UNIVERSITA' DEGLI STUDI DI PERUGIA
DIPARTIMENTO DI FILOSOFIA SCIENZE SOCIALI UMANE E DELLA FORMAZIONE
Corso di Laurea in Scienze per l'Investigazione e la Sicurezza

# 5. INDICI DI DISUGUAGLIANZA

Prof. Maurizio Pertichetti

Statistica sociale

# 5. INDICI DI DISUGUAGLIANZA

Un valore centrale, utile a precisare la "posizione" di una distribuzione di frequenze , vale a dire la "tendenza centripeta" dei dati, da solo tuttavia non è sufficiente a caratterizzarla perché non consente di percepire quali siano le effettive grandezze dei valori da cui esso è ricavato. Si sa che due distribuzioni possono avere uguale media, ma essere composte da valori diversi.

Consideriamo il seguente esempio di tre studenti che hanno superato ciascuno tre esami:

| | | | |
|---|---|---|---|
| A | 24 | 24 | 24 |
| B | 23 | 24 | 25 |
| C | 18 | 24 | 30 |

è facile vedere che il voto medio e quello mediano per ciascun studente è pari a 24.

Sintetizzando un insieme di dati con un unico valore centrale si vengono in sostanza a perdere informazioni che pure la distribuzione conteneva.

Ne discende allora che è utile un ulteriore elemento che tenga conto anche del modo con il quale i valori si distribuiscono tra le diverse modalità del carattere, ovvero della loro **disuguaglianza**.

Gli **indici di disuguaglianza** ci dicono appunto quanto possono essere diversi i valori di una distribuzione.

I valori assunti dagli indici risultano diversi se si misura la disuguaglianza:

- delle singole modalità rispetto ad un valore centrale. Ovvero si calcola determinando gli scostamenti, o scarti, tra le modalità del carattere e una sua media poi sintetizzati a loro volta come media. I questo caso la disuguaglianza viene intesa come **dispersione**;

- oppure tra tutte le modalità considerate a due a due, ovvero la disuguaglianza reciproca. Si calcola determinando le differenze medie, cioè le differenze in valore assoluto tra le modalità del carattere che poi vengono sintetizzate a loro volta sempre come media. I questo caso la disuguaglianza viene intesa come **variabilità**.

I conseguenti indici che si ottengono, e ciò vale non solo per quelli di disuguaglianza ma in generale per tutti quelli in generale utilizzati dalla statistica, si distinguono in:

- **indici assoluti**, che sono espressi nella stessa unità di misura del fenomeno in esame e non si adattano a consentire confronti ;
- **indici relativi**, tra i quali gli **indici normalizzati**, che prescindono dall'unità di misura del fenomeno esaminato e sono particolarmente adatti per effettuare confronti tra distribuzioni diverse. Si ottengono rapportando un indice assoluto ad una media o al suo massimo ed assumono valori compresi tra 0 ed 1.

Caratteristiche di un indice di disuguaglianza :

- assumere solo valori positivi perchè non ha senso parlare di dispersione o variabilità negative o valori nulli quando tutti i termini della distribuzione sono uguali fra loro $X_1 = X_2 = \ldots = X_k$ ;
- assumere un livello minimo quando tutti i casi sono uguali e un livello massimo quando ogni caso è differente dagli altri, ovvero registrare valori crescenti all'aumentare della disuguaglianza ciò in quanto una misura di questa deve essere tanto più grande quanto maggiore è la differenza fra i dati.

**Il campo di variazione**

Un indice assoluto della variabilità di una successione di dati, di immediata percezione e assai semplice da calcolarsi, è rappresentato dal **campo di variazione** (o **range**), che è dato dalla differenza tra il valore massimo e il valore minimo della successione. Di fatto costituisce l'ampiezza dell'intervallo dei dati. In simboli:

$$\omega = x_{max} - x_{min}$$

L'indice in questione è poco utilizzato in quanto prende in considerazione solo la dispersione esistente tra i valori estremi della distribuzione, per cui, oltre a non tener conto di tutte le informazioni su una variabile statistica, risente di eventuali valori anomali nei dati.

Data la seguente serie: 1 2 3 6 9 10 15
- Il valore più alto è 15, il più basso 1
- Il range è dato dalla differenza tra i due valori R = 15-1 = 14

Data la seguente serie: -11 -2 3 9 10 18
- Il valore più alto è 18, il più basso -11
- Il range è dato dalla differenza tra i due valori R = 18 - (-11) = 18 + 11 = 29

Il campo di variazione, che è espresso nella stessa unità di misura dei dati, tanto più è piccolo tanto più i dati sono concentrati, viceversa tanto più è grande tanto più i dati sono dispersi.

**Gli scostamenti medi**

Questi indici di variabilità, anche se poco usati nella pratica, coinvolgono nel loro calcolo tutte le determinazioni della variabile considerata e sono: lo **scostamento medio dalla media aritmetica** e lo **scostamento medio dalla mediana**.

Questi indici sono calcolati considerando i valori assoluti degli scarti, in quanto nel caso della media aritmetica, come sappiamo, la media degli scarti, presi con il loro segno, è zero.

Lo **scostamento medio dalla media aritmetica** è un indice di variabilità dato dalla media aritmetica dei valori assoluti degli scarti dalla media aritmetica, ovvero a seconda che si abbia una successione di dati o una distribuzione di frequenza:

$$S_\mu = \frac{\sum_{i=1}^{k} |x_i - \mu|}{n} \qquad S_\mu = \frac{\sum_{i=1}^{k} |x_i - \mu| n_i}{n}$$

Esempio di calcolo dello scostamento medio dalla media aritmetica

Data la seguente successione: 4 5 15 23 28

$$\mu = \frac{4 + 5 + 15 + 23 + 28}{5} = \frac{75}{5} = 15{,}00$$

$$S_\mu = \frac{|4-15| + |5-15| + |15-15| + |23-15| + |28-15|}{5} = \frac{42}{5} = 8{,}4$$

Il valore così ricavato indica che i dati della distribuzione si discostano, dalla loro media aritmetica, mediamente di 8,4 unità in più o in meno.

**La varianza, lo scostamento quadratico medio e la devianza**

La **varianza** di un insieme di dati o di una distribuzione di frequenza è una misura di dispersione che si ottiene come **media dei quadrati degli scarti dalla media aritmetica**. In simboli, ovvero a seconda che si abbia una successione di dati o una distribuzione di frequenza:

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{n} \qquad \sigma^2 = \frac{\sum_{i=1}^{k} (xi - \mu)^2 n_i}{n}$$

E' anche possibile usare una formula semplificata ottenuta sviluppando la precedente (sviluppo che qui si omette):

$$\sigma^2 = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - \mu^2$$

La varianza presenta, tuttavia, un notevole inconveniente nel senso che è espressa attraverso il quadrato dell'unità di misura delle osservazioni, per cui se le osservazioni ad esempio sono in metri, la varianza è espressa in metri al quadrato. Motivo per cui non è mai possibile rappresentare su uno stesso diagramma la varianza e la distribuzione delle osservazioni.

Per ovviare all'inconveniente anzidetto, si preferisce usare la radice quadrata della varianza e ottenere un importante indice di variabilità, tra tutti il più utilizzato, denominato **scostamento quadratico medio** o **deviazione standard**. In simboli, ovvero a seconda che si abbia una successione di dati o una distribuzione di frequenza:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \mu)^2}{n}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{k} (xi - \mu)^2 n_i}{n}}$$

Lo scostamento quadratico medio o deviazione standard è un indice altamente rappresentativo del maggiore o minore addensamento dei dati intorno al loro valore medio. E' preferibile in generale allo scostamento semplice medio perché da risalto e permette di considerare anche le variazioni più piccole delle distribuzioni, ovvero rappresenta una misura della variabilità più sensibile.

Da ultimo infine consideriamo il numeratore della varianza in quanto si presenta come un'altra misura della dispersione denominata **devianza**. Per un carattere X la sua espressione analitica, a seconda che si abbia una successione di dati o una distribuzione di frequenza::

$$D(X) = \sum_{i=1}^{k} (x_i - \mu)^2 \qquad D(X) = \sum_{i=1}^{k} (xi - \mu)^2 n_i$$

Esempio di calcolo dello scostamento medio dalla media aritmetica

Data la seguente successione:

3 5 15 23 28

$$\mu = \frac{3 + 5 + 15 + 23 + 28}{5} = \frac{74}{5} = 14{,}80$$

| $x_i$ | $x_i - \mu$ | $(x_i - \mu)^2$ |
|---|---|---|
| 3 | -11,8 | 139,24 |
| 5 | -9,8 | 96,04 |
| 15 | 0,2 | 0,04 |
| 23 | 8,2 | 67,24 |
| 28 | 13,2 | 174,24 |
| totale | | 476,8 |

$$\sigma^2 = \frac{D(X)}{n} = \frac{476,8}{5} = 95,36$$

$$\sigma = \sqrt{\sigma^2} = 9,77$$

Data la seguente successione:

3,9 8,9 4,8 5,0 9,9

$$\mu = \frac{3,9 + 8,9 + 4,8 + 5,0 + 9,9}{5} = \frac{32,5}{5} = 6,50$$

| $x_i$ | $x_i - \mu$ | $(x_i - \mu)^2$ |
|---|---|---|
| 3,9 | -2,6 | 6,76 |
| 8,9 | 2,4 | 5,76 |
| 4,8 | -1,7 | 2,89 |
| 5 | -1,5 | 2,25 |
| 9,9 | 3,4 | 11,56 |
| totale | | 29,22 |

$$\sigma^2 = \frac{D(X)}{n} = \frac{29,22}{5} = 5,84$$

$$\sigma = \sqrt{\sigma^2} = 2,42$$

Data la seguente distribuzione di frequenze:

| $x_i$ | $n_i$ |
|---|---|
| 173 | 14 |
| 178 | 18 |
| 183 | 28 |
| 188 | 33 |
| 193 | 17 |
| 198 | 15 |
| tot | 125 |

$\mu = 186$

| $x_i - \mu$ | $(x_i - \mu)^2$ | $(x_i - \mu)^2 n_i$ |
|---|---|---|
| -12,64 | 159,77 | 2.236,77 |
| -7,64 | 58,37 | 1.050,65 |
| -2,64 | 6,97 | 195,15 |
| 2,36 | 5,57 | 183,80 |
| 7,36 | 54,17 | 920,88 |
| 12,36 | 152,77 | 2.291,54 |
| totale | | 6.878,80 |

| $x_i^2 n_i$ |
|---|
| 419.006 |
| 570.312 |
| 937.692 |
| 1.166.352 |
| 633.233 |
| 588.060 |
| 4.314.655 |

$$\sigma^2 = \frac{D(X)}{n} = \frac{6.878,80}{125} = 55,03$$

$$\sigma^2 = \frac{\Sigma x_i^2 n_i}{n} - \mu^2 = \frac{4.314.655}{125} - 34.462 = 55,03 \qquad \sigma = \sqrt{\sigma^2} = 7,418$$

Data la seguente distribuzione di frequenze:

| $x_i$ | $n_i$ |
|---|---|
| 1 | 60 |
| 2 | 80 |
| 3 | 30 |
| 4 | 25 |
| 5 | 5 |
| tot | 200 |

| $x_i n_i$ | $x_i - \mu$ | $(x_i - \mu)^2$ | $(x_i - \mu)^2 n_i$ |
|---|---|---|---|
| 60 | -1,18 | 1,38 | 82,84 |
| 160 | -0,18 | 0,03 | 2,45 |
| 90 | 0,83 | 0,68 | 20,42 |
| 100 | 1,83 | 3,33 | 83,27 |
| 25 | 2,83 | 7,98 | 39,90 |
| 435 | totale | | 228,88 |

| $x_i^2 n_i$ |
|---|
| 60 |
| 320 |
| 270 |
| 400 |
| 125 |
| 1.175 |

$\mu = 435 / 200 = 2,18$

$$\sigma^2 = \frac{D(X)}{n} = \frac{228,88}{200} = 1,14 \qquad \sigma^2 = \frac{\Sigma x_i^2 n_i}{n} - \mu^2 = \frac{1.175}{200} - 4,73 = 1,14$$

$$\sigma = \sqrt{\sigma^2} = 1,070$$

**Le differenze medie**

A volte, per indicare la variabilità dei dati, può essere più utile far riferimento alle differenze esistenti fra l'uno e l'altro dato che non ai loro scostamenti da un valore medio.

Le **differenze medie** sono indici di mutua variabilità che esaminano le differenze esistenti, in valore assoluto, fra ciascun dato e tutti gli altri $|x_i - x_j|$ e ne operano una sintesi tramite una opportuna media.

Le distanze sono calcolate in valore assoluto, in quanto considerando i valori algebrici la sommatoria si annullerebbe.

Quando il confronto tra tutte le modalità è fatto non considerando la differenza di ciascuna modalità con se stessa, per cui i confronti sono **n(n-1)**, e la sintesi è effettuata facendo la media arimetica delle differenze si ottiene la **differenza semplice media**:

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{n(n-1)}$$

Quando il confronto tra tutte le modalità è fatto tenendo conto anche della differenza di una modalità con se stessa, per cui i confronti sono $\mathbf{n^2}$, e la sintesi è effettuata facendo la media arimetica delle differenze allora si parla **differenza semplice media con ripetizione**:

$$\Delta_R = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{n^2}$$

I due indici sono legati dalla relazione $\Delta = \Delta_R\,(n/n-1)$ per cui ogni volta che si calcolerà l'uno il risultato potrà estendersi all'altro.

Ai fini del calcolo le due formule possono essere scritte:

$$\Delta = \frac{4\sum_{j=1}^{n}(ix_i)}{n(n-1)} - \frac{2\,\mu(n+1)}{(n-1)} \qquad \Delta_R = \frac{4\sum_{j=1}^{n}(ix_i)}{n^2} - \frac{2\,\mu(n+1)}{n}$$

La differenza semplice assume:

- **valore minimo** quando tutti i dati sono uguali tra loro per cui $\mathbf{\Delta_{min} = 0}$ ;
- **valore massimo** quando n-1 unità statistiche non possiedono alcuna modalità del carattere mentre l'n-esima possiede l'intero ammontare del carattere che è uguale a nμ, per cui si ha $\mathbf{\Delta_{max} = 2\mu}$

Per distribuzioni di frequenza la differenza semplice media media può essere calcolata applicando le formule:

$$\Delta = \frac{2}{n(n-1)}\sum_{i=1}^{k-1} n_i'(n-n_i')(x_{i+1}-x_i) \qquad \Delta_R = \frac{2}{n^2}\sum_{i=1}^{k-1} n_i'(n-n_i')(x_{i+1}-x_i)$$

Esempio di calcolo della differenza semplice media.

Data la seguente successione:

33 45 98 55 24

| $x_i$ | i | $ix_i$ |
|---|---|---|
| 24 | 1 | 24 |
| 33 | 2 | 66 |
| 45 | 3 | 135 |
| 55 | 4 | 220 |
| 98 | 5 | 490 |
| 255 | | 935 |

$$\mu = \frac{255}{5} = 51$$

$$\Delta = \frac{4 \sum_{j=1}^{n} (ix_i)}{n(n-1)} - \frac{2\mu(n+1)}{(n-1)} = \frac{4 * 935}{5*4} - \frac{2*51*6}{4} = 187 - 153 = 34{,}0$$

$$\Delta_R = \frac{4 \sum_{j=1}^{n} (ix_i)}{n^2} - \frac{2\mu(n+1)}{n} = \frac{4 * 935}{5*5} - \frac{2*51*6}{5} \quad 150 - 122 = 27{,}2$$

**Il coefficiente di variazione**

In realtà operare confronti sulla deviazione standard non è di grande aiuto, perché essa dipende fortemente dalla media dei dati su cui è stata calcolata. In questo senso un indice relativo assai utilizzato, se i valori della distribuzione sono positivi o almeno la media risulta maggiore di zero, è il **coefficiente di variazione**, definito come rapporto tra scostamento quadratico medio o deviazione standard e media aritmetica. E' in sostanza un numero puro, non espresso in alcuna unità di misura, che consente di effettuare confronti fra distribuzioni diverse per fenomeni omogenei. La sua espressione analitica è:

$$Cv = \frac{\sigma}{\mu} = \frac{\sqrt{\frac{\Sigma (x_i - \mu)^2 n_i}{n}}}{\mu}$$

Questo coefficiente di variazione, espresso in genere in termini percentuali moltiplicando Cv per 100, è indipendente dall'unità di misura, ovvero è un numero puro utilizzato, sia per misurare la variazione media del fenomeno in rapporto alla sua media aritmetica, sia per confrontare la variabilità relativa di un fenomeno in circostanze differenti (ad esempio, la variabilità della distribuzione per età tra le varie regioni, la distribuzione dei redditi per paesi e per anno, la variabilità del peso rispetto al sesso, ...).

Il coefficiente di variazione, inoltre, è necessario come già detto tutte le volte che si intende confrontare la variabilità di due fenomeni espressi in unità di misure diverse (ad esempio, la variabilità del peso rispetto a quella dell'altezza, la variabilità dei consumi di carburante rispetto alla variabilità dell'usura dei pneumatici per una determinata marca di autovetture, ecc.).

Esempio di calcolo del coefficiente di variazione

Date le seguenti distribuzioni:

| $x_i$ | $n_i$ | $x_i*n_i$ | $(x_i-\mu)$ | $(x_i-\mu)^2$ | $(xi-\mu)^2 n_i$ |
|---|---|---|---|---|---|
| 1 | 2 | 2 | -2 | 4 | 8,00 |
| 2 | 2 | 4 | -1 | 1 | 2,00 |
| 3 | 2 | 6 | 0 | 0 | 0,00 |
| 4 | 2 | 8 | 1 | 1 | 2,00 |
| 5 | 2 | 10 | 2 | 4 | 8,00 |
| | 10 | 30 | | | 20,00 |

$\mu = 30/10 =$ 3

SQM = $\sigma = \sqrt{\Sigma(xi-\mu)^2 n_i / n} =$ 1,414

Cv = $\sigma/\mu =$ 0,471 (*100 = 47,1 % )

| $x_i$ | $n_i$ | $x_i*n_i$ | $(x_i-\mu)$ | $(x_i-\mu)^2$ | $(xi-\mu)^2 n_i$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | -2 | 4 | 0,00 |
| 2 | 0 | 0 | -1 | 1 | 0,00 |
| 3 | 10 | 30 | 0 | 0 | 0,00 |
| 4 | 0 | 0 | 1 | 1 | 0,00 |
| 5 | 0 | 0 | 2 | 4 | 0,00 |
| | 10 | 30 | | | 0,00 |

$\mu = 30/10 =$ 3

SQM = $\sigma = \sqrt{\Sigma(xi-\mu)^2 n_i / n} =$ 0,000

Cv = $\sigma/\mu =$ 0,000 (*100 = 0 % )

Date le seguenti distribuzioni :

Un fenomeno riferito ai maschi

| $x_i$ | $n_i$ | $x_i n_i$ | $x_i-\mu$ | $(x_i-\mu)^2$ | $(x_i-\mu)^2 n_i$ |
|---|---|---|---|---|---|
| 34,5 | 81 | 2.794,5 | -11,4 | 129,2 | 10.464,3 |
| 44,5 | 31 | 1.379,5 | -1,4 | 1,9 | 57,9 |
| 54,5 | 36 | 1.962,0 | 8,6 | 74,5 | 2.683,6 |
| 64,5 | 35 | 2.257,5 | 18,6 | 347,2 | 12.152,8 |
| | 183 | 8.393,5 | | | 25.358,5 |

$\mu =$ 45,866 $\sigma = \sqrt{\frac{\Sigma (x_i - \mu)^2 n_i}{n}} = \frac{25.358,5}{183} =$ 11,772

$Cv = \frac{\sigma}{\mu} = \frac{11,772}{45,866} =$ 0,257 (*100 = 25,7 % )

Stesso fenomeno riferito alle femmine

| $x_i$ | $n_i$ | $x_i n_i$ | $x_i-\mu$ | $(x_i-\mu)^2$ | $(x_i-\mu)^2 n_i$ |
|---|---|---|---|---|---|
| 34,5 | 30 | 1.035,0 | -18,5 | 343,2 | 10.294,8 |
| 44,5 | 42 | 1.869,0 | -8,5 | 72,7 | 3.052,1 |
| 54,5 | 36 | 1.962,0 | 1,5 | 2,2 | 78,4 |
| 64,5 | 75 | 4.837,5 | 11,5 | 131,7 | 9.876,4 |
| | 183 | 9.703,5 | | | 23.301,6 |

$\mu =$ 53,025 $\sigma = \sqrt{\frac{\Sigma (x_i - \mu)^2 n_i}{n}} = \frac{23.301,6}{183} =$ 11,284

$Cv = \frac{\sigma}{\mu} = \frac{11,284}{53,025} =$ 0,213 (*100 = 21,3 % )

Per l'esempio fatto i maschi presentano una variazione media del fenomeno intorno alla media aritmetica pari al 25% contro il 21% delle femmine.

Alcuni valori particolari del CV che possono essere utili nello studio di una distribuzione di dati:

- CV = 0 in questo caso la deviazione standard è pari a 0. Tutti i dati sono uguali tra loro e la media può essere considerata come un indice perfetto per rappresentarli.
- CV ≥ 0.5 in questo caso la deviazione standard è più della metà della media. La media, in questo caso, non può essere considerata un buon indice per rappresentare i dati.
- CV ≤ 0.5 in questo caso la deviazione standard è meno della metà della media. La media, in questo caso, può essere considerata un buon indice per rappresentare i dati.

**Indice di eterogeneità**

Per disporre di indici di disuguaglianza utilizzabili con qualsiasi tipo di carattere (anche qualitativo non ordinabile) occorre che la definizione dell'indice coinvolga solo le frequenze delle diverse modalità, senza richiedere relazioni di ordine fra le modalità stesse. Un esempio è fornito dall'**indice di eterogeneità di Gini**.

$$G = 1 - \sum_{i=1}^{k} f_i^2$$

E' un indice assoluto di eterogeneità in quanto è massimo ossia pari a 1-(1/k) quando le modalità hanno tutte la medesima frequenza o, se si vuole, quando le frequenze sono equidistribuite tra tutte le modalita, mentre è minimo (ossia nullo e quindi c'è massima omogeneità) quando tutte le frequenze si addensano in una sola modalità .

Esempio di calcolo dell'indice assoluto

| $x_i$ | $n_i$ | f | $f^2$ |
|---|---|---|---|
| 1 | 2 | 0,20 | 0,04 |
| 2 | 2 | 0,20 | 0,04 |
| 3 | 2 | 0,20 | 0,04 |
| 4 | 2 | 0,20 | 0,04 |
| 5 | 2 | 0,20 | 0,04 |
| | 10 | 1,00 | 0,20 |

| $x_i$ | $n_i$ | f | $f^2$ |
|---|---|---|---|
| 1 | 0 | 0,00 | 0,00 |
| 2 | 0 | 0,00 | 0,00 |
| 3 | 10 | 1,00 | 1,00 |
| 4 | 0 | 0,00 | 0,00 |
| 5 | 0 | 0,00 | 0,00 |
| | 10 | 1,00 | 1,00 |

$G = 1 - \Sigma f^2$ $\quad 1 - 0,20 = 0,80$ $\quad 1 - 1,00 = 0,00$

$k = 5$ $\quad max = 1 - \frac{1}{k} = 0,8$ $\quad min = 0,00$

Per rendere pero confrontabili fra loro due indici calcolati su due diversi caratteri con frequenze $n_i$ diverse delle modalità occorre utilizzare un indice relativo, che si ottiene dividendo l'indice assoluto per il massimo valore che esso può assumere.

Nel caso dell'indice di Gini, essendo $G_{max} = 1-(1/k)$ , l'**indice normalizzato $G_N$** si ottiene:

$$G_N = \frac{G}{G_{max}} = \frac{G}{1-(1/k)} = G \frac{k}{(k-1)}$$

L'indice così ottenuto è un indice relativo che varia tra 0 che è il minimo e 1 che è il massimo.

$$0 \leq G_N \leq 1$$

Riprendendo l'esempio precedente :

- l'indice in corrispondenza della massima eterogeneità era G=0,8 per cui:
  $G_N$ = G * k/(k-1) = 0,80*5/4 = 0,80*1,25 = **1,0**
- l'indice in corrispondenza della minima eterogeneità era G=0 per cui:
  $G_N$ = G * k/(k-1) = 0*5/4 = **0**