IL VALORE DELLE DIFFERENZE SESSUALI

dal punto di vista biometrico

NOTA

del Dott. FERNANDO DE HELGUERO

I metodi biometrici trovano una applicazione in molti problemi di antropologia fisica e la loro precisione porta un valido contributo in molte ricerche: senza entrare nel campo della *correlazione* e della *eredità* mi basta ricordare la memoria del Pearson sulla ricostruzione della statura per mezzo delle ossa lunghe (1) e la recente polemica sulle formule che dànno la capacità cranica in funzione di misure lineari (2).

Io voglio mostrare in questa nota come essi trovino una applicazione anche per valutare il valore sessuale dei caratteri, siano essi misure dirette, ovvero indici.

Ricordo il metodo che si segue comunemente in antropologia per lo studio delle differenze sessuali: si misura un dato carattere in un gran numero di individui dell'un sesso e dell'altro, se ne fanno le medie, maschile e femminile, e si osservano i valori estremi. Poi si confrontano le medie (ovvero gli estremi) risolvendo la proporzione

$$\text{Media femminile} : \text{Media maschile} :: x : 1000.$$

Si ha così quello che si dice *media femminile prendendo quella*

(1) Karl Pearson, *Mathematical Contributions to the Theory of Evolution*, Phil. Trans. of the R. Soc. of London A 192, 1899.

(2) V. John Beddoe, *A method of estimating Skull-capaeity from peripheral measures*, in Journal of Anth. Inst., Vol. XXXIV, 1904.

maschile uguale a 1000. Talvolta anzichè 1000 si prende 100 od altro numero.

Ciò fatto si considera come misura del valore sessuale del carattere in esame la differenza fra il numero così ottenuto e 1000. Per esempio, di due caratteri, le cui medie femminili riferite a quelle maschili come 1000, siano 880 e 900, si dice senz'altro che il primo ha maggior valore dell'altro come carattere sessuale.

Questo metodo è inesatto sia che si confrontino i valori estremi delle serie maschile e femminile, sia che si confrontino le medie.

Per i valori estremi ciò è evidente, perchè essi mentre non possono certo considerarsi come rappresentanti dell'intiera serie, non hanno nemmeno alcun valore statistico in quanto dipendono dal numero di individui studiati: infatti misurando delle serie poco numerose è molto difficile di trovare gli stessi valori estremi dati da serie molto numerose. Questa è la ragione del poco valore statistico dei metodi somatometrici del prof. Camerano che sono basati quasi esclusivamente su questi elementi.

La media invece non ha questo difetto: se la serie è abbastanza numerosa la media è completamente indipendente dal numero di individui misurati. Ma per essa resta l'altro difetto accennato per gli estremi: nemmeno la media può prendersi come rappresentante di tutta una serie e sarebbe inutile spendere troppe parole su ciò ricordando la spiritosa demolizione che ne ha fatto il Bertrand nella introduzione del suo *Calcul des probabilités*, se non fosse rimasta in molti antropologi che adoperano metodi statistici quella falsa idea: basta che ricordi l'importanza che si dà al tipo medio ed il nessun conto in cui è tenuta la variabilità.

La rappresentazione completa della grandezza di un carattere in un gruppo è data dalla *variazione* del carattere: sia essa data sotto la forma grafica di *seriazione*, sia espressa analiticamente sotto forma della equazione di una curva che poco si discosti dal poligono empirico della seriazione.

Della variazione è elemento importantissimo la media, ma non ne è l'unico: essa considera anche come gli individui si raggruppano intorno alla media, cioè tien conto della *variabilità* del carattere. Per dimostrare nel nostro caso particolare che la media non è sufficiente per giudicare del valore sessuale di un carattere farò un esempio.

Fig: 1.ª

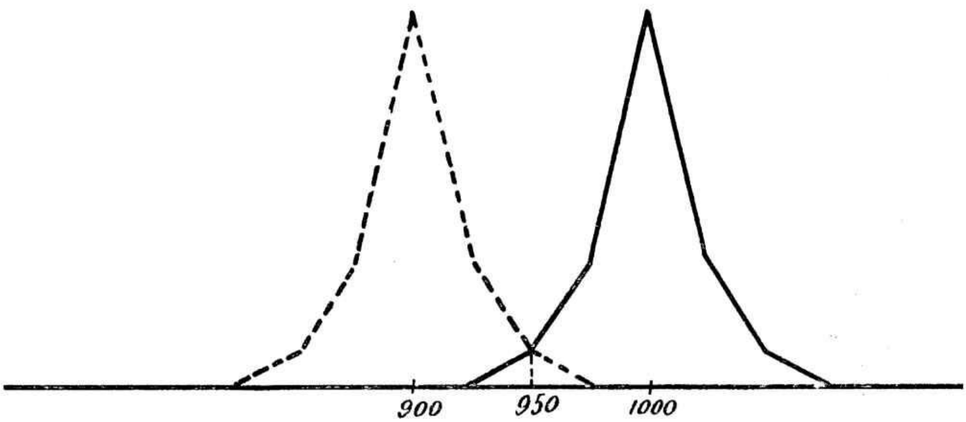
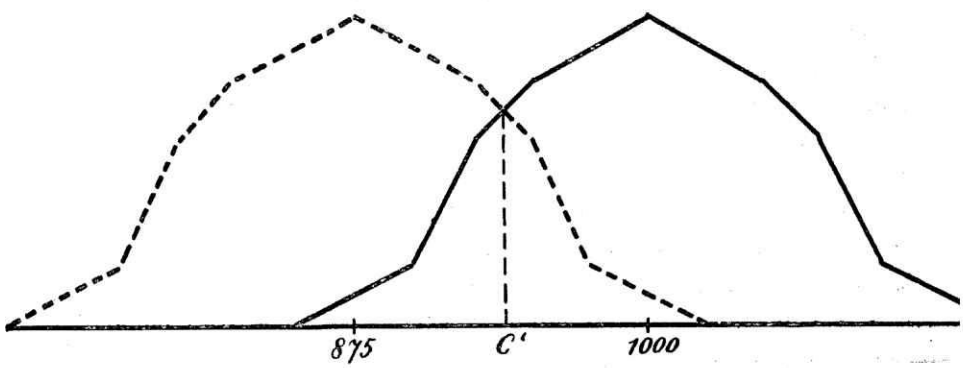


Fig: 2.ª



Supponiamo due caratteri che diano le seriazioni rappresentate nelle figure 1 e 2, entrambe costruite prendendo le medie maschili uguali a 1000. Le seriazioni maschili son tracciate con linea piena, quelle femminili son tratteggiate.

Le medie femminili sono 900 e 875, perciò, secondo il criterio comune, dovremmo dire il secondo carattere migliore del primo per la differenziazione dei sessi. Ma osservando la prima figura si noti che quasi tutti i maschi hanno un valore superiore a 950 e quasi tutte le femmine un valore inferiore. Diremo 950 *valore critico*. Il carattere avente una tale seriazione ha un grande valore come indice sessuale.

Altrettanto non può dirsi del carattere rappresentato nella seconda figura poichè qualunque valore si scelga (anche C′ che come dimostrerò è il più opportuno) non potremo certo asserire che questo separa i maschi dalle femmine poichè molti maschi han valori inferiori di molte femmine.

In altre parole il metodo usato dagli antropologi sarebbe ottimo se tutti i maschi e tutte le femmine presentassero per quel carattere il valore della rispettiva media, o ne differissero pochissimo, ma nella pratica noi abbiamo una grande variabilità sia nei maschi che nelle femmine, per cui molti maschi sono inferiori alla media femminile e molte femmine superiori della media maschile. Quanto maggiore è questa variabilità tanto minore è il valore sessuale del carattere, per cui talvolta la minore variabilità può compensare, come nel caso schematico sopra esposto, la minore differenza delle medie.

Inoltre il metodo usato in antropologia non ci dà modo di trovare il valore critico, cioè quel valore per il quale, assumendo come femmine tutti gli individui con valore inferiore (o superiore) e come maschi gli altri, l'errore che si commette sia minimo.

Per l'importanza di un tale valore che si assume oggi empiricamente in antropologia basta che citi l'*indice ilio-pelvico* od *indice sessuale* del prof. Sergi.

Il problema che mi propongo è appunto questo: Dato il valore di un carattere (misura od indice) in una serie di maschi e di femmine cercare il *valore critico*. L'*errore percentuale* che si commette considerando come femmine o maschi tutti gli individui che superano o no questo valore ci darà un ottimo elemento per giudicare della bontà di un carattere come indice sessuale anche se non vogliamo servirci del valore critico per la determinazione del sesso in nuovi individui.

Il problema è duplice: 1° Ricerca del valore critico; 2° misura dell'errore percentuale nel senso sopra espresso. In questa ricerca noi ammettiamo che la variabilità segua la legge normale: supponiamo note le medie b_1 e b_2 femminile e maschile e σ_1 σ_2 le rispettive *deviazioni normali* che misurano la variabilità. Per non interrompere con inopportune digressioni questa nota esporrò in una appendice il modo di calcolare questi numeri data la statistica.

In una seconda appendice dò la trattazione matematica del problema del quale ecco la soluzione:

Il valore critico è dato dalla formula

$$x = \frac{b_1 \sigma_2^2 - b_2 \sigma_1^2 \pm \sigma_1 \sigma_2 \sqrt{(b_1 - b_2)^2 - 4{\cdot}6052 (\sigma_2^2 - \sigma_1^2)\, lg \frac{\sigma_1}{\sigma_2}}}{\sigma_2^2 - \sigma_1^2}$$

dove il logaritmo è a base 10 e vale uno solo dei segni $\pm$, il superiore o l'inferiore, a seconda dei casi, e ciò non resta mai dubbio nella pratica.

Trovato x possiamo calcolare separatamente i due errori che si commettono prendendo come maschi tutti gli individui superiori ad x e come femmine gli altri.

I due errori son dati da

$$E_1 = \frac{5000 - \Theta_1 \left(\frac{x - b_1}{\sigma_1}\right)}{200} \quad \text{e} \quad E_2 = \frac{5000 - \Theta_1 \left(\frac{b_2 - x}{\sigma_2}\right)}{200},$$

dove $\Theta_1(x)$ è l'integrale normale il cui valore si trova nelle tavole (p. es: Davenport, *Statistical Methods, etc.*, Tav. IV, pag. 55).

L'errore percentuale si ha allora facendo la somma dei due errori $E_1 + E_2$.

Per chiarire ciò con un esempio prenderò i dati del Morselli (¹) relativi al peso del cranio e della mandibola.

Per il peso del cranio si trova:

Per le femmine $b_1 = 526$ gr. $\sigma_1 = 106.1$
Per i maschi $b_2 = 597$ gr. $\sigma_2 = 113.8$

(¹) Morselli, *Sul peso del cranio e della mandibola in rapporto col sesso.* Archivio d'Antropologia, V, p. 149.

Il valore critico si trova $x = 572$ grammi.

Se consideriamo come maschi tutti i crani superiori a 572 gr. commettiamo un errore $E_1 = 16.62\ ^0/_0$. Assumendo come femminili tutti i crani inferiori a 572 grammi si commette un errore $E_2 = 20.66\ ^0/_0$. In complesso si ha un errore totale $E = 37.28\ ^0/_0$.

Similmente per il peso della mandibola:

$$\begin{array}{llll} \text{Per le femmine} & \sigma_1 = 12.6 & & b_1 = 64.0 \\ \text{Per i maschi} & \sigma_2 = 18.0 & & b_2 = 82.4. \end{array}$$

Il valore critico è 75.6.

$$\begin{array}{lrl} \text{I due errori sono} & E_1 = & 8.87\ ^0/_0 \\ & E_2 = & 17.62\ ^0/_0 \\ \hline \text{L'errore totale è} & E = & 26.49\ ^0/_0. \end{array}$$

Come si vede il peso della mandibola è miglior indice sessuale del peso del cranio, come del resto avrebbero fatto supporre le medie femminili riferite a quelle maschili come 1000: esse sono 777 per il peso della mandibola e 881 per il peso dei crani.

Questo stesso metodo può naturalmente applicarsi per lo studio delle differenze sessuali in tutta la scala zoologica, per esempio per la determinazione del valore critico nel peso dei bozzoli del Bombyx mori, valore che si assume empiricamente negli allevamenti dei bachi da seta e che pure ha molta importanza per la separazione dei sessi.

Questo stesso metodo può anche applicarsi allo studio oltrechè delle differeze di sesso, a quelle di specie, di varietà, ecc., ma l'ho esposto sotto questa forma perchè specialmente per i sessi non può trovarsi per alcun carattere un valore critico assoluto, esistendo sempre per qualunque carattere un certo numero di individui che presentano dei valori più lontani dalla loro media che non alcuni individui del sesso opposto. In tal caso è necessario ricorrere al metodo statistico e non può supplire nè la somatometria, nè il metodo della seriazione.

APPENDICE I.

Determinazione dei parametri di una curva normale.

Data una serie di valori corrispondenti alle misure di uno stesso carattere in una serie di individui per trovare i parametri o costanti della variazione (cioè la *media* e la *deviazione normale)* si procede nel modo seguente:

Si fa la *seriazione* dei numeri dati *(varianti)* raccogliendo in classi quelli uguali o poco diversi fra di loro. Dicesi *frequenza (f)* di una classe il numero di varianti che essa contiene; *grandezza* di una classe il valore medio (o comune) dei numeri che in essa figurano. Così se misurando il peso di una serie di mandibole maschili ne troviamo 14 che hanno un peso compreso fra 60 e 70 grammi diremo che la classe di grandezza 65 ha la frequenza uguale a 14.

Si scelga ora una classe vicina alla media (indichiamo Y_m la sua grandezza) e ad essa si faccia corrispondere il numero zero, mentre alle classi successive si danno i numeri 1, 2... ordinatamente, e a quelle precedenti i numeri negativi — 1, — 2...: diremo questi numeri *grandezze relative* delle classi e le indicheremo con Y. Sia d la differenza fra le grandezze assolute di due classi successive ed n il numero totale di varianti.

Calcoliamo

$$\nu_1 = \frac{\Sigma\,(Yf)}{n} \quad \text{e} \quad \nu_2 = \frac{\Sigma\,({}^2 Y f)}{n},$$

dove il simbolo Σ indica che devonsi fare i prodotti $(Y.f)$ o $(Y.^2 f)$ per ogni singola classe e poi si devono sommare. La media è data allora da

$$M = Y_m + \nu_1\, d,$$

e la deviazione normale da

$$\sigma = d\,\sqrt{\nu_2 - \nu_1^2 + \tfrac{1}{6}}.$$

Come esempio calcoliamo i parametri della statistica del peso delle mandibole nei ♀ data dal Morselli. Si hanno le classi:

Grandezza:	40-50	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130.
Frequenza:	1	11	14	18	24	15	10	6	1

Allora $d = 10$, $n = 100$. Prendiamo $Y_m = 85$; allora si ha il quadro:

Grandezza assoluta	f	Y	$(Y.f)$	$(Y.^2 f)$
45	1	— 4	— 4	+ 16
55	11	— 3	— 33	+ 99
65	14	— 2	— 28	+ 56
75	18	— 1	— 18	+ 18
85	24	0	0	0
95	15	1	+ 15	+ 15
105	10	2	+ 20	+ 40
115	6	3	+ 18	+ 54
125	1	4	+ 4	+ 16
	$n = 100$		$\nu_1 = \frac{-26}{100}$	$\nu_1 = \frac{314}{100}$

Perciò $\nu_1 = -0{,}26 \qquad \nu_2 = 3{,}14$.

La media è $M = 85 - 2{,}6 = 82{,}4$.

La deviazione normale è $\sigma = 10\,\sqrt{3{,}14 - 0{,}0676 + 0{,}1667} = 18{,}0$.

APPENDICE II.

Trattazione analitica del problema.

Siano $\frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-b_1)^2}{2\sigma_1^2}}$, $\frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-b_2)^2}{2\sigma_2^2}}$ le leggi di variazione del carattere in esame rispettivamente nelle femmine e nei maschi

Si chiede di cercare il valore x_0 di x che rende minima l'espressione

$$\int_{x_0}^{\infty} \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-b_1)^2}{2\sigma_1^2}} dx + \int_{-\infty}^{x_0} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-b_2)^2}{2\sigma_2^2}} dx.$$

L'espressione può mettersi sotto la forma:

$$y = \frac{1}{\sqrt{2\pi}} \left\{ -\int_{\infty}^{x_0} \frac{1}{\sigma_1} e^{-\frac{(x-b_1)^2}{2\sigma_1^2}} dx + \int_{-\infty}^{x_0} \frac{1}{\sigma_2} e^{-\frac{(x-b_2)^2}{2\sigma_2}} dx. \right.$$

Per cercare il suo minimo basta porre uguale a zero la sua derivata

$$\frac{dy}{dx_0} = -\frac{1}{\sigma_1} e^{-\frac{(x_0-b_1)^2}{2\sigma_1^2}} + \frac{1}{\sigma_2} e^{-\frac{(x_0-b_2)^2}{2\sigma_2^2}} = 0;$$

da cui $\frac{1}{\sigma_1} e^{-\frac{(x_0-b_1)^2}{2\sigma_1^2}} = \frac{1}{\sigma_2} e^{-\frac{(x_0-b_2)^2}{2\sigma_2^2}}$

Prendendo i logaritmi a base e:

$$\log \sigma_1 + \frac{(x_0-b_1)^2}{2\sigma_1^2} - \log \sigma_2 - \frac{(x_0-b_2)^2}{2\sigma_2^2} = 0$$

$$\sigma_2^2 (x_0 - b_1)^2 - \sigma_1^2 (x_0 - b_2)^2 + 2\sigma_1^2 \sigma_2^2 \log \frac{\sigma_1}{\sigma_2} = 0$$

e sviluppando

$$x_0^2 (\sigma_2^2 - \sigma_1^2) - 2 x_0 (b_1 \sigma_2^2 - b_2 \sigma_1^2) + (b_1^2 \sigma_2^2 - b_2^2 \sigma_1^2) + \\ + 2\sigma_1^2 \sigma_2^2 \log \frac{\sigma_1}{\sigma_2} = 0.$$

E risolvendo:

$$x_0 = \frac{b_1 \sigma_2^2 - b_2 \sigma_1^2 \pm \sqrt{\sigma_1^2 \sigma_2^2 (b_1 - b_2)^2 - 2\sigma_1^2 \sigma_2^2 (\sigma_2^2 - \sigma_1^2) \log \frac{\sigma_1}{\sigma_2}}}{\sigma_2^2 - \sigma_1^2}$$

e ponendo i logaritmi decimali invece dei neperiani:

$$x_o = \frac{b_1 \sigma_2^2 - b_2 \sigma_1^2 \pm \sigma_1 \sigma_2 \sqrt{(b_1 - b_2)^2 - 4{,}6052 (\sigma_2^2 - \sigma_1^2) \log \frac{\sigma_1}{\sigma_2}}}{\sigma_2^2 - \sigma_1^2}$$

Questa formula si semplifica se si prende come origine la media femminile, ciò che può sempre farsi. Allora $b_1 = 0$ e $b_2 = b$ esprime la differenza fra le medie. La formula diviene:

$$x_0 = \frac{- b \sigma_1^2 \pm \sigma_1 \sigma_2 \sqrt{b^2 - 4{,}6052 (\sigma_2^2 - \sigma_1^2) \log \frac{\sigma_1}{\sigma^2}}}{\sigma_2^2 - \sigma_1^2}.$$

Si osservi che geometricamente il valore critico x_0 è l'ascissa del punto di incontro delle due curve e questo spiega l'esistenza del doppio segno, poichè le intersezioni sono due ma una cade molto lontana dalle medie e non interessa il nostro problema.

Se non si avessero empiricamente le curve normali maschile e femminile ma si avesse una sola curva dimorfica potremmo scomporla coi metodi dal Pearson e da me proposti per ciò (V. *Per la risoluzione delle curve dimorfiche*, HELGUERO, nelle Memorie della R. Acc. dei Lincei, 1906, pag. 163-203).

L'errore E_1 è dato da:

$$\frac{1}{\sigma_1 \sqrt{2\pi}} \int_{x_0}^{\infty} e^{-\frac{(x - b_1)^2}{2 \sigma_1^2}} dx = E_1 \text{ ed analogamente } E_2.$$

Possiamo scrivere

$$E_1 = \frac{1}{\sqrt{\pi}} \int_{\frac{x_0 - b_1}{\sqrt{2} \sigma_1}}^{\infty} e^{-z^2} dx$$

ponendo $\frac{x - b_1}{\sigma_1 \sqrt{2}} = z$, da cui $\frac{dx}{\sigma_1 \sqrt{2}} = dz$.

Se per poter applicare la tavola data dal BERTRAND *(Calcul des probabilités)*

poniamo $\Theta(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt$ si ha

$$E_1 = \left\{ 1 - \Theta\left(\frac{x_0 - b_1}{\sigma_1 \sqrt{2}}\right) \right\} \frac{100}{4}, \text{ ovvero}$$

$$E_1 = \left\{ 1 + \Theta\left(\frac{b_1 - x_0}{\sigma_1 \sqrt{2}}\right) \right\} \frac{100}{4},$$

secondo che $x_0 \gtrless b_1$.

E analogamente

$$E_2 = \left\{ 1 - \Theta \left(\frac{b_2 - x_0}{\sigma_2} \right) \right\} \cdot 25 \quad \text{ovvero}$$

$$E_2 = \left\{ 1 + \Theta \left(\frac{x_0 - b_2}{\sigma_2} \right) \right\} \cdot 25$$

secondo che $\qquad b_2 > x_0 \qquad$ ovvero $\qquad b_2 < x_0$.

Se invece si vuol far uso della Tav. IV del DAVENPORT (pag. 55 dello *Statistical Methods*, ecc.) che dà i valori dell'integrale di probabilità normale cioè di

$$\frac{10000}{\sqrt{2\pi}} \int_0^t e^{-\frac{1}{2}t^2} dt = \Theta_1$$

allora gli errori son dati da

$$E_1 = \frac{5000 - \Theta_1 \left(\frac{x_0 - b_1}{\sigma_1} \right)}{200} \quad \text{e} \quad E_2 = \frac{5000 - \Theta_1 \left(\frac{b_2 - x_0}{\sigma_2} \right)}{200}$$

se come è ordinariamente $b_1 < x_0 < b_2$. Se invece fosse $x_0 < b_1$

$$E_1 = \frac{5000 + \Theta_1 \left(\frac{b_1 - x_0}{\sigma_1} \right)}{200}$$

ed E_2 resta lo stesso, e se fosse $x_0 > b_2$, E_1 resta lo stesso ed E_2 diviene

$$E_2 = \frac{5000 + \Theta_1 \left(\frac{x_0 - b_2}{\sigma_2} \right)}{200}.$$

Le prime di queste formole son quelle riportate nella nota. Abbiamo messo queste e non quelle che richiedono l'uso della tavola del Bertrand perchè dànno dei risultati più esatti.