## 5.3. Indici di variabilità

### 5.3.1. Intervalli di variazione

Sostituire una distribuzione con un valore medio, per quanto esso possa esserne rappresentativo, causa comunque una forte perdita di informazione. Diventa perciò necessario trovare strumenti che esprimano in modo sintetico qual è l'attitudine del carattere studiato ad assumere modalità differenti, a variare. La variabilità, d'altronde, è una caratteristica della realtà. La ragion d'essere della statistica sta proprio nella variabilità dei fenomeni e nel conseguente tentativo di esprimerla e spiegarla.

Per i caratteri quantitativi, la più semplice misura di variabilità è la differenza fra il valore massimo e il valore minimo presentato dal carattere. Essa è indicata con V ed è chiamata campo di variazione. Rispetto alla distribuzione unitaria, se $\xi_{(1)}$ è il minimo e $\xi_{(n)}$ è il massimo valore delle modalità, si ha:

$$V = \xi_{(n)} - \xi_{(1)} .$$

L'intervallo compreso fra min(X) e max(X) contiene tutte le n osservazioni, mentre la sua ampiezza V esprime la differenza massima fra due modalità osservate. V è sensibile alla eventuale presenza di osservazioni anomale. Rispetto alle distribuzioni unitarie contenute nella tabella 1 si ottiene rapidamente che il campo di variazione del voto di maturità è V=60/60-36/60=24/60.

Tab. 1 - Studenti presenti alla lezione di statistica del 6-10-1994
per voto alla maturità

| | Voto alla maturità | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 36 | 37 | 38 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 48 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 60 | n.i | Tot. |
| N. studenti | 5 | 1 | 8 | 8 | 2 | 12 | 1 | 8 | 2 | 3 | 6 | 8 | 1 | 8 | 1 | 7 | 1 | 3 | 1 | 5 | 17 | 1 | 109 |

Rispetto alla distribuzione di frequenze espresse in classi, il calcolo di V può presentare difficoltà se la classe iniziale e finale sono aperte. Per ovviare a tali problemi si può costruire un intervallo che non dipenda dai dati che si trovano in posizione estrema. Occorre perciò riferirsi a valori caratteristici, interni all'intervallo $\xi_{(1)} \vdash\!\dashv \xi_{(n)}$ . In particolare si possono utilizzare i quartili: $Q_1$ e $Q_3$ e costruire $V_1$, la differenza interquartile:

$$V_1 = Q_3 - Q_1$$

L'intervallo $Q_1 \vdash\!\dashv Q_3$ , per la definizione dei quartili, non contiene meno del 50% delle osservazioni, mentre negli intervalli $\xi_{(1)} \vdash Q_1$ e $Q_3 \dashv \xi_{(n)}$ al massimo vi sono rispettivamente il 25% delle unità.

I calcoli già effettuati per ottenere i quartili (cfr. § 5.2.3) consentono di ottenere rapidamente i valori di $V_1$.

Rispetto al voto alla maturità si ha: V1= 54,5/60-42/60= 12,5/60.

Fra 42/60 e 54,5/60 cadono non meno del 50% delle osservazioni (in effetti, controllando sulla tabella 15, l'intervallo contiene il 52,8% degli studenti).

Rispetto alla statura delle studentesse, si ha: $V_1$ = cm (169,81-161,62)= cm 8,19
l'area dell'istogramma delimitata da $Q_1$ e $Q_3$ (cfr. Fig. 1) è pari esattamente al 50% della distribuzione.

### 5.3.2. Scostamenti medi

#### 5.3.2.1. Concetti e problemi di calcolo

Gli intervalli di variazione sono costruiti con particolari valori caratteristici osservati. Si può invece richiedere che la variabilità venga espressa utilizzando tutte le n osservazioni. Si consideri la seguente distribuzione unitaria dei voti riportati all'esame di statistica da 10 studenti:

28 21 24 24 20 25 27 27 24 24.

Come già messo in evidenza, si può sintetizzare la distribuzione calcolando il voto medio. Si ottiene così:

$$\mu = \frac{20 + 21 + 24 \cdot 4 + 25 + 27 \cdot 2 + 28}{10} = 24{,}4 \,.$$

24,4/30 compensa nel collettivo i voti alti e quelli bassi, ma ogni unità differisce da questo valore fittizio, non fosse altro perché, come voto di esame, 24,4 è inesistente, essendo i voti espressi su una scala numerica formata dai numeri interi da 1 a 30.

Rappresentando i voti come punti su di un segmento, sul quale viene indicata anche la media aritmetica, la diversità fra ciascun voto $\xi_i$ e $\mu$ è indicata da un segmento (Fig. 1).

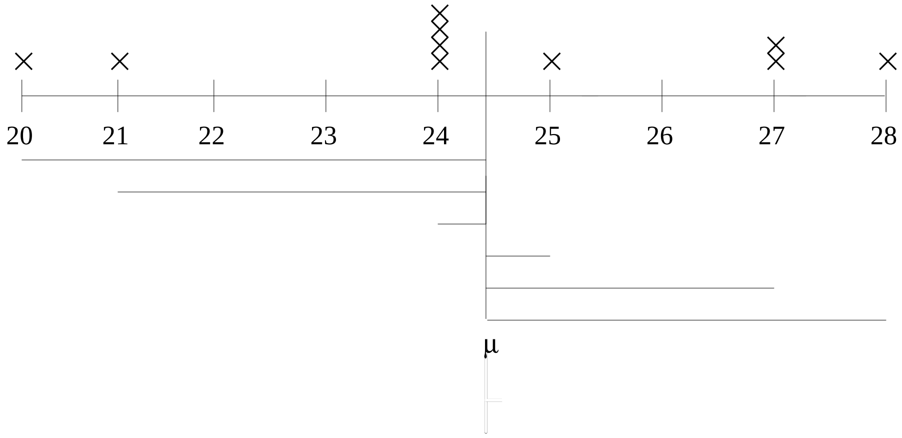


Fig. 1 - Distribuzione degli studenti secondo il voto riportato all'esame di statistica e rappresentazione degli scarti dal voto medio

Ciò che interessa per esprimere la variabilità sono dunque le n quantità $|\xi_i - \mu|$. Se $|\xi_i - \mu| = 0$, per tutti gli i, non vi è variabilità, mentre quanto più le diversità $|\xi_i - \mu|$ sono grandi, tanti più i valori $\xi_i$ sono lontani da $\mu$ e il carattere è variabile. Come misura di variabilità si può pertanto proporre una media di potenze di ordine r delle differenze in modulo $|\xi_i - \mu|$.

Per r=1, si ha lo scostamento semplice medio dalla media aritmetica ($S_\mu$), che per distribuzioni unitarie ha l'espressione:

$$S_\mu = \frac{\sum_{i=1}^{n} |\xi_i - \mu|}{n} \,;$$

e per le distribuzioni di frequenze è:

$$S_\mu = \frac{\sum_{t=1}^{k} |x_t - \mu| \cdot n_t}{n} = \sum_t |x_t - \mu| \, f_i = \frac{\sum_{t=1}^{k} |x_t - \mu| \cdot p_t}{100} \,.$$

Per r=2 si ha la media quadratica degli scostamenti, in tal modo si "valorizzano le diversità grandi", mentre "penalizzano le diversità piccole". La misura proposta è lo scostamento quadratico medio dalla media aritmetica che per distribuzioni unitarie è espresso da:

$$\sigma = \sqrt{\frac{\sum_i (\xi_i - \mu)^2}{n}} \,.$$

Per distribuzioni di frequenze si ha:

$$\sigma = \sqrt{\frac{\sum_t (x_t - \mu)^2 \cdot n_t}{n}} = \sqrt{\sum_t (x_t - \mu)^2 \cdot f_t} = \sqrt{\frac{\sum_t (x_t - \mu)^2 \cdot p_t}{100}} .$$

Se anziché sintetizzare la distribuzione con $\mu$ si usa la mediana m, si possono calcolare le diversità $|\xi_i - m|$. Per misurare la variabilità, in questo caso, si può utilizzare lo scostamento semplice medio dalla mediana ($S_m$). Esso, per distribuzioni unitarie, è espresso da:

$$S_m = \frac{\sum_i |\xi_i - m|}{n} .$$

Per distribuzioni di frequenze si ha:

$$S_m = \frac{\sum_t |x_t - m| \cdot n_t}{n} = \sum_t |x_t - m| \cdot f_t = \frac{\sum_t |x_t - m| \cdot p_t}{100} .$$

Il calcolo degli scostamenti medi non pone problemi di calcolo nuovi rispetto a quelli già affrontati per la media aritmetica.

a) Distribuzione unitaria

Si voglia calcolare la variabilità della distribuzione dei voti di 10 studenti all'esame di statistica

28 21 24 24 20 25 27 27 24 24,

utilizzando $S_\mu$, $\sigma$ e $S_m$.

Il voto medio della distribuzione è 24,4 , la mediana è 24.

È opportuno disporre i calcoli nel modo seguente:

Tab. 2 - Prospetto per il calcolo di $S_\mu$, $\sigma$ e $S_m$
Distribuzione unitaria

| Voto all'esame di statistica $\xi_i$ | $\lvert\xi_i - 24{,}4\rvert$ | $(\xi_i - 24{,}4)^2$ | $\lvert\xi_i - 24\rvert$ |
|---|---|---|---|
| 20 | 4,4 | 19,36 | 4 |
| 21 | 3,4 | 11,56 | 3 |
| 24 | 0,4 | 0,16 | 0 |
| 24 | 0,4 | 0,16 | 0 |
| 24 | 0,4 | 0,16 | 0 |
| 24 | 0,4 | 0,16 | 0 |
| 25 | 0,6 | 0,36 | 1 |
| 27 | 2,6 | 6,76 | 3 |
| 27 | 2,6 | 6,76 | 3 |
| 28 | 3,6 | 12,96 | 4 |
| Totale | 18,8 | 58,4 | 18 |

Si ha pertanto:

$$S_\mu = \frac{18{,}8}{10} = 1{,}88$$

$$\sigma = \sqrt{\frac{58{,}4}{10}} = 2{,}42$$

$$S_m = \frac{18}{10} = 1{,}8.$$

$S_\mu$ indica che mediamente nella distribuzione un voto differisce dalla media aritmetica di $\pm$1,88. $\sigma$, "potenziando" gli scarti grandi e "sminuendo" i piccoli, informa che un voto differisce dalla media aritmetica di $\pm$2,42. $S_m$, invece, indica che mediamente un voto differisce dalla mediana di $\pm$1,8. Tutte le misure di variabilità sono espresse in 30-esimi.

b) Distribuzione di frequenze per un carattere discreto con modalità espresse singolarmente

La distribuzione unitaria del voto all'esame di statistica dei 10 studenti, già trattata al punto a), può essere trasformata in distribuzione di frequenze. Di tale distribuzione sono già state calcolate la media aritmetica e la mediana. Per determinare $S_\mu$, $\sigma$ e $S_m$ è opportuno predisporre i calcoli nel prospetto seguente:

Tab. 3 - Prospetto per il calcolo di $S_\mu$, $\sigma$ e $S_m$
Distribuzione di frequenze

| Voto all'esame di statistica $x_t$ | Frequenza $n_t$ | $\lvert x_t-24{,}4 \rvert$ | $(x_t-24{,}4)^2$ | $\lvert x_t-24 \rvert$ | $\lvert x_t-24{,}4 \rvert n_t$ | $(x_t-24{,}4)^2 n_t$ | $\lvert x_t-24 \rvert n_t$ |
|---|---|---|---|---|---|---|---|
| 20 | 1 | 4,4 | 19,36 | 4 | 4,4 | 19,36 | 4 |
| 21 | 1 | 3,4 | 11,56 | 3 | 3,4 | 11,56 | 3 |
| 24 | 4 | 0,4 | 0,16 | 0 | 1,6 | 0,64 | 0 |
| 25 | 1 | 0,6 | 0,36 | 1 | 0,6 | 0,36 | 1 |
| 27 | 2 | 2,6 | 6,76 | 3 | 5,2 | 13,52 | 6 |
| 28 | 1 | 3,6 | 12,96 | 4 | 3,6 | 12,96 | 4 |
| Totale | 10 | | | | 18,8 | 58,4 | 18 |

Si ha pertanto:

$$S_\mu,=\frac{18{,}8}{10}=1{,}88$$

$$\sigma=\sqrt{\frac{58{,}4}{10}}=2{,}42$$

$$S_m=\frac{18}{10}=1{,}8.$$

Le misure di variabilità assumono, ovviamente, gli stessi valori già ottenuti per la distribuzione unitaria, con gli stessi significati.

c) Distribuzione di frequenze rispetto ad un carattere con modalità espresse in classi

Si riprenda ora la tab. 4, che si riferisce alla distribuzione degli studenti universitari per classe di voto alla maturità e si calcolino le misure di variabilità $S_\mu$, e $\sigma$.

Tab. 4 - Studenti presenti alla lezione di statistica del 6-10-1994
per classe di voto alla maturità
(prospetto di calcolo della media aritmetica)

| Classe di voto | $n_t$ | $\frac{{}_tc_i+{}_tc_s}{2}={}_tx$ | $n_t \cdot {}_tx$ |
|---|---|---|---|
| 36 | 5 | 36 | 180 |
| 37 $\vdash\!\dashv$ 43 | 32 | 40 | 1.280 |
| 44 $\vdash\!\dashv$ 54 | 44 | 49 | 2.156 |
| 55 $\vdash\!\dashv$ 59 | 10 | 57 | 570 |
| 60 | 17 | 60 | 1.020 |
| Totale | 108 | | 5.206 |

E' necessario, anche in questo caso, ricorrere ai valori centrali di ogni classe. I calcoli si disporranno come nella tabella che segue.

Tab. 5 - Prospetto per il calcolo di $S_\mu$ e $\sigma$
Distribuzione di frequenze per carattere con modalità espresse in classi

| Classe di voto | $n_t$ | $_tx$ | $\lvert {_tx}-48{,}20 \rvert n_t$ | $(_tx-48{,}20)^2 n_t$ |
|---|---|---|---|---|
| 36 | 5 | 36 | 61 | 744,2 |
| 37⊢⊣43 | 32 | 40 | 262,4 | 2.151,68 |
| 44⊢⊣54 | 44 | 49 | 35,2 | 28,16 |
| 55⊢⊣59 | 10 | 57 | 88 | 774,4 |
| 60 | 17 | 60 | 200,6 | 2.367,08 |
| Totale | 108 | | 647,2 | 6.065,52 |

Si ottiene pertanto:

$$S_\mu = \frac{647{,}8}{108} = 6$$

$$\sigma = \sqrt{\frac{6065{,}52}{108}} = 7{,}49.$$

Se anziché operare sulla distribuzione di frequenze si fosse operato sulla distribuzione unitaria, i valori di $S_\mu$ e $\sigma$ sarebbero stati diversi, poiché la distribuzione unitaria e quella di frequenze danno informazioni diverse, e la costruzione delle classi riduce l'informazione e quindi la variabilità. In effetti, rispetto alla distribuzione unitaria, si ha:

$$S_\mu = 6{,}88$$
$$\sigma = 7{,}8.$$

Entrambi gli scostamenti sono stati calcolati tenendo presente che la media del voto alla maturità in tale distribuzione è 48,39.

**5.3.2.2 Proprietà degli scostamenti medi. Relazioni con la media aritmetica**

Per la proprietà dei valori medi potenziati di ordine r di crescere al crescere dell'ordine, fra lo scostamento semplice medio dalla media aritmetica e lo scostamento quadratico medio dalla stessa media vale la relazione:

$$S_\mu \leq \sigma.$$

Poiché inoltre si può dimostrare che lo scostamento semplice medio dalla mediana gode della proprietà di essere il minimo fra tutti gli scostamenti semplici al variare del valore medio di riferimento, fra gli scostamenti medi introdotti vale la relazione:

$$S_m \leq S_\mu \leq \sigma.$$

Il rispetto della diseguaglianza indicata è verificabile rapidamente negli esempi sviluppati nel paragrafo precedente.

Fra gli scostamenti proposti $\sigma$ è il più grande, eppure è la più utilizzata e la più nota misura di variabilità. A giustificazione di ciò va tenuta presente l'importante proprietà dello scostamento quadratico medio dalla media aritmetica di essere il minimo tra gli scostamenti quadratici da qualsiasi altro termine di riferimento. La prova della proprietà è rapida. Sia h un qualsiasi termine di confronto per le modalità del carattere, lo scostamento quadratico da h ($^2S_h$), in distribuzioni di frequenze, è espresso da:

$$^2S_h = \sqrt{\frac{\sum_t (x_t - h)^2 n_t}{n}} = \sqrt{\sum_t (x_t - h)^2 f_t}\ .$$

Operando, per semplicità, sul quadrato dello scostamento si ottiene:

$$
{}^2S_h^2 = \sum_t (x_t - h)^2 f_t = \sum_t (x_t - \mu + \mu - h)^2 f_t = \sum_t \left[(x_t - \mu) + (\mu - h)\right]^2 f_t = \\
\sum_t (x_t - \mu)^2 f_t + 2 \cdot (\mu - h) \sum_t (x_t - \mu) \cdot f_t + (\mu - h)^2 .
$$

Nell'ultimo termine dell'uguaglianza, il doppio prodotto si elimina per la proprietà della media aritmetica di annullare la somma degli scarti e perciò si ha:

$$
{}^2S_h^2 = \sum_t (x_t - \mu)^2 f_t + (\mu - h)^2 ,
$$

ossia:

$$
{}^2S_h^2 = \sigma^2 + (\mu - h)^2 .
$$

In tale relazione l'unico termine che dipende da h è $(\mu - h)^2$ che si annulla per $\mu = h$, perciò il minimo di ${}^2S_h{}^2$ è $\sigma^2$ .

La media aritmetica, che rende nulla la somma degli scarti e minima la somma degli scarti al quadrato, può perciò essere intesa come il miglior parametro di centralità e $\sigma$ come la migliore misura di variabilità corrispondente (va però osservato che ciò è vero in ordine 2, l'ordine della geometria euclidea).

Lo scostamento quadratico medio si può esprimere utilizzando i valori medi potenziati di ordine r=1 ed r=2. E' infatti:

$$
\sigma^2 = \frac{\sum_t (x_t - \mu)^2 n_t}{n} = \sum_t (x_t - \mu)^2 f_t = \sum_t x_t^2 f_t - 2\mu \cdot \sum_t x_t f_t + \mu^2 \sum_t f_t = \sum_t x_t^2 f_t - 2\mu^2 + \mu^2 = \sum_t x_t^2 f_t - \mu^2 ,
$$

ossia:

$$
\sigma = \sqrt{M_2^2 - \mu^2} ,
$$

dove $M_2$ è la media quadratica della distribuzione ed $M_2^2$ è il suo quadrato. Si dice perciò che lo scostamento quadratico medio dalla media aritmetica è la radice quadrata del quadrato della media quadratica meno il quadrato della media aritmetica. Questa proprietà è utilizzata per il calcolo di $\sigma$ ed è quella adottata, ad esempio, dalle calcolatrici tascabili per ottenere lo scostamento.

La traslazione dell'origine non modifica lo scostamento, infatti non modifica la posizione reciproca dei punti sul segmento (cfr. Fig. 1). Una traslazione in $x_o$ trasforma ciascuna modalità in $x_t + x_o$, lo scostamento quadratico medio dalla media aritmetica dello scarto trasformato (${}_o\sigma$) diviene allora:

$$
{}_o\sigma = \sqrt{\frac{\sum_t \left[x_t + x_o - (\mu + x_o)\right]^2 n_t}{n}} = \sqrt{\frac{\sum_t (x_t - \mu)^2 n_t}{n}} = \sigma .
$$

La variazione dell'unità di misura influisce sullo scostamento. Una variazione dell'unità di misura trasforma $x_t$ in c $x_t$, con c positivo, di conseguenza, con riferimento allo scostamento della distribuzione il cui carattere ha mutato unità di misura (${}_m\sigma$) si ha:

$$
{}_m\sigma = \sqrt{\frac{\sum_t (cx_t - c\mu)^2 n_t}{n}} = c\sqrt{\frac{\sum_t (x_t - \mu)^2 n_t}{n}} = c\sigma .
$$

Per effetto di una trasformazione lineare la modalità $x_t$ diviene: $cx_t + x_o$ e lo scostamento del carattere trasformato linearmente (${}_l\sigma$) diviene:

$$
{}_l\sigma = c\sigma .
$$

Rispetto a quella particolare trasformazione lineare che è la standardizzazione, che come detto fa passare da $x_t$ a $\frac{x_t - \mu}{\sigma} = z_t$ , essendo immediato scrivere $z_t = \frac{x_t}{\sigma} - \frac{\mu}{\sigma}$ , per la relazione precedente, si ha

${}_{l}\sigma = \frac{\sigma}{\sigma} = 1$. In conclusione la standardizzazione di un carattere quantitativo consente di ottenere una nuova distribuzione di media 0 e $\sigma = \sigma^2 = 1$.

**5.3.2.3. Varianza e devianza**

Lo scostamento quadratico medio dalla media aritmetica è stato indicato con σ. Il suo quadrato, $\sigma^2$, è detto varianza (Var). A sua volta il numeratore della varianza è noto come devianza (Dev). E' opportuno esplicitare le relazioni indicate. Si ha:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\text{Var}} = \sqrt{\frac{\text{Dev}}{n}} = \sqrt{\frac{\sum_t (x_t - \mu)^2 n_t}{n}} .$$

Segue anche che:

$$\sigma^2 = \text{Var} = M_2^2 - \mu^2 = \frac{\sum_t x_t^2}{n} - \mu^2 ,$$

e quindi:

$$\text{Dev} = n\text{Var} = n\sigma^2 = \sum_t x_t^2 n_t - n\mu^2 .$$

Per modificazioni dell'unità di misura la varianza ${}_c\sigma^2$ diviene: ${}_c\sigma^2 = c^2\ \sigma^2$, ed analogamente per trasformazioni lineari la varianza della distribuzione il cui carattere è stato trasformato diviene:

$${}_l\sigma^2 = c^2\ \sigma^2 .$$

**5.3.2.4. Informazioni sulla distribuzione fornite dalla conoscenza di μ e σ**

Come si è già osservato, μ è una sintesi di n osservazioni del carattere quantitativo X effettuate su un collettivo di n unità statistiche. Riferendosi per semplicità alla distribuzione unitaria e ponendo le osservazioni su di un segmento avente per estremi $\xi_{(1)}$ ed $\xi_{(n)}$, μ è un punto dell'intervallo $\xi_{(1)} \vdash\dashv \xi_{(n)}$. σ, invece, lunghezza media - in termini di media quadratica - dei segmenti $\xi_i \vdash\dashv \mu$ e $\mu \vdash\dashv \xi_i$, è rappresentabile con un segmento che ha un estremo in μ e l'altro in μ+σ o in μ - σ. Noti μ e σ, si può perciò costruire un intorno di μ di ampiezza kσ, con k positivo, in modo che $|\xi_i - \mu| < k\sigma$ oppure $|\xi_i - \mu| \geq k\sigma$. Nota la distribuzione del carattere X, la frequenza assoluta o relativa con cui si presentano osservazioni che rispettano l'una o l'altra diseguaglianza si ottiene immediatamente e con esattezza. Così rispetto alla distribuzione degli studenti per voto alla maturità (tab. 1) si trova che μ=48,39/60, mentre σ è 7,8/60.

Posto k=1,4 si ha:

$$\text{freq}\{|\xi_i - 48{,}39| < 1{,}4 \cdot 7{,}8\} = 1 - \frac{23}{108} = 0{,}79 ,$$

dove con la notazione freq si indica la frequenza relativa. Si può perciò dire che il 79% degli studenti ha un voto di maturità che appartiene all'intervallo 37,47–59,31.

Quando la distribuzione non è nota, un problema analogo viene risolto, per k>1, grazie alla diseguaglianza di Bienaymé-Cebycev. In questo caso, tuttavia, la mancanza di informazioni sulla forma della distribuzione, non consente più di ottenere come soluzione un unico valore, bensì un intervallo di valori plausibili per la frequenza relativa cercata. La diseguaglianza di Bienaymé-Cebycev con riferimento alla distribuzione del carattere X dice che per k>1:

$$\text{freq}\{|\xi_i - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

e pertanto:

$$\text{freq}\{|\xi_i - \mu| < k\sigma\} \geq 1 - \frac{1}{k^2} .$$

Per la prima diseguaglianza le osservazioni non interne all'intorno di μ di raggio kσ sono tutt'al più $1/k^2$. Poiché la risposta individua l'intervallo $0 \vdash\!\dashv 1/k^2$, tanto più il rapporto $1/k^2$ è piccolo, tanto più alto è il valore, l'importanza, dell'informazione ottenuta.

Per la seconda diseguaglianza le osservazioni interne all'intorno di μ di raggio kσ sono almeno $1-1/k^2$. Poiché la risposta fornisce l'intervallo $1-1/k^2 \vdash\!\dashv 1$, l'informazione ottenuta è tanto migliore quanto più $1-1/k^2$ è grande.

Se rispetto alla distribuzione degli studenti per voto di maturità si ammette di sapere solamente che μ=48,39 e σ=7,8, posto k=1,4, si può calcolare qual'è la frequenza minima delle osservazioni interne all'intervallo 37,47–59,31. Si ottiene così che la frequenza cercata è $1-(1/1,4^2)=0,49$, perciò la frequenza effettiva appartiene all'intervallo $0,49 \vdash\!\dashv 1$. Tale intervallo comprende in effetti il risultato precedentemente ottenuto sulla distribuzione osservata.

Da quanto detto emerge che quanto più si considera un piccolo intorno di μ, tanto più è ampio l'intervallo che contiene le frequenze relative delle osservazioni non interne all'intorno e ridotto è il valore dell'informazione ottenuta (inoltre non esiste risposta per k=1). Al contrario, quanto più l'intorno di μ è grande, tanto meno frequenti sono le osservazioni non interne all'intorno, e quindi l'intervallo fornito come risposta è piccolo e il valore dell'informazione ottenuta è grande.

**5.3.2.5. Coefficiente di variazione**

In base alle considerazioni fin qui svolte, σ si può intendere come una misura del rischio di vedere i valori rilevati allontanarsi dalla loro media. Tale rischio è tanto più alto quanto più σ è "grande" e viceversa. Resta da capire come fare a misurare la "grandezza" di σ. In effetti σ dipende dall'ordine medio di grandezza del carattere, ossia dalla media. Se ad esempio μ è dell'ordine delle unità, non ci si attende che σ sia dell'ordine delle migliaia; così come se μ è dell'ordine dei milioni, non ci si attende che σ sia dell'ordine delle decine. Utilizzando μ come unità di misura di σ si ha il coefficiente di variazione. Esso espresso nella forma di percentuale è:

$$C = \frac{\sigma}{\mu}100$$

L'indice ha il minimo in 0 - che si realizza se e solo se le unità presentano tutte la stessa modalità, diversa da 0 - ma il suo massimo non è definito. C esprime σ come percentuale rispetto alla media ed è adimensionale. Fra due distribuzioni con lo stesso σ quella meno variabile è la distribuzione con μ maggiore.

BIBLIOGRAFIA

M. FRAIRE (1994), *Metodi di analisi multidimensionale dei dati. Aspetti statistici ed applicazioni informatiche*, CISU, Roma.

G. LETI (1983), *Statistica descrittiva*, il Mulino, Bologna, 1983.

M. G. OTTAVIANI (1993), I libri di testo e la statistica: metodi di analisi e confronto di testi. *INDUZIONI***, 6**, pp.93-120.