$

$$\chi^2_{(k-1)} = -\frac{2.302585}{c}\left(f \log s^2 - \sum_{i=1}^{k} f_i \log s_i^2\right).$$

E' agevole dimostrare che il test sfrutta il confronto fra la media aritmetica e quella geometrica delle s^2 con i pesi f_i. In effetti, la quantità:

$$\log s_g^2 = \frac{\sum_{i=1}^{k} f_i \log s_i^2}{\sum_{i=1}^{k} f_i}$$

, cioè: il logaritmo della media geometrica delle s_i^2 risulta: $f \log s_g^2 = \sum_{i=1}^{k} f_i \log s_i^2$. Sostituendo questo risultato nel test, si ha:

$$\chi^2_{(k-1)} = \frac{2.302585 f}{c} \log \frac{s^2}{s_g^2}.$$

Rimane da aggiungere, per la precisione, che nel caso di campioni piuttosto numerosi (ma questo non può riguardare il nostro caso), si ricava un valore di c = 1, per cui le relazioni precedenti si semplificano. Inoltre, nella eventualità che i campioni fossero tutti di uguale numerosità (ed anche questo non caratterizza il nostro fenomeno), il test si modificherebbe in:

$$\chi^2_{(k-1)} = -\frac{2.302585}{c}\left(k \log s^2 - \sum_{i=1}^{k} \log s_i^2\right), \text{ poiché si avrebbe:}$$

$n_1 = n_2 = n_3 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots n_k$

$f_1 = f_2 = f_3 = \ldots\ldots\ldots\ldots\ldots\ldots\ldots f_k$

$f_i = k f_o$, per cui $c = 1 + \frac{k+1}{3 k f_o}$.

Comunque, quale che sia l'alternativa in situazioni del genere, il problema della verifica di omogeneità delle varianze sarà risolto in 4 fasi:

1) viene scelto preliminarmente un certo livello di significatività α (od anche: $(1 - \alpha)$ = livello di fiducia);

2) viene formulata l'*ipotesi nulla*;

3) viene calcolato il valore di $\chi^2_{(k-1)}$ sperimentale

4) viene verificata l'alternativa di accettazione o rifiuto dell'*ipotesi nulla*.

Pertanto, gli elementi di definizione di cui sopra saranno dati, per il fenomeno che ci interessa, da:

1) *Livello di fiducia*
95%

2) *Ipotesi nulla*
$H_o : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma_u^2$, cioè: non esiste diffe-

renza fra i diversi gradi di dispersione delle età degli atleti in funzione della loro classificazione nelle relative grandi aree geografiche a cui appartengono, contro l'ipotesi contraria:

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \sigma_4^2 \neq \sigma_5^2 \neq \sigma_6^2 \neq \sigma_u^2$$

3) *Calcolo del* $\chi^2_{(k-1)}$ *sperimentale,* dove: (k - 1) = gradi di libertà

4) *Verifica di accettazione o rifiuto*

a) accettazione in caso di: $\chi^2_{(5)} < \chi^2_{(t)}$

b) rifiuto in caso di: $\chi^2_{(5)} > \chi^2_{(t)}$, dove:

$\chi^2_{(5)}$ = parametro sperimentale del fenomeno

$\chi^2_{(t)}$ = parametro critico del tabulato (per n. 5 gradi di libertà al 95% di intervallo di fiducia).

Vediamo, ora, lo sviluppo analitico relativo all'intero campione delle n. 108 performances secondo i dati della Tabella n. 5. Si avrà allora:

$$c = 1 + \frac{1}{3\,(6-1)}(0.79386 - \frac{1}{102}) = 1 + \frac{1}{15}\,(0.79386 - 0.00980) =$$

$$= 1.05227;$$

$$s^2 = \text{varianza "pooled"} = \frac{1274.43551}{102} = 12.49446, \text{ per cui:}$$

$$\chi^2_{(t)} = \frac{2.302585}{1.05227}\,[102\,(\log 12.49446) - 107.34127] =$$

$$= 2.18821\,(111.86518 - 107.34127) = \mathit{9.89926}$$

$\chi^2_{(t)}$ = valore teorico ricavato dal tabulato per n. 5 gradi di libertà ed al 95% di intervallo di fiducia = *11.070.*

Poiché 9.89926 < 11.070, deve essere accettata l'ipotesi di omogeneità delle varianze allo 0.05 di significatività. Nel caso, invece, di n. 29 performances ricavate dalle sole prime estrazioni relative ad ogni singola nazionalità facente parte del campione (Tabella n. 6) sarà:

$$c = 1 + \frac{1}{3\,(6-1)}(3.31250 - \frac{1}{23}) = 1 + \frac{1}{15}(3.31250 - 0.04348) =$$

$$= 1.21793;$$

$$s^2 = \frac{310.24112}{23} = 13.48874, \text{ per cui:}$$

$$\chi^2_{(t)} = \frac{2.302585}{1.21793}[23\,(\log 13.48874) - 21.97982] =$$

$$= 1.89057\,(25.98934 - 21.97982) = \mathit{7.58028}.$$

$\chi^2_{(t)}$ = valore teorico ricavato dal tabulato per n. 5 gradi di libertà ed al 95% di intervallo di fiducia = *11.070.*

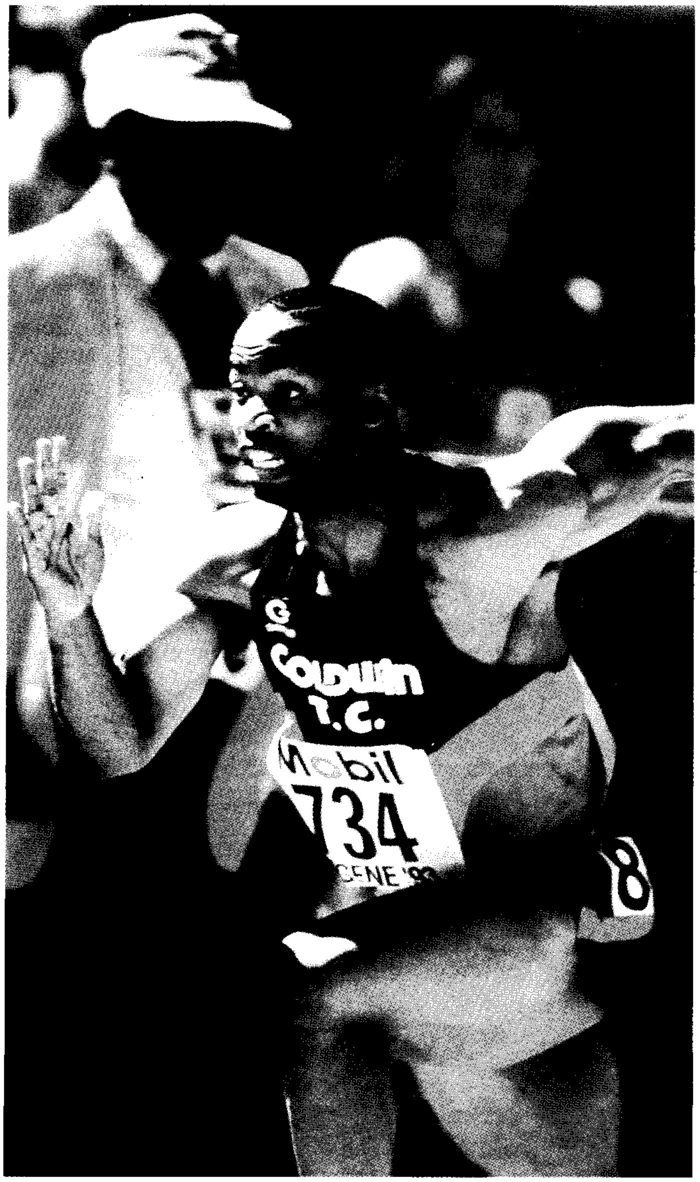


TABELLA N. 5

DATI STATISTICI PER LA VERIFICA DI OMOGENEITÀ DELLE VARIANZE CAMPIONARIE DELLE ETÀ DEGLI ATLETI SUDDIVISI PER GRANDI AREE (Campione: n. 108 atleti)

N.	Campioni	Numerosità	s_i^2	f_i	$1/f_i$	$\log s_i^2$	$f_i \log s_i^2$	$s_i^2 f_i$
1	Europa continentale	46	13.05169	45	0.02222	1.11567	50.20515	587.32605
2	Asia nord-orientale	15	10.20952	14	0.07143	1.00900	14.12600	142.93328
3	Africa centro-merid.	15	18.74286	14	0.07143	1.27283	17.81962	262.40004
4	America settentrionale	23	5.71937	22	0.04545	0.75735	16.66170	125.82614
5	America centrale	4	6.25000	3	0.33333	0.79588	2.38764	18.75000
6	Oceania	5	34.30000	4	0.25000	1.53529	6.14116	137.20000
		108		102	0.79386		107.34127	1274.43551

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} = \text{varianza campionaria corretta}$$

n = numerosità

$f_i = (n_i - 1)$

TABELLA N. 6

DATI STATISTICI PER LA VERIFICA DI OMOGENEITÀ DELLE VARIANZE CAMPIONARIE DELLE ETÀ DEGLI ATLETI SUDDIVISI PER GRANDI AREE (Atleti primi estratti di ogni singola nazionalità dell'intero campione)

N.	Campioni	Numerosità	s_i^2	f_i	$1/f_i$	$\log s_i^2$	$f_i \log s_i^2$	$s_i^2 f_i$
1	Europa continentale	17	10.43382	16	0.06250	1.01844	16.29509	166.94112
2	Asia nord-orientale	2	8.00000	1	1.00000	0.90309	0.90309	8.00000
3	Africa centro-merid.	5	5.70000	4	0.25000	0.75587	2.73049	28.80000
4	America settentrionale	2	0.00000	1	1.00000	—	—	0.00000
5	America centrale	1	0.00000	0	—	—	—	0.00000
6	Oceania	2	112.50000	1	1.00000	2.05115	2.05115	112.50000
		29		23	3.31250		21.97982	310.24112

TABELLA N. 7

DATI STATISTICI PER LA VERIFICA DI UNA IPOTESI DI ASSENZA DI OMOGENEITÀ DELLE VARIANZE CAMPIONARIE DELLE ETÀ DEGLI ATLETI SUDDIVISI PER GRANDI AREE GEOGRAFICHE (Campione: n. 108 atleti)

N.	Campioni	Numerosità	s_i^2	f_i	$1/f_i$	$\log s_i^2$	$f_i \log s_i^2$	$s_i^2 f_i$
1	Europa continentale	46	12.50000	45	0.02222	1.09691	49.36095	562.50000
2	Asia nord-orientale	15	9.50000	14	0.07143	0.97772	13.68808	133.00000
3	Africa centro-merid.	15	17.70000	14	0.07143	1.24797	17.47158	247.80000
4	America settentrionale	23	5.40000	22	0.04545	0.73239	16.11258	118.80000
5	America centrale	4	7.80000	3	0.33333	0.89209	2.67627	23.40000
6	Oceania	5	43.20000	4	0.25000	1.63548	6.54192	172.80000

Poiché 7.58028 < 11.070 deve essere accettata, anche in questo caso, l'ipotesi di omogeneità delle varianze allo 0.05 di significatività.
Del resto non poteva essere altrimenti, considerata la medesima provenienza dei dati presi in esame in questa fattispecie. Il valore ricavato 7.58028 è servito, comunque, ad evidenziare che a livello di élite (il primo estratto di ogni nazionalità è anche il primo per valore tecnico, considerando che la progressione del *"passo sistematico"* di estrazioni dei dati procede in rapporto diretto con la diminuzione del tasso tecnico della prestazione atletica) il fenomeno di omogeneità delle varianze in funzione delle grandi aree geografiche è più evidente che non quello a livello di massa. Questo è il significato della disuguaglianza 7.58028 < 11.071 rispetto all'altra 9.89926 < 11.070.
Tutto ciò, infatti, porta alla conclusione che il fenomeno presenta effettiva omogeneità delle varianze delle 6 grandi aree geografiche, nel senso della loro "compatibilità", con la stima della *vera* varianza globalmente intesa (cioè: dell'intera popolazione dei dati ovvero, in altri termini, delle 1292 rilevazioni statistiche).
È interessante notare, fra l'altro (e ciò è un altro aspetto che emerge dalla ripetizione del calcolo effettuato con il diverso criterio), che nel caso di tutte le 108 performances del campione suddiviso per grandi aree geografiche si verifica un campo di variazione di:
a) età media: 32.60 - 26.91 = 5.69 anni
pari a 5 anni e 8 mesi
b) deviazione-standard: 5.85662 - 2.50000 = 3.35662
pari a 3 anni e 4 mesi
mentre nel caso delle *29 prime performances*, relative alle singole nazionalità facenti parte del campione, si ha:

1) Europa continentale:
età media = 29.94000
pari ad anni 30 netti (a meno di 3 giorni)
deviazione-standard = 3.23014
pari ad anni 3 ed 8 mesi

2) Asia nord-orientale:
età media 28.00000
pari ad anni 28 netti
deviazione-standard = 2.00000
pari ad anni 2 netti

3) Africa centro-meridionale:
età media = 29.20000
pari ad anni 29 e 2 mesi
deviazione-standard = 2.38747
pari ad anni 2 e 5 mesi

4) America settentrionale:
età media = 24.00000
pari ad anni 24 netti
deviazione-standard = 0.00000
pari ad anni 0

5) America centrale:
età media = 24.00000
pari ad anni 24 netti
deviazione-standard = 0.00000
pari ad anni 0

6) Oceania:
età media = 34.50000
pari ad anni 34 e 1/2 netti
deviazione standard = 10.60660
pari ad anni 10 e 7 mesi

i cui valori mostrano, praticamente, quasi equivalenza fra i fenomeni di élite e di massa per Europa Continentale, Asia nord-orientale, Africa centro-meridionale ed America settentrionale, mantre per quanto riguarda l'Oceania si può osservare che l'età del primo atleta estratto di nazionalità neozelandese, Jack Foster (1932), che ha corso, all'età di 42 anni, la maratona di Christchurch in 2:11:21.6 nel 1974, ha pesantemente influenzato sui relativi parametri, senza tuttavia creare le premesse per poter respingere ,come si è visto, l'ipotesi di *omoscedascità* delle varianze al livello di fiducia proposto. Non sarà inutile sottolineare, inoltre, che i valori relativi all'area dell'America Centrale (il messicano Martin Pitayo (1964, che ha corso, all'età di 24 anni, la gara in 2:12:02 a Chicago, è l'unico atleta che ne fa parte) risultano ovviamente scarsamente rappresentativi sotto l'aspetto del criterio di élite geografica.
La constatazione di questi specifici aspetti, comunque, vuole significare semplicemente, alla luce soprattutto del risultato dei due tests effettuati in contrapposizione, che quanto finora detto è più marcato se riferito al caso delle descritte 29 migliori performances del campione. In effetti, sebbene gli atleti restino così differenziati per età, valore tecnico e provenienza geografica (e quindi per caratteristiche antropometriche, scuola, tradizione nazionale, predisposizione specifica alla gara, metodologia di allenamento, ecc.), tutti, all'interno della propria classificazione geografica, rispettano, in relazione alle variazioni di età ed ai vari parametri di valutazione connessi, standard di massima efficienza atletica pressoché equivalente e comune all'*universo finito* nel suo insieme considerato; ciò che naturalmente potrebbe bene non risultare affatto scontato in altre categorie di atleti, gare e performances. In sostanza, il complesso dei fattori che presiedono al valore tecnico manifesta caratteristiche strutturali di età che non si diversificano significativamente in ragione delle grandi aree geografiche considerate, nonostante la raccolta dei dati tecnici possa, a prima vista, far presumere propio il contrario (una poco attenta osservazione delle deviazioni-standard è quella che può più facilmente indurre all'errore di valutazione).

Un ulteriore significato di questo tipo di ricerca può quindi ravvisarsi nel fatto che qualora si volesse in qualche modo avvalorare la tesi dell'esistenza di caratteri discriminatori della qualità atletica da attribuire ad uno specifico arco di età (nel senso che questo dipenda tendenzialmente dal tipo di suddivisione geografica effettuata) necessiterebbe una verifica contraria a quella emersa dall'analisi, cioè in pratica un valore di $\chi^2_{(t)}$ superiore a quello critico, già visto, pari a 11.071, relativamente a 5 gradi di libertà con un intervallo di fiducia de 95%.
Come si è visto, però, i predetti valori rientrano nel regime delle oscillazioni campionarie che non contraddicono il reale fenomeno dell'intera popolazione dei dati, con ciò avallando la tesi che indica ininfluente, ai fini della specifica efficienza atletica, l'appartenenza ad una determinata grande area geografica dei soggetti autori della performances in questione.
Quanto appena detto, peraltro, potrà essere reso ancora più chiaro con una ipotesi di differenti *varianze corrette* e (quindi di relative differenti deviazioni-standard). L'esempio che in proposito risulta nella Tabella n. 7 si basa naturalmente sulle medesime 108 prestazioni atletiche, modificando però volutamente i valori delle varianze in questione, al fine di evidenziare, giust'appunto, la caratteristica contraria a quella definita dall'analisi del fenomeno effettivo.
Si osservi, allora, la nuova ipotetica situazione della dispersione di età nella Tabella n. 7 i cui dati non sembrano discostarsi poi di molto da quelli della effettiva realtà. In fin dei conti, tale situazione avrebbe potuto essere benissimo anche quella veramente rappresentativa del fenomeno, viste le variazioni delle relative deviazioni-standard, apparentemente lievi.
Vediamo, però , il valore dell'indice $\chi^2_{(t)}$ che risulta secondo tale ipotesi:

$$c = \frac{1}{3\,(6-1)}\left(0.79386 - \frac{1}{102}\right) = 1 + \frac{1}{15}\,(0.79386 - 0.00980) =$$

$$= 1.05227$$

$$s^2 = \frac{1246.29927}{102}\quad 12.21863\text{, per cui sarà:}$$

$$\chi^2_{(t)} = \frac{2.302585}{1.05227}\,[102\,(\log 12.21863) - 104.98812] =$$

$$= 2.18821\,(110.87630 - 104.98812) = \mathbf{12.88457}.$$

Poiché in questo caso teorico risulta: 12.88457 > 11.071, dovrebbe essere rifiutata l'ipotesi di omogeneità delle varianze di gruppo allo 0.05 di significatività. Una situazione del genere comporterebbe conseguentemente una esistenza di dispersione di età tale da differenziare in modo significativo le caratteristiche strutturali in questione proprio in ragione delle 6 grandi aree geografiche. In altre parole, ciò vorrebbe dire che l'ingresso in lista degli atleti così classificati geograficamente sarebbe necessariamente condizionato da diverse medie, deviazioni-standard e profondità di campi di variazioni di età, ovviamente caratterizzati da valori più marcatamente antitetici. Esattamente il contrario della effettiva realtà, appunto, avendo l'analisi dimostrato, in risposta ad entrambi i quesiti inizialmente proposti, che i parametri valutati non sono attribuibili al fattore età, quand'anche questo fosse posto in funzione delle grandi aree geografiche considerate, bensì ad oscillazioni campionarie *non in contraddizione* con la realtà del fenomeno.
Più precisamente, si può bene affermare che, contrariamente alle apparenze, la distribuzione degli atleti nella lista *all-time* della gara esaminata non esprime significativamente relazione né fra le età degli atleti ed i corrispondenti valori tecnici da questi conseguiti, né fra le variabilità riscontrate dal parametro *età* all'interno di ogni singola grande area geografica. È allora conseguenziale, in proposito, una domanda: "Poiché l'estrazione casuale ha proposto ben 46 atleti della sola Europa Continentale contro 62 complessivi appartenenti alle altre 5 residue grandi aree geografiche (distribuiti perdipiù in modo non omogeneo), non potrebbe ciò risultare, in qualche aspetto interpretativo, contraddittorio con le conclusioni suggerite dall'analisi?". La risposta è negativa in quanto, come si può facilmente prendere atto dallo sviluppo della formula di M.S. Barlett, un eventuale fenomeno con gruppi di uguale numerosità altro non è che un caso particolare del fenomeno generico relativo a gruppi di differente numerosità, quali, appunto, quelli trattati in questa sede. Del resto, anche l'osservazione del criterio informatore dei valori di sommatoria delle Tabelle n. 5, n. 6 e n. 7 ne rende chiaro, per altra via, il significato.
Pertanto, in considerazione delle risultanze espresse dall'analisi, la presenza (e prevalenza) di certi determinati atleti in lista non può essere attribuita ad una *diversa e più efficiente* struttura di classe della loro età e della loro appartenenza a certe specifiche grandi aree geografiche (questo è il significato emerso dall'analisi in funzione dei parametri stabiliti in premessa), motivo per cui il fenomeno può avere sì spiegazioni discriminanti i suoi molteplici aspetti tecnici e statistici, ma in senso interpretativo *estraneo* alle problematiche affrontate, cioè esso è da ricondurre, in pratica, ad un eventuale esame di *altri* rappresentativi parametri di valutazione. In effetti, l'indagine non consente, così come è stata condotta, interpretazioni di significato diverso, sia in relazione agli specifici argomenti considerati che alle modalità della loro trattazione.
Comunque, resta ancora da evidenziare un fatto, assoluta-

mente da non sottovalutare come utile premessa a quanto appresso si dirà. L'analisi ha preso in considerazione, come si è visto, solo dati statistici aggiornati al 30 Giugno 1992 ed è stata piuttosto prevalente opinione, finora, che per le gare di corsa di fondo e gran fondo fosse ritenuto più predisposto il soggetto di età atletica matura e magari di certe determinate provenienze geografiche, ma come si è potuto notare in questi ultimi tempi (i recentissimi Campionati Mondiali di Stoccarda '93 ne hanno dato ampia conferma) i fatti hanno dimostrato che nelle gare in questione, e la maratona ne è anche interessata, atleti giovani e molto giovani - ed anche di varia provenienza geografica - si sono trovati ad avere praticamente le stesse *chances* tecniche degli altri per esprimersi pure ai massimi livelli (tra l'altro, il fenomeno mostra chiara tendenza ad una crescente sua accentuazione). In quest'ottica di osservazione, sono assai sintomatiche le ancora più recenti performances raggiunte, nel settore femminile, da diverse giovanissime atlete cinesi sulle lunghe distanze, sebbene tale fenomeno si presti, proprio per la sua dirompente eccezionalità, ad innescare ovviamente polemiche su certe precise connotazioni tecniche circa il grado della loro naturale attendibilità.
C'è da sottolineare, infine, che alla sopracitata data di aggiornamento il fenomeno a prima vista non era poi così evidente, anzi appariva abbastanza rischioso azzardare una ipotesi del genere, tuttavia la reale consistenza dei suoi elementi costitutivi, sebbene piuttosto nascosta, ha ugualmente trovato modo di poter esprimere e puntualizzare, a livello strettamente analitico, le effettive caratteristiche tecniche proprie delle iniziali ipotesi di studio effettuate in tale direzione.
A questo punto, potrebbero risultare certamente interessanti le conoscenze dei vari aspetti storico-statistici anche della gara di maratona del settore femminile, in quanto una eventuale ricerca in tal senso, seppure non pienamente conforme, sarebbe senz'altro in grado di produrre, con le relative proprie acquisizioni, utili ed importanti elementi rappresentativi del fenomeno, specie nel campo dell'analisi comparata di regressione e correlazione.
Sempre a proposito di gare di fondo e gran fondo non sarebbe sicuramente meno interessante una trattazione anche per quanto riguarda la gara maschile e femminile di marcia dei 10 km. (su pista) e quella maschile dei 50 km. (su strada), specialità che non smuovono di certo adeguati consensi e favori da parte delle grandi masse del pubblico sportivo, ma che tuttavia, specie nell'ambito del settore maschile, hanno da sempre gratificato l'atletica italiana di numerosi e prestigiosi riconoscimenti ai più alti vertici dell'agonismo internazionale.
Di ciò se ne potrà eventualmente riparlare in un auspicabile prossimo futuro. Necessita comunque precisare che, risultando analisi del genere direttamente assoggettate alla inderogabile disponibilità di dati statistici di ampia ed opportuna consistenza, rimane solo da augurarsi che quanto prima possibile se ne possa creare, in questa sede, l'occasione giusta.

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

BIBLIOGRAFIA

1) GIARDINA BASILIO: *Statistica per aziende e ricercatori.* Ediz. Franco Angeli, Milano 1987.
2) RICCI FAUSTO: *Statistica ed elaborazioni statistiche delle informazioni.* Ediz. Zanichelli, Bologna 1989.
3) SALVATORE DOMINICK: *Statistica ed econometria.* Ediz. ETAS Libri, Milano 1985.
4) SPIEGEL MURRAY R.: *Probabilità e statistica.* Ediz. ETAS Liberi, Milano 1976.
5) VIANELLI SILVIO: *Metodologia descrittiva e della ricerca empirica.* Ediz. Calderini, Bologna 1978.