SUI METODI PROPOSTI PER IL CALCOLO DELLA DIFFERENZA MEDIA

In: «*Metron*», Roma, 1931, Vol. IX, n. 1, pp. 47-52

METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

Vol. IX - N. 1. **1-II-1931.**

B. DE FINETTI

Sui metodi proposti per il calcolo della differenza media

ROMA
AMMINISTRAZIONE DEL «METRON»
R. UNIVERSITÀ — ISTITUTO DI STATISTICA

BRUNO DE FINETTI

Sui metodi proposti per il calcolo della differenza media

I.

Da quando, nel 1912, il GINI [6] (1) propose di introdurre nella metodologia statistica la *differenza media* come un *indice di variabilità* (2), mettendo in luce diversi vantaggi che essa presentava, molti Autori si sono occupati dei metodi per facilitarne il calcolo. La differenza delle notazioni e altre diversità inessenziali rendono però malagevole, almeno a prima vista, un confronto, e lasciano l'impressione che esista un gran numero di metodi diversi; era perciò opportuno esaminarli, mettendo in luce ciò che v'è in ciascuno di essi di distinto, ed è ciò che qui mi propongo.

II.

Per calcolare la differenza media si tratta, in sostanza, di calcolare la somma di tutte le differenze $|x_i - x_j|$; i vari metodi escogitati

(1) I richiami tra [ ] rimandano alla lista dei lavori citati.

(2) Per uno scopo affatto diverso, la differenza media era stata già considerata, molti anni addietro, dagli astronomi JORDAN, VON ANDRAE e HELMERT, in parecchi lavori di cui il GINI ebbe conoscenza durante la pubblicazione di [7], e citò ivi in nota a pag. 58. Detti AA. se n'erano occupati a proposito di una questione di calcolo delle probabilità relativa alla teoria degli errori di osservazione. Il JORDAN, in una prima memoria [1], proponeva di valutare la precisione di un insieme di osservazioni ugualmente attendibili basandosi sulle $\frac{n\,(n-1)}{2}$ differenze di esse due a due anzichè sulle n differenze dalla media, credendo di ottenere un vantaggio dell'accresciuto numero di dati. Ciò non è, evidentemente, che illusorio, i dati *indipendenti* essendo sempre soltanto $(n-1)$; ebbe così origine una discussione, iniziata da VON ANDRAE [2], cui ri-

si propongono appunto di apportare delle semplificazioni a questa somma.

I metodi essenzialmente distinti sono quattro.

1. Metodo GINI [6]

che si basa su una scomposizione delle differenze $|x_i - x_j|$ a seconda dell'intervallo che interessano, ma considerando soltanto gli intervalli fra due termini simmetrici della graduatoria.

Formula : $$\sum_{ij}^{n} |x_i - x_j| = \sum_i d_{i,n-i+1} |x_i - x_{n-i+1}|$$

ove $d_{i,n-i+1} = |n + 1 - 2i|$ è la distanza graduale.

2. Metodo DE FINETTI-PACIELLO [13]

che si basa su una scomposizione analoga, ma considerando i singoli intervalli successivi $|x_2 - x_1|$, $|x_3 - x_2|$, ..., $|x_n - x_{n-1}|$.

Formula : $$\sum_{ij} |x_i - x_j| = 2 \sum_{i=1}^{n-1} i\,(n - i)\,(x_{i+1} - x_i) .$$

3. Metodo GINI- CZUBER [7]-[8]

Questo metodo, proposto contemporaneamente (nel 1914) dai detti Autori, e ottenuto per vie del tutto diverse e indipendenti, è di gran lunga il più usato, e ad esso si riducono; con insignificanti modificazioni, quelli di molti Autori.

---

spose il JORDAN [3] provocando una seconda risposta di VON ANDRAE [4] e infine l'intervento di HELMERT che dedicò alla questione una sua memoria [5].

In questi lavori si incontrano, come il GINI stesso avvertiva nella ricordata citazione, le due seguenti espressioni della differenza media: in [4] la formula corrispondente al primo metodo di GINI, e in [1] la formula

$$\sum_{ij}^{n} \left| x_i - x_j \right| = 2\,n \left(\sum_{i=1}^{n} a_i - n\,a_1\right) - \sum_{i=1}^{n-1} (n - i)^2\,(a_{i+1} - a_i),$$

che, al pari della simmetrica (non segnata da JORDAN)

$$\sum_{ij}^{n} \left| x_i - x_j \right| = 2\,n \left(n\,a_n - \sum_{i=1}^{n} a_i\right) - \sum_{i=1}^{n-1} i^2\,(a_{i+1} - a_i),$$

si ricava facilmente dalla formula del secondo metodo (DE FINETTI-PACIELLO) sviluppando rispettivamente i $(n - i) = n\,(n - i) - (n - i)^2$ e $i\,(n - i) = i\,n - i^2$.

Quanto allo scopo, già accennato, di tali lavori, si comprende facilmente che esso è ben lontano da quello dello studio statistico della variabilità, nel quale dunque, come s'è asserito, spetta al GINI il merito d'aver introdotta la differenza media come un utile indice.

Si tratta, in sostanza, di eseguire separatamente la somma dei termini positivi e negativi. La formulazione più semplice è quella dello CZUBER [8]: pongasi

$$
\begin{array}{ll}
s_1 = x_1 & s'_1 = x_n \\
s_2 = x_1 + x_2 & s'_2 = x_n + x_{n-1} \\
\ldots\ldots & \ldots\ldots \\
s_n = x_1 + x_2 + \ldots + x_n & s'_n = x_n + x_{n-1} + \ldots + x_1 \\
S = s_1 + s_2 + \ldots + s_n & S' = s'_1 + s'_2 + \ldots + s'_n \,;
\end{array}
$$

risulta $\Sigma_{ij} |x_i - x_j| = 2\,(S' - S)$ (α).

Si osservi che si può scrivere anche

$$S = \sum_{1}^{n}{}_i (n - i + 1)\, x_i \qquad S' = \sum_{1}^{n}{}_i\, i x_i \qquad (*).$$

Notando che $S + S' = (n + 1) \sum_{1}^{n}{}_i x_i = n\,(n + 1)\, M$ ove $M =$ media aritmetica di $x_1, x_2, \ldots x_n$, si hanno le ovvie identità

$$2\,(S' - S) = 2\,(S' + S) - 4\,S = 2\,n\,(n + 1)\,\mathrm{M} - 4\,S \qquad (\beta)$$

e $$2\,(S' - S) = 4\,S' - 2\,(S' + S) = 4\,S' - 2n\,(n+1)\,M \qquad (\gamma).$$

Diremo procedimento (α^*), (β^*), (γ^*) quello corrispondente alle stesse formule (α), (β), (γ) quando però le S, S' si calcolino col procedimento $(*)$ anzichè con quello originario dello CZUBER. Questa terminologia permette facilmente di identificare tutte le possibili varianti di dettaglio.

La formula data da GINI [7] non è che la (γ^*); è ottenuta peraltro per via completamente diversa, e cioè partendo dal rapporto di concentrazione R, legato alla differenza media dalla nota relazione $\Delta_R = 2\,MR$.

Il MORTARA [11] riporta la (α) (la sua [3c]; p. 152) colla dimostrazione dello CZUBER, e la (γ) (la sua [3f], p. 153) con una dimostrazione inutilmente complicata.

Il PIETRA, in [9] scrive la (α) in una forma che rende necessario considerare separatamente un certo termine che è invece conglobato in quella di CZUBER; la complicazione, per sè stessa inutile, è però giustificata dal fatto che egli vuol definire la differenza media come media aritmetica degli scostamenti semplici medi dei diversi termini. Questa proprietà era stata già rilevata dal GINI in [6] (pp. 35-36), ed è anzi interessante osservare che le formule (26) e (27) ivi ottenute contengono già, in effetto, la (α) in cui però la S si calcola col metodo (*) e la S' con quello formulato due anni più tardi da CZUBER. Nell'altro lavoro del PIETRA [10] è riportata la (α) nella forma e colla dimostrazione dello CZUBER; egli dà poi una disposizione di calcolo (che dice « self explanatory », e quindi non spiega) ispirata alla formula (γ^*).

Il VINCI [12] dà, in sostanza, la (α^*), scrivendola però in forma diversa: come somma di differenze anzichè come differenza di somme: ciò può essere vantaggioso o svantaggioso, per la rapidità dei calcoli, a seconda dei casi e del tipo di macchine usato.

Il DE GLERIA [14] dà infine come « nuova formula » la (β), e, nella seconda nota [15] una disposizione di calcolo (β^*), che è perfettamente simmetrica a quella (γ^*) di PIETRA [10], e quindi ad essa perfettamente equivalente dal punto di vista della rapidità dei calcoli.

4. Metodo grafico (GINI).

La relazione $\Delta_R = 2\,MR$, già richiamata, permette infine di ottenere Δ da R determinato graficamente, misurando l'area della curva di concentrazione (GINI, [7]).

III.

Fare apprezzamenti sulla maggiore o minore convenienza delle diverse formule dal punto di vista pratico è cosa assai delicata e difficile. Tanto più che bisognerebbe tener conto di molti fattori diversi da caso a caso (numero dei termini, ordine di grandezza dei valori, delle loro differenze e delle frequenze, tipo di macchina usato, ecc.), e che non si può mai escludere l'intervento di fattori soggettivi.

Ciò non è però un motivo sufficiente per non dire la propria opinione, pur avvertendo onestamente che si tratta di un'opinione e che non ha nessuna pretesa di costituire un giudizio. Io ritengo che, per il metodo GINI-CZUBER, la forma più conveniente sia la (γ^*) adottata dal GINI, rispondente alla formula

$$\Sigma_{ij}\,|x_i - x_j| = 4\,\Sigma_i^{n}{}_{1}\, i x_i - 2n\,(n+1)\,M\,.$$

Scartando il primo metodo di GINI, che non si presta agevolmente al caso generale in cui i termini sono ripetuti con frequenze diverse, rimarrebbero in lizza, fra le formule analitiche, quella ora riscritta e quella recentemente proposta, insieme al sig. U. PACIELLO, da me. Dato che esse sono totalmente diverse, è assai difficile giudicare a priori della maggiore o minore semplicità d'applicazione. Gli esperimenti fatti dal sig. U. PACIELLO in occasione di vari lavori dell'Ufficio Matematico dell'Istituto Centrale di Statistica riuscirono favorevoli al nuovo metodo, ma non è da escludere, date le contrarie asserzioni del DE GLERIA, che ciò fosse dovuto a circostanze particolari. Non c'è del resto fretta di concludere, nè motivo di desiderare il successo a una formula piuttosto che all'altra, sia pure la propria. Solo è augurabile che tutte vengano ripetutamente usate, in modo da lasciar decidere, con piena cognizione di causa, alla pratica.

Debbo dire infine che, per semplicità di scrittura, ho qui riportate le formule riferendomi al caso in cui i termini x_1, $x_2 \ldots x_n$, sono distinti (anzichè ripetuti, ciascuno con una frequenza f_1, $f_2 \ldots f_n$) ; ho fatto così dopo essermi sincerato che ciò nulla toglieva di sostanziale al raffronto che mi ero proposto. L'unico leggermente danneggiato — nei confronti con gli altri — da questa semplificazione è il metodo del VINCI, che è però in ogni caso meno semplice del (γ *).

Il metodo grafico, usando un buon planimetro, permette di raggiungere un ottimo grado di approssimazione, come provarono ripetuti esperimenti eseguiti, specialmente dall'ing. D'ANCONA, presso l'Istituto Centrale di Statistica. Data la radicale diversità fra un metodo grafico e uno analitico, è ancor meno il caso di fare raffronti: la preferenza per l'uno o l'altro tipo di procedimenti dipenderà infatti piuttosto da circostanze esteriori come la maggiore o minore abitudine a calcoli grafici e numerici, e la maggiore o minore comodità di eseguirli.

LAVORI CITATI

[1]. W. JORDAN, *Ueber die Bestimmung der Genauigkeit mehrfach wiederholter Beobachtungen einer Unbekannten.* — « Astronomische Nachrichten », Bd. 74, n. 1766-67, 1869.

[2]. VON ANDRAE. *Schreiben des Herrn Geheimen Etatsrath von Andrä am der Herausgeber.*— Ibidem, n. 1770, 1869.

[3]. W. JORDAN. *Ueber die Bestimmung des mittleren Fehlers durch Wiederholung der Beobachtungen.* — Ibidem, n. 1886.

[4]. VON ANDRAE. *Ueber die Bestimmung des wahrscheinlichen Fehlers durch die gegebenen Differenzen vom gleichgenauen Beobachtungen einer Unbekannten.* — Ibidem, n. 1889.

[5]. HELMERT. Die Berechnung des wahrscheinlichen Beobachtungsfehlers aus der ersten Potenzen der Differenzen gleichgenauer direkter Beobachtungen. — Ibidem, n. 2097.

[6]. C. GINI. *Variabilità e Mutabilità.* — « Studi economico-giuridici pubblicati per cura della Facoltà di Giurisprudenza della R. Università di Cagliari » Anno III, p. II, Bologna, Cuppini, 1912.

[7]. Id. id. *Sulla misura della concentrazione e della variabilità dei caratteri.* — « Atti R. Ist. Veneto S. L. A. », T. LXXIII, P. II, Venezia, 1914.

[8]. E. CZUBER. *Beitrag zur Theorie statistischer Reihen.* — « Versicherungswissenschaftlichen Mitteilungen », N. F., 9. Bd., 2 Heft, Vienna, 1914.

[9]. G. PIETRA. *Delle relazioni tra gli indici di variabilità.* — « Atti R. Ist. Veneto S. L. A. », T. LXXIV, P. II, Venezia, 1915.

[10]. Id. id. *The theory of statistical relations with special reference to cyclical series.* — « Metron », Vol. IV, N. 2-4, Ferrara, 1925.

[11]. G. MORTARA. *Statistica Metodologica.* — Città di Castello, Tip. « Leonardo da Vinci », 1922.

[12]. F. VINCI. *Statistica Metodologica.* — Padova, La Litotipo, 1924.

[13]. B. DE FINETTI e U. PACIELLO. — *Calcolo della differenza media.* — « Metron », Vol. VIII, N. 3, Roma, 1930.

[14]. A. DE GLERIA. *Una abbreviazione nel calcolo della differenza media.* — « Rivista Italiana di Statistica », Anno I, N. 4, Città di Castello, 1929.

[15]. Id. id. *Sul calcolo della differenza media.* — « Rivista Italiana di Statistica », Anno II, N. 2, Città di Castello, 1930.