Sulla media e la varianza di un campione (*)

di G. Pompili
Roma

1. – Posizione del problema

Sia M una massa finita formata da H elementi $e_1, e_2, \cdots, e_H$ a cui competono ordinatamente certi valori $x_1, x_2, \cdots, x_H$ di un dato carattere quantitativo X.

Alla massa M resta in tal modo associata una variable casuale $\mathfrak{X}$ che assume i valori $x_1, x_2, \cdots, x_H$ — non necessariamente tutti distinti — ciascuno con la stessa probabillità $\frac{1}{H}$. Se poi h_1 degli H valori x_i coincidono in un certo valore x'_1, h_2 coincidono in $x'_2, \cdots, h_k$ $(k \leqslant H)$ coincidono nel valore x'_k, potremo anche dire che la v. c. (sigla di «variabile casuale») $\mathfrak{X}$ assume i valori $x'_1, x'_2, \cdots, x'_k$ con probabilità $p_1 = \frac{h_1}{H}$, $p_2 = \frac{h_2}{H}, \cdots, p_k = \frac{h_k}{H}$.

Secondo un simbolismo da tempo in uso indicheremo con m_r il momento r-mo della v. c. $\mathfrak{X}$:

$$m_r = M(\mathfrak{X}^r) = \frac{1}{H}\sum_1^H {}_i\, x_i^r \qquad (r = 1, 2, \cdots);$$

tra questi momenti è particolarmente importante quello del primo ordine m_1, che è comunemente chiamato «media».

Indicheremo poi con $\overline{m}_r$ il momento r-mo rispetto alla media m_1:

$$\overline{m}_r = M([\mathfrak{X} - m_1]^r) = \frac{1}{H}\sum_1^H {}_i\,(x_i - m_1)^r \quad (r = 2, 3, \cdots);$$

tra questi momenti è particolarmente notevole quello di second'ordine che indicharemo con μ_x^2 e chiameremo, secondo una terminologia ormai diffusa, *varianza* della v. c. $\mathfrak{X}$.

Ciò premesso immaginiano di avere un campione della nostra v. c. costruito scegliendo a caso N elementi della massa M, con un procedimento rispetto a cui tutti gli elementi della massa si trovino nelle stessi condizioni, e raccogliendo poi i valori che il carattere X assume in corrispondenza agli elementi scelti.

Sopra questo campione, con qualche avvertenza che chiariremo nel seguito, si definiscono due parametri ${}_cm_1$ e ${}_c\mu^2$, chiamati rispettivamente *media* e *varianza* del campione, i quali, al variare del campione stesso, descrivono due v. c. che è interessante studiare, determinandone in particolare, la media e la varianza.

In vista di tale problema è opportuno precisar meglio il metodo di scelta che, come è noto, può essere di due tipi fondamentali:

I) *estrazione a caso con ripetizione* (schema bernoulliano) in cui si rimette di volta in volta l'elemento estratto nella massa,

II) *estrazione a caso senza ripetizione* (schema dell'estrazione in blocco) in cui l'elemento estratto non viene ricollocato nella massa.

Il problema che ci siamo prospettato nelle righe precedenti è già stato risolto sia nello schema bernoulliano che in quello dell'estrazione in blocco (1).

Tuttavia in sede di sistemazione espositiva dei risultati si può raggiungere una notevole semplificazione basandosi sull'impiego sistematico della

(*) Ricevuto nell' Ottobre 1951.

(1) Per le formule relative allo schema di BERNOULLI ved. un qualunque buon trattato di «Statistica» o «Calcolo delle probabilità»; per quelle relative allo schema dell'estrazione in blocco ved.: AL. A. TSCHUPROW: *On the mathematical expectation of the moments of frequency distributions in the case of correlated observations*, Metron, II, 1923; J. S. NEYMAN: *Contribution to the Theory of Small Samples drawn from a finite Population*, Biometrica, XVII, 1925; A. E. R. CHURCH: *On the means and squared standard-deviations of small samples from any population*, Biometrica, XVIII, 1926.

v. c. delle prove ripetute associata allo schema d'estrazione, il che rende, per così dire, automatico il procedimento per ricavare le formule.

Ho già esposto completamente questo metodo d'impostare e risolvere il problema nel caso bernoulliano dell'estrazione a caso con ripetizione [1] ; in questa Nota impiego lo stesso metodo anche nel caso dell'estrazione in blocco, cosa che non mi ero ancora deciso a fare non tanto per difficoltà concettuali — perchè mi sarebbe bastato seguire pedissequamente quanto avevo già fatto per il caso bernoulliano — quanto per una certa laboriosità di calcoli veramente scoraggiante.

Ma esigenze di simmetria e amore di completezza mi hanno spinto ad introdurre anche questo argomento nel mio corso di «Calcolo delle probabilità» inducendomi così ad affrontar la fatica che, con qualche opportuno accorgimento che il Lettore noterà, ho cercato di rendere minima.

Aggiungo che i campioni ottenuti con scelta a caso senza ripetizione sono già statti studiati sistematicamente da G. Gini [2] il quale ha stabilito risultati fondamentali sopra le medie combinatorie potenziate di tali campioni.

2. — La v. c. dell'estrazione in blocco

I risultati delle prove ripetute N volte, con estrazione a caso senza ripetizione, nell'ipotese di k alternative, descrivono, al variare del grupo di N prove, la v. c. multipla $(\mathfrak{D}_1, \mathfrak{D}_2, \cdots, \mathfrak{D}_k)$, detta «dell'estrazione in blocco», che è $(k-1)$—pla e non k—pla, come potrebbe sembrare, perchè le k v. c. $\mathfrak{D}_i$ sodisfano alla relazione identica:

$$\mathfrak{D}_1 + \mathfrak{D}_2 + \cdots + \mathfrak{D}_k = N. \tag{2.1}$$

I momenti fattoriali di tali v. c. sono noti [3] :

$$\begin{aligned} m_{(r_1, \ldots, r_k)} &= M\,[\mathfrak{D}_1(\mathfrak{D}_1 - 1) \cdots \\ &\cdots (\mathfrak{D}_1 - r_1 + 1)\,\mathfrak{D}_2(\mathfrak{D}_2 - 1) \cdots (\mathfrak{D}_2 - r_2 + 1) \cdots \\ &\cdots \mathfrak{D}_k(\mathfrak{D}_k - 1) \cdots (\mathfrak{D}_k - r_k + 1)] = \\ &= \frac{N(N-1)\cdots(N-r+1)}{H(H-1)\cdots(H-r+1)}\, h_1(h_1 - 1) \cdots \\ &\cdots (h_1 - r_1 + 1)\, h_2(h_2 - 1) \cdots (h_2 - r_2 + 1) \cdots \\ &\cdots h_k(h_k - 1) \cdots (h_k - r_k + 1) \end{aligned} \tag{2.2}$$

dove si è posto $r = r_1 + r_2 \cdots + r_k$.

Mediante i momenti fattoriali si possono poi calcolare subito i momenti ordinari e quelli rispetto alla media, ricorrendo a formule assai facilmente ricavabili e che qui, per comodità del Lettore, mi limito a riportare per i momenti fino al quart'ordine.

Diamo anzitutto le formule che legano i momenti ordinari a quelli fattoriali:

$$\left\{\begin{array}{lllll} m_1 & = m_{(1)} & & & \\ m_2 & = m_{(2)} & + m_{(1)} & & \\ m_{1,1} & = m_{(1,1)} & & & \\ m_3 & = m_{(3)} & + 3\,m_{(2)} & + m_{(1)} & \\ m_{2,1} & = m_{(2,1)} & + m_{(1,1)} & & \\ m_{1,1,1} & = m_{(1,1,1)} & & & \\ m_4 & = m_{(4)} & + 6\,m_{(3)} & + 7\,m_{(2)} & + m_{(1)} \\ m_{3,1} & = m_{(3,1)} & + 3\,m_{(2,1)} & + m_{(1,1)} & \\ m_{2,2} & = m_{(2,2)} & + m_{(2,1)} & + m_{(1,2)} & + m_{(1,1)} \\ m_{2,1,1} & = m_{(2,1,1)} & + m_{(1,1,1)} & & \\ m_{1,1,1,1} & = m_{(1,1,1,1)} & & & \\ \cdots\cdots & & & & \end{array}\right. \tag{2.3}$$

I momenti rispetto alla media sono poi legati ai momenti rispetto all'origine dalle formule seguenti:

$$\left\{\begin{array}{ll} \overline{m}_2 & = m_2 - m_1^2 \\ \overline{m}_{1,1} & = m_{1,1} - m_{1,0}\,m_{0,1} \\ \overline{m}_3 & = m_3 - 3m_1 m_2 + 2m_1^3 \\ \overline{m}_{2,1} & = m_{2,1} - m_{2,0}m_{0,1} + 2m_{1,0}^2 m_{0,1} - 2m_{1,0}m_{1,1} \\ \overline{m}_{1,1,1} & = m_{1,1,1} - m_{1,1,0}\,m_{0,0,1} + \\ & \quad - m_{1,0,1}\,m_{0,1,0} - m_{0,1,1}\,m_{0,0,1} + \\ & \quad + 2m_{1,0,0}\,m_{0,1,0}\,m_{0,0,1} \\ \overline{m}_4 & = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4 \\ \overline{m}_{3,1} & = m_{3,1} - 3\,m_{2,1}\,m_{1,0} + 3\,m_{1,1}\,m_{1,0}^2 + \\ & \quad - m_{3,0}m_{0,1} + 3m_{2,0}m_{1,0}m_{0,1} - 3m_{1,0}^3 m_{0,1} \\ \overline{m}_{2,2} & = m_{2,2} - 2m_{2,1}\,m_{0,1} - 2m_{1,2}\,m_{1,0} + \\ & \quad + 4m_{1,1}\,m_{1,0}\,m_{0,1} + m_{2,0}\,m_{0,1}^2 + \\ & \quad + m_{0,2}\,m_{1,0}^2 - 3m_{1,0}^2\,m_{0,1}^2 \\ \overline{m}_{2,1,1} & = m_{2,1,1} - 2\,m_{1,1,1}\,m_{1,0,0} + \\ & \quad - m_{2,1,0}\,m_{0,0,1} - m_{2,0,1}\,m_{0,1,0} + \\ & \quad + 2\,m_{1,1,0}\,m_{1,0,0}\,m_{0,0,1} + \\ & \quad + 2m_{1,0,1}\,m_{1,0,0}\,m_{0,1,0} + m_{1,0,0}^2\,m_{0,1,1} + \\ & \quad + m_{2,0,0}\,m_{0,0,1}\,m_{0,1,0} + \\ & \quad - 3\,m_{1,0,0}^2\,m_{0,1,0}\,m_{0,0,1} \\ \overline{m}_{1,1,1,1} & = m_{1,1,1,1} - m_{1,1,1,0}\,m_{0,0,0,1} + \\ & \quad - m_{1,1,0,1}\,m_{0,0,1,0} - m_{1,0,1,1}\,m_{0,1,0,0} + \\ & \quad - m_{0,1,1,1}\,m_{1,0,0,0} + \\ & \quad + m_{1,1,0,0}\,m_{0,0,1,0}\,m_{0,0,0,1} + \\ & \quad + m_{1,0,1,0}\,m_{0,1,0,0}\,m_{0,0,0,1} + \\ & \quad + m_{1,0,0,1}\,m_{0,1,0,0}\,m_{0,0,1,0} + \\ & \quad + m_{0,1,1,0}\,m_{1,0,0,0}\,m_{0,0,0,1} + \\ & \quad + m_{0,1,0,1}\,m_{1,0,0,0}\,m_{0,0,1,0} + \\ & \quad + m_{0,0,1,1}\,m_{1,0,0,0}\,m_{0,1,0,0} + \\ & \quad - 3m_{1,0,0,0}\,m_{0,1,0,0}\,m_{0,0,1,0}\,m_{0,0,0,1} \\ \cdots\cdots & \end{array}\right. \tag{2.4}$$

---

[1] Ved.: G. Pompilj: *Complementi di calcolo delle probabilità*, Veschi, Roma, 1948; pagg. 148 e seg..

[2] Ved.: C. Gini: *Le medie dei campioni*, Metron, XV, 1950, G. Pompilj: *Sulle medie combinatorie potenziate dei campioni*, Rend. Sem. Mat. di Padova. XVIII. 1948.

[3] Ved.: G. Pompilj: *Complementi di Calcolo delle Probabilità*, Veschi, Roma, 1948, pag. 135.

In particolare dalla prima delle (2.3) si ricava, in base alla (2.2), che la media della v. c. $\mathfrak{O}_i$ è Np_i; in simboli:

$$(2.5) \qquad M(\mathfrak{O}_i) = Np_i\,.$$

Possiamo quindi definire le v. c. scarto:

$$(2.6) \qquad \mathfrak{N}_i = \mathfrak{O}_i - Np_i \qquad (i=1,2,\cdots,k)$$

che soddisfano alla relazione lineare identica:

$$(2.7) \qquad \mathfrak{N}_1 + \mathfrak{N}_2 + \cdots + \mathfrak{N}_k = 0\,.$$

Conviene dire subito che il calcolo dei momenti della v. c. $(\mathfrak{N}_1, \mathfrak{N}_2, \cdots, \mathfrak{N}_k)$, cioè dei momenti rispetto alla media della v. c. dell'estrazione in blocco, calcolo da farsi attraverso le (2.3) e le (2.4), è però tuttavia laborioso; ciò mi ha indotto a condurre avanti i calcoli in un caso particolare, tuttavia completamente sufficiente per la soluzione del problema che ci siamo posto; il caso particolare in questione corrisponde a $k=H$ e quindi $h_1=h_2=\cdots=h_H=1$.

In queste condizioni il momento fattoriale $m_{(r_1,r_2,\ldots,r_H)}$ è nullo, a meno che non sia $r_i\,(i=1,2,\cdots,H)$ eguale a 0 a 1, nel qual caso, se r delle r_i sono eguali ad 1 e le rimanenti $H-r$ sono nulle, in base alla (2.2), si ha che il valore del momento fattoriale è dato da:

$$(2.8) \qquad \frac{N(N-1)\cdots(N-r+1)}{H(H-1)\cdots(H-r+1)}\,.$$

In queste condizioni si ottengono subito i momenti della v. c. dell'estrazione in blocco:

$$(2.9) \quad \begin{cases} M(\mathfrak{O}_i^r) = \dfrac{N}{H} \\[2ex] M(\mathfrak{O}_i^r\,\mathfrak{O}_j^s) = \dfrac{N(N-1)}{H(H-1)} \\[2ex] M(\mathfrak{O}_i^r\,\mathfrak{O}_j^s\,\mathfrak{O}_w^t) = \dfrac{N(N-1)(N-2)}{H(H-1)(H-2)} \\[2ex] M(\mathfrak{O}_i^r\,\mathfrak{O}_j^s\,\mathfrak{O}_w^t\,\mathfrak{O}_z^u) = \dfrac{N(N-1)(N-2)(N-3)}{H(H-1)(H-2)(H-3)}\,. \\[2ex] \cdots\cdots \end{cases}$$

Approfittando ora delle (2.4) e delle (2.9) possiamo scrivere le espressioni dei momenti rispetto alla media. Prima però converrà fare due posizioni che ci permetteranno una scrittura più compatta:

$$(2.10) \qquad \alpha = \frac{N}{H}\,, \qquad \beta = 1-\alpha = \frac{H-N}{H}$$

Dopo di che i momenti rispetto alla media della v. c. dell'estrazione in locco assumono la forma seguente:

$$(2.11) \quad \begin{cases} M(\mathfrak{N}_i^2) = \alpha\beta \\[1ex] M(\mathfrak{N}_i\,\mathfrak{N}_j) = -\dfrac{\alpha\beta}{H-1} \\[1ex] M\,\mathfrak{N}_i^3) = \alpha\beta(\beta-\alpha) \\[1ex] M(\mathfrak{N}_j^2\,\mathfrak{N}_j) = -\dfrac{\alpha\beta(\beta-\alpha)}{H-1} \\[1ex] M(\mathfrak{N}_i\,\mathfrak{N}_j\,\mathfrak{N}_w) = 2\dfrac{\alpha\beta(\beta-\alpha)}{(H-1)(H-2)} \\[1ex] M(\mathfrak{N}_i^4) = \alpha\beta(1-3\alpha\beta) \\[1ex] M(\mathfrak{N}_i^3\,\mathfrak{N}_j) = -\dfrac{\alpha\beta(1-3\alpha\beta)}{H-1} \\[1ex] M(\mathfrak{N}_i^2\,\mathfrak{N}_j^2) = \alpha\beta\left[\alpha\beta - \dfrac{(\beta-\alpha)^2}{H-1}\right] \\[1ex] M(\mathfrak{N}_i^2\,\mathfrak{N}_j\,\mathfrak{N}_w) = \dfrac{\alpha\beta}{H-1}\left[2\dfrac{(\beta-\alpha)^2}{H-2} + \beta^2\right] \\[1ex] M(\mathfrak{N}_i^2\,\mathfrak{N}_j\,\mathfrak{N}_w\,\mathfrak{N}_z) = 3\dfrac{\alpha\beta\{(N-1)\beta+(\beta-\alpha)(2\alpha-\beta)\}}{(H-1)(H-2)(H-3)} \\[1ex] \cdots \end{cases}$$

3. — Le v. c. «momento»

Per poter approfittare, nei successivi calcoli, delle formule (2.11), stabilite alla fine del numero precedente, converrà dire, riprendendo il discorso del n. 1, che negli H elementi della massa $\mathbf{M}$ il carattere X assume ordinatamente i valori $x_1, x_2, \cdots, x_H$.

In tali condizioni i momenti della corrispondente v. c. $\mathfrak{X}$ sono dati, come si è visto da:

$$(3.1) \qquad m_r = M(\mathfrak{X}^r) = \frac{1}{H}\sum_1^H{}_i\, x_i^r \qquad (r=1,2,\cdots)$$

e in modo del tutto analogo si definiscono i momenti rispetto alla media:

$$(3.2) \quad \overline{m}_r = M([\mathfrak{X}-m_1]^r) = \sum_1^H{}_i \frac{1}{H}(x_i-m_1)^r \quad (r=2,3,\cdots)$$

Scegliendo a caso senza ripetizione, come già si è detto, N elementi della massa $\mathbf{M}$ si otterranno le N determinazioni seguenti del carattere X: x_{j_1}, $x_{j_2}, \cdots, x_{j_N}$ le quali formeranno un campione della v. c. $\mathfrak{X}$, campione ottenuto con un'estrazione in blocco.

In termini della corrispondente v. c. dell'estrazione in blocco possiamo dire che la determinazione $(o_1, o_2, \cdots, o_H)$, relativa al nostro campione è for mata da N numeri 1 e $H-N$ numeri 0, così distribuiti:

$$o_i = \begin{cases} 1 & se \ \ i = j_1, j_2, \cdots, j_N \\ 0 & negli \ altri \ casi. \end{cases} \tag{3.3}$$

*

Come momento r-mo del nostro campione possiamo prendere la quantità:

$$_c m_r = \frac{1}{N} \sum_s^N{}_1 \, x_{j_s}^r \tag{3.4}$$

che in base alle (3.3), si può anche scrivere in quest' altra forma

$$_c m_r = \frac{1}{N} \sum_i^N{}_1 \, x_i^r \, o_i \tag{3.5}$$

che è per noi assai comoda perchè permette di scrivere subito l'espressione esplicita della v. c. $\mathfrak{M}_r$ descritta da $_c m_r$ al variare del campione:

$$\mathfrak{M}_r = \frac{1}{N} \sum_i^H{}_1 \, x_i^r \, \mathfrak{O}_i \, . \tag{3.6}$$

*

Il valor medio di questa v. c. $\mathfrak{M}_r$, *descritta dal momento* r-*mo* $_c\mathrm{m}_r$ *del campione, coincide con il momento* r-*mo* m_r *della v. c.* $\mathfrak{X}$ *di partenza.*

La dimostrazione è immediata; basta infatti ricordare che, in base alla prima delle (2.9) è:

$$M(\mathfrak{O}_i) = \frac{N}{H},$$

dopo di che si ha subito:

$$M(\mathfrak{M}^r) = \frac{1}{N} \sum_i^H{}_1 \, x_i^r \, M(\mathfrak{O}_i) = \frac{1}{H} \sum_i^H{}_1 \, x_i^r = m_r \tag{3.7}$$

c. v. d.

*

Possiamo infine definire la v. c. scarto:

$$\mathfrak{S}_r = \mathfrak{M}_r - m_r = \frac{1}{N} \sum_i^H{}_1 \, x_i^r \, \mathfrak{N}_i \tag{3.8}$$

di cui ci interessano i successivi momenti.

Questi momenti si calcolano in modo concettualmente assai semplice secondo un schema che ho già avuto occasione di esporre [1], ma che tuttavia nel nostro caso porta a calcoli piuttosto laboriosi; per il qual motivo mi limiterò a ricorrere a tale schema generale solo per il calcolo di alcuni momenti, mentre per altri, che pure interessano, mi servirò di particolari accorgimenti.

I calcoli che seguono mi sembrano abbastanza chiari; e pertanto mi limiterò a riportarli senza una parola di spiegazione.

$$\left\{ \begin{aligned} M(\mathfrak{S}_r \mathfrak{N}_j) &= \frac{1}{N} \Bigg[\sum_i^{i \neq j} x_i^r \, M(\mathfrak{N}_i \mathfrak{N}_j) + \\ &\quad + x_j^r \, M(\mathfrak{N}_j^2) \Bigg] = \frac{\beta}{H-1} (x_j^r - m_r) \\ M(\mathfrak{S}_r \mathfrak{N}_j \mathfrak{N}_t) &= \frac{1}{N} \Bigg[\sum_i^{i \neq j,t} x_i^r \, M(\mathfrak{N}_i \mathfrak{N}_j \mathfrak{N}_t) + \\ &\quad + x_j^r \, M(\mathfrak{N}_j^2 \mathfrak{N}_t) + x_t^r \, M(\mathfrak{N}_j \mathfrak{N}_t^2) \Bigg] = \\ &= \frac{\beta(\beta-\alpha)}{(H-1)(H-2)} (2m_r - x_i^r - x_t^r) \\ M(\mathfrak{S}_r \mathfrak{N}_j^2) &= \frac{1}{N} \Bigg[\sum_i^{i \neq j} x_i^r \, M(\mathfrak{N}_i \mathfrak{N}_j^2) + \\ &\quad + x_j^r \, M(\mathfrak{N}_j^3) \Bigg] = \frac{\beta(\beta-\alpha)}{H-1} (x_j^r - m_r) \\ M(\mathfrak{S}_r^2 \mathfrak{N}_j) &= \frac{1}{N} \Bigg[\sum_i^{i \neq j} x_i^r \, M(\mathfrak{S}_r \mathfrak{N}_i \mathfrak{N}_j) + \\ &\quad + x_j^r \, M(\mathfrak{S}_r \mathfrak{N}_j^2) \Bigg] = \frac{\beta(\beta-\alpha)}{\alpha(H-1)(H-2)} \cdot \\ &\quad \cdot [x_j^{2r} - m_{2r} - 2 m_r (x_j^r - m_r)] \, . \end{aligned} \right. \tag{3.9}$$

Mediante questi valori medi si calcolano poi subito i due momenti seguenti:

$$\left\{ \begin{aligned} M(\mathfrak{S}_r^2) &= \frac{1}{N} \sum_i^H{}_1 \, x_i^r \, M(\mathfrak{S}_r \mathfrak{N}_i) = \\ &= \frac{H-N}{H-1} \cdot \frac{1}{N} (m_{2r} - m_r^2) \\ M(\mathfrak{S}_r^2 \mathfrak{S}_u) &= \frac{1}{N} \sum_i^H{}_1 \, x^u \, M(\mathfrak{S}_r^2 \mathfrak{N}_i) = \\ &= \frac{(H-N)(H-2N)}{(H-1)(H-2)} \cdot \frac{1}{N^2} [m_{2r+u} + \\ &\quad - 2m_{r+u} m_r - m_{2r} m_u + 2 m_r^2 m_u] \, . \end{aligned} \right. \tag{3.10}$$

Un altro momento che c'interessa calcolare è $M(\mathfrak{S}_1^4)$; ma seguendo la strada fin'ora percorsa andremmo incontro a calcoli piuttosto laboriosi; ricorreremo perciò ad altro ordine di considerazione, e, per rendere i calcoli più semplici, li eseguireno nella ipotesi *non restrittiva* che sia $m_1 = 0$.

[1] G. Pompilj: *Sulla media geometrica e sopra un indice di mutabilità calcolati mediante un campione*, Mem. Soc. Italiana delle Scienze, s. III, 1947.

Indichiamo con $\mathfrak{X}_1, \mathfrak{X}_2, \cdots, \mathfrak{X}_N$ le v. c. descritte dalle determinazioni assunte dalla v. c. $\mathfrak{X}$ nelle successive N prove con le quali viene construito il campione.

Queste N v. c. non sono certo indipendenti perchè, eseguendosi le N prove secondo lo schema dell'estrazione a caso senza ripetizione, il risultato di una prova è ovviamente influenzato dai risultati dele prove precedenti; inoltre esse sono formalmente eguali alla v. c. $\mathfrak{X}$, intendendosi con questo che hanno in comune con tale v. c. la funzione di ripartizione.

I momenti del quart'ordine della v. c. $(\mathfrak{X}_1, \mathfrak{X}_2, \cdots, \mathfrak{X}_N)$ si calcolano abbastanza facilmente e il calcolo è ancor più facilitato dall'ipotesi posta che sia:

$$m_1 = \frac{1}{H} \sum_1^H{}_i \, x_i = 0 . \tag{3.11}$$

Anzitutto possiamo subito scrivere:

$$M(\mathfrak{X}_r^4) = M(\mathfrak{X}^4) = \overline{m}_4 \tag{3.12}$$

mentre per il calcolo dei momenti misti occorre tener presente che due v. c. $\mathfrak{X}_r$ con indice in basso diverso non possono mai assumere contemporaneamente lo stesso valore x_i; avremo pertanto:

$$(3.13) \left\{ \begin{aligned}
M(\mathfrak{X}_r^3 \mathfrak{X}_s) &= \frac{1}{H(H-1)} \sum_1^H{}_i \sum^{j \neq i}{}_j \, x_j^3 x_i = \\
&= \frac{1}{H(H-1)} \sum_1^H{}_i \, x_i (H\overline{m}_3 - x_i^3) = - \frac{\overline{m}_4}{H-1} \\
M(\mathfrak{X}_r^2 \mathfrak{X}_s^2) &= \frac{1}{H(H-1)} \sum_1^H{}_i \sum^{j \neq i}{}_j \, x_j^2 x_i^2 = \\
&= \frac{1}{H(H-1)} \sum_1^H{}_i \, x_i^2 (H\overline{m}_2 - x_i^2) = \frac{H\overline{m}_2^2 - \overline{m}_4}{H-1} \\
M(\mathfrak{X}_r^2 \mathfrak{X}_s \mathfrak{X}_t) &= \frac{1}{H(H-1)(H-2)} \sum_1^H{}_i \sum^{j \neq i}{}_j \sum^{w \neq i,j}{}_w \\
& x_w^2 x_j x_i = \frac{1}{H(H-1)(H-2)} \sum_1^H{}_i \sum^{j \neq i}{}_j x_j x_i \, \cdot \\
& (H\overline{m}_2 - x_j^2 - x_i^2) = \frac{1}{H(H-1)(H-2)} \cdot \\
& \cdot \sum^H{}_i \, x_i (-H\overline{m}_2 x_i + 2x_i^3 - H\overline{m}_3) = \\
&= - \frac{H\overline{m}_2^2 - 2\overline{m}_4}{(H-1)(H-2)} \\
M(\mathfrak{X}_r \mathfrak{X}_s \mathfrak{X}_t \mathfrak{X}_u) &= \frac{2}{H(H-1)(H-2)(H-3)} \cdot \\
& \cdot \sum^H{}_i \sum^{j \neq i}{}_j \sum^{w \neq i,j}{}_w \sum^{z \neq i,j,w}{}_z \, x_z x_w x_i x_j = \\
&= \frac{1}{H(H-1)(H-2)(H-3)} \sum_1^H{}_i \sum^{j \neq i}{}_j \sum^{w \neq i,j}{}_w \\
& x_w x_j x_i (-x_w - x_i - x_j) = \\
&= \frac{1}{H(H-1)(H-2)(H-3)} \sum_1^H{}_i \sum^{j \neq i}{}_j \\
& x_j x_i (2x_j^2 + 2x_j x_i + 2x_i - H\overline{m}_2) = \\
&= \frac{1}{H(H-1)(H-2)(H-3)} \sum^H{}_i \, x_i (-6x_i^3 + \\
& + 3Hx_i \overline{m}_2 + 2\overline{m}_3) = 3 \frac{H\overline{m}_2^2 - 2\overline{m}_4}{(H-1)(H-2)(H-3)} .
\end{aligned} \right.$$

Ciò posto si osservi che la v. c. scarto $\mathfrak{S}_1$, nelle nostre ipotesi, non è altro che la media aritmetica delle N v. c. $\mathfrak{X}_i$, cioè:

$$\mathfrak{S}_1 = \frac{1}{N} (\mathfrak{X}_1 + \mathfrak{X}_2 + \cdots + \mathfrak{X}_N) \tag{3.14}$$

di modo che si può subito scrivere:

$$\begin{aligned}
\mathfrak{S}_1^4 = \frac{1}{N_2^4} [\Sigma \mathfrak{X}_r^4 + 4 \Sigma \mathfrak{X}_r^3 \mathfrak{X}_s + 6 \Sigma \mathfrak{X}_r^2 \mathfrak{X}_s^2 + \\
+ 12 \Sigma \mathfrak{X}_r^2 \mathfrak{X}_s \mathfrak{X}_w + 24 \Sigma \mathfrak{X}_r \mathfrak{X}_s \mathfrak{X}_w \mathfrak{X}_z]
\end{aligned} \tag{3.15}$$

dove i 5 sommatori a secondo membro comprendono rispettivamente $N, 2\binom{N}{2}, \binom{N}{2}, 3\binom{N}{3}, \binom{N}{4}$ termini.

Passando infine, nella (3.12), ai valori medi, possiamo scrivere:

$$\begin{aligned}
M(\mathfrak{S}_1^4) = \frac{1}{N^4} \Bigg[N M(\mathfrak{X}_r^4) + 8\binom{N}{2} M(\mathfrak{X}_r^3 \mathfrak{X}_s) + \\
+ 6\binom{N}{2} M(\mathfrak{X}_r^2 \mathfrak{X}_s^2) + 36\binom{N}{3} M(\mathfrak{X}_r^2 \mathfrak{X}_s \mathfrak{X}_w) + \\
+ 24\binom{N}{4} M(\mathfrak{X}_r \mathfrak{X}_s \mathfrak{X}_w \mathfrak{X}_z) \Bigg] ,
\end{aligned} \tag{3.16}$$

da cui ricordando le (3.11), si ottiene infine:

$$\begin{aligned}
M(\mathfrak{S}_1^4) = \frac{H-N}{H-1} \cdot \frac{1}{N^3} \Bigg[\overline{m}_4 + \\
+ 3 \frac{(H-N-1)(N-1)}{(H-2)(H-3)} (H\overline{m}_2^2 - 2\overline{m}_4) \Bigg] .
\end{aligned} \tag{3.17}$$

*

Chiuderemo questo numero così pieno di formule elencando solo quelle che c'interesseranno per il seguito, calcolate nell'ipotesi, che per i nostri scopi non resulterà restrittiva, che sia $m_1 = 0$:

$$(3.18)\quad \begin{cases} M(\mathfrak{S}_1^2) = \dfrac{H-N}{H-1}\cdot\dfrac{\overline{m}_2}{N} \\ M(\mathfrak{S}_2^2) = \dfrac{H-N}{H-1}\cdot\dfrac{\overline{m}_4-\overline{m}_2^2}{N} \\ M(\mathfrak{S}_1^2\,\mathfrak{S}_2) = \dfrac{H-N}{H-1}\cdot\dfrac{H-2N}{H-2}\cdot\dfrac{\overline{m}_4-\overline{m}_2^2}{N^2} \\ M(\mathfrak{S}_1^4) = \dfrac{H-N}{H-1}\cdot\dfrac{1}{N^3}\left[\overline{m}_4 + \right. \\ \qquad\left. + 3\dfrac{(H-N-1)(N-1)}{(H-2)\,(H-3)}(H\overline{m}_2^2-\overline{m}_4)\right]. \end{cases}$$

4. — La v. c. «media»

Della v. c. $\mathfrak{M}_1$ descritta dalla media ${}_cm_1$ del campione ci sbrigheremo in due parole. Infatti, per la (3.7), si può subito scrivere:

$$(4.1)\qquad M\,\mathfrak{M}_1 = m_1$$

e, in base alla prima delle (3.18) si può dare l'espressione della varianza:

$$(4.2)\qquad \mu^2_{\mathfrak{M}_1} = M(\mathfrak{S}_1^2) = \frac{H-N}{H-1}\,\frac{\mu^2_{\mathfrak{X}}}{N}.$$

5. — La v. c. «varianza»

Nel caso di un campione ottenuto secondo lo schema bernoulliano, cioè con estrazione a caso senza ripetizione, per definire la varianza del campione stesso si introduce un opportuno fattore di correzione in modo che la v. c. descritta da tale varianza abbia per media proprio la varianza della v. c. di partenza; ed effettivamente come varianza del campione si prende non già la media aritmetica dei quadrati degli scarti dalla media del campione medesimo, ma tale media aritmetica moltiplicata per il fattore $\dfrac{N}{N-1}$.

Anche nel caso dell'estrazione in blocco converrà definire la varianza ${}_c\mu^2$ del campione in modo che la media della v. c. descritta da tale parametro coincida con la varianza della v. c. di partenza.

A tale scopo bisogna porre:

$$(5.1)\qquad {}_c\mu^2 = \frac{H-1}{H}\cdot\frac{1}{N-1}\sum_1^N{}_s\,(x_{j_s} - {}_cm_1)^2$$

che si può anche scrivere in quest'altra maniera per i nostri scopi più comoda:

$$(5.2)\qquad {}_c\mu^2 = \frac{H-1}{H}\cdot\frac{1}{N-1}\sum_1^H{}_i\,(x_i - {}_cm_1)^2\,o_i\,.$$

La (5.2) permette di dare subito l'espressione esplicita della v. c. $\overline{\mathfrak{M}}_2$ descritta dalla varianza ${}_c\mu^2$ al variare del campione:

$$(5.3)\qquad \overline{\mathfrak{M}}_2 = \frac{H-1}{H}\cdot\frac{1}{N-1}\sum_1^H{}_i\,(x_i - \mathfrak{M}_1)^2\,\mathfrak{S}_i$$

che possiamo anche scrivere in quest'altra maniera:

$$(5.4)\qquad \overline{\mathfrak{M}}_2 = \frac{H-2}{H}\cdot\frac{N-1}{N}\,[\mu^2_{\mathfrak{X}} + \mathfrak{S}_2 - 2m_1\,\mathfrak{S}_1 - \mathfrak{S}_1^2]\,.$$

Vogliamo anzitutto controllare che la v. c. $\overline{\mathfrak{M}}_2$, sopra definita, è invariante di fronte ai cambiamenti d'origine, la cosa è praticamente pacifica trattandosi di varianza, ma è facile eseguire direttamente il controllo. Se infatti ai valori x_i si somma la quantità a ottenendo i nuovi valori $x'_i = x_i + a$ che appartengono alla v. c. $\mathfrak{X}' = \mathfrak{X} + a$ e se indichiamo con l'apice i momenti e le v. c. relativi a questa nuova v. c. si ha:

$$m'_1 = m_1 + a$$
$$\mu^2_{\mathfrak{X}'} = \mu^2_{\mathfrak{X}}$$
$$\mathfrak{S}'_1 = \mathfrak{S}_1$$
$$\mathfrak{S}'_2 = \mathfrak{S}_2 + 2a\,\mathfrak{S}_1$$

e quindi:

$$(5.5)\qquad \overline{\mathfrak{M}}'_2 = \frac{H-1}{H}\cdot\frac{N}{N-1}\,[\mu^2_{\mathfrak{X}'} + \mathfrak{S}'_2 - 2m'_1\,\mathfrak{S}'_1 + $$
$$+ \mathfrak{S}'^2_1] = \frac{H-1}{H}\cdot\frac{N}{N-1}\,[\mu^2_{\mathfrak{X}} + \mathfrak{S}_2 - 2m_1\,\mathfrak{S}_1 - \mathfrak{S}_1^2] = \overline{\mathfrak{M}}_2$$

c. v. d.

Dalla (5.5) segue in particolare che possiamo, senza perdere nulla in generalità, portare l'origine nella media o, quel che è la stessa cosa, supporre $m_1 = 0$ dopo di che la (5.4) si srcive piùs emplicemente così:

$$(5.6)\qquad \overline{\mathfrak{M}}_2 = \frac{H-1}{H}\cdot\frac{N}{N-1}\,[\mu^2_{\mathfrak{X}} + \mathfrak{S}_2 - \mathfrak{S}_1^2]\,.$$

*

Possiamo ora controllare, ricordando le (3.18), che la media della v. c. $\overline{\mathfrak{M}}_2$ coincide con la varianza della v. c. di partenza; si ha infatti:

$$(5.7)\qquad M(\overline{\mathfrak{M}}_2) = \frac{H-1}{H}\cdot\frac{N-1}{N}\,[\mu^2_{\mathfrak{X}} + M(\mathfrak{S}_2) -$$
$$- M(\mathfrak{S}_1^2) = \mu^2_{\mathfrak{X}}$$

c. v. d.

*

Sottraendo dalla v. c. $\overline{\mathfrak{M}}_2$ il suo valor medio $\mu^2_{\mathfrak{X}}$ otteniamo la v. c. scarto:

$$\bar{\mathfrak{S}}_2 = \bar{\mathfrak{M}}_2 - \mu^2_{\mathfrak{x}} = \frac{H-1}{H} \cdot \frac{N}{N-1} \left[\frac{H-N}{N(H-1)} \mu^2_{\mathfrak{x}} + {}_2\mathfrak{S} - \mathfrak{S}_1^2 \right] \tag{5.8}$$

di cui interessa in modo particolare il secondo momento, che ci dà la varianza della v. c. $\bar{\mathfrak{M}}_2$ stessa.

Per calcolare rapidamente tale secondo momento conviene, approfitando delle (3.18), determinare preventivamente i due valori medi seguenti:

$$\left\{\begin{aligned} M(\bar{\mathfrak{S}}_2 \mathfrak{S}_2) &= \frac{H-1}{H} \cdot \frac{N}{N-1} \cdot \left[\frac{H-N}{N(H-1)} \mu^2_{\mathfrak{x}} M(\mathfrak{S}_2) + M(\mathfrak{S}_2^2) + \right. \\ &\quad \left. - M(\mathfrak{S}_1^2 \mathfrak{S}_2) \right] = \frac{H-N}{H-2} \cdot \frac{1}{N} (\bar{m}_4 - \bar{m}_2^2) . \\ M(\mathfrak{S}_2 \mathfrak{S}_1^2) &= \frac{H-1}{H} \cdot \frac{N}{N-1} \cdot \left[\frac{H-N}{N(H-N)} \mu^2_{\mathfrak{x}} M(\mathfrak{S}_1^2) + M(\mathfrak{S}_2 \mathfrak{S}_1^2) + \right. \\ &\quad \left. + M \mathfrak{S}_1^4) \right] = \frac{H-N}{H} \cdot \frac{1}{N(N-1)} \left\{ \frac{H(N-1)}{(H-1)(H-2)} \bar{m}_2^2 + \right. \\ &\quad + \frac{H-2N}{H-2} \bar{m}_4 - \frac{1}{N} \left[\bar{m}_4 + \right. \\ &\quad \left. \left. + 3 \frac{(H-N-1)(N-1)}{(H-2)(H-3)} \cdot (H\bar{m}_2^2 - \bar{m}_4) \right] \right\} . \end{aligned}\right. \tag{5.9}$$

Mediante i precedenti valori medi (5.9) possiamo poi calcolare, nel modo più semplice, il secondo momento della v. c. scarto $\bar{\mathfrak{S}}_2$:

$$\begin{aligned} \mu^2_{\bar{\mathfrak{M}}_2} &= M(\bar{\mathfrak{S}}_2^2) = \frac{H-1}{H} \cdot \\ &\cdot \frac{N}{N-1} \left| \frac{H-N}{N(H-1)} \mu^2_{\mathfrak{x}} M(\bar{\mathfrak{S}}_2) + M(\bar{\mathfrak{S}}_2 \mathfrak{S}_2) + \right. \\ &\left. - (M \bar{\mathfrak{S}}_2 \mathfrak{S}_1^2) \right| = \frac{H-1}{H} \cdot \frac{H-N}{H-1} \left\{ \frac{1}{H-2} (\bar{m}_4 - \bar{m}_2^2) + \right. \\ &- \frac{1}{H(N-1)} \left[\frac{H(N-1)}{(H-1)(H-2)} \bar{m}_2^2 + \frac{H-2N}{H-2} \bar{m}_4 + \right. \\ &\left. \left. - \frac{1}{N} \left(\bar{m}_4 + 3 \frac{(H-N-1)(N-1)}{(H-2)(H-3)} (H\bar{m}_2^2 - \bar{m}_4) \right) \right] \right\} . \end{aligned} \tag{5.10}$$