

## **6. LA CONCENTRAZIONE**

Prof. Maurizio Pertichetti

Statistica sociale

## 6. LA CONCENTRAZIONE

La concentrazione si definisce come una espressione strettamente collegata al tema della variabilità. Più in particolare si può dire che la concentrazione esprime il significato di variabilità interpretandolo concettualmente come disuguaglianza. Il campo di applicazione riguarda esclusivamente i caratteri quantitativi cosiddetti trasferibili. Un carattere è detto trasferibile se ha senso immaginare che **la sua** intensità globale o una sua parte sia attribuibile (anche solo idealmente) ad una sola o a poche unità. Sono variabili trasferibili: reddito, patrimonio, diritti di possesso, quote di mercato, salari, ecc. (non trasferibili sono quelle che riguardano aspetti intrinseci delle unità: altezza, peso, età, ecc.). Se i valori della variabile sono livelli raggiungibili da qualsiasi unità ed ha un senso la loro somma o aggregazione allora lo studio di concentrazione è plausibile.

La concentrazione si rappresenta attraverso un indice che misura il grado di concentrazione del carattere e una particolare rappresentazione grafica chiamata diagramma di Lorenz.

Dato un carattere X, si può affermare che esso ha un alto grado di concentrazione se molta parte della sua intensità totale è attribuita a pochi fra i casi oggetto di studio del fenomeno. Diversamente si può affermare che esso ha una ridotta concentrazione quando la sua intensità complessiva è equamente ripartita fra tutti i casi. Pertanto, si è in presenza di:

- **concentrazione nulla** (o **equidistribuzione**) quando tutte le unità dispongono in parti uguali dell'intensità totale;
- **concentrazione massima** quando una sola unità dispone interamente dell'intensità totale.

Per esempio, se la ricchezza di una data collettività si trova per molta parte ad essere posseduta da un numero esiguo di individui, mentre tutti gli altri ne sono privi o ne posseggono poca, si dice che quella ricchezza è fortemente concentrata. I dati statistici delle ricchezze individuali presentano in tal caso dei valori nulli, dei valori bassi e dei valori molto elevati, con scarti rilevanti l'uno all'altro e la relativa successione sarà perciò caratterizzata da una grande variabilità.

Per contro, se la ricchezza totale di quella collettività fosse ripartita in parti uguali fra tutti gli individui, i singoli valori della ricchezza individuale formerebbero una successione di dati costanti. Nulla sarebbe in tal caso la variabilità e nulla la concentrazione.

Il ricorso ad un esempio può essere utile ad agevolare la comprensione del concetto e dei diversi passaggi nella elaborazione dei dati. Si immagini che per un certo anno l'ammontare totale della ricchezza prodotta da una collettività formata da 5 famiglie corrisponda al valore di 10.000.000 e che essa si distribuisca tra ciascuna delle suddette famiglie secondo intensità diverse così come riportate nella prima tabella.

| $x_i$<br>ordinati | $n_i$ |
|-------------------|-------|
| 500.000           | 1     |
| 1.000.000         | 1     |
| 1.500.000         | 1     |
| 2.000.000         | 1     |
| 5.000.000         | 1     |
| 10.000.000        | 5     |

si calcolino per i dati di partenza le frequenze cumulate assolute

| $x_i'$<br>cum | $n_i'$<br>cum |
|---------------|---------------|
| 500.000       | 1             |
| 1.500.000     | 2             |
| 3.000.000     | 3             |
| 5.000.000     | 4             |
| 10.000.000    | 5             |

e quindi le frequenze cumulate relative

| $p_i =$<br>$n_i' \text{ cum} / n$ | $q_i =$<br>$x_i' \text{ cum} / q$ |
|-----------------------------------|-----------------------------------|
| 0,200                             | 0,050                             |
| 0,400                             | 0,150                             |
| 0,600                             | 0,300                             |
| 0,800                             | 0,500                             |
| 1,000                             | 1,000                             |

Data la prima tabella, dove la ricchezza è distribuita unitariamente tra le unità statistiche in valore assoluto, la seconda tabella, dove i valori permangono ancora in valore assoluto, ci dice che la prima famiglia nell'anno preso in considerazione possiede l'equivalente di 500.000, le prime due famiglie l'equivalente di 1.500.000, le prime tre famiglie l'equivalente di 3.000.000, ecc., fino al totale di tutte e 5 le famiglie nella cui disponibilità vi è complessivamente l'ammontare corrispondente a 10.000.000.

La terza tabella, che opera la trasformazione dei valori assoluti della seconda tabella in valori relativi uniformando le unità di misura delle  $p_i$  e delle  $q_i$ , così da poterle rappresentare su uno stesso diagramma, ci dice (interpretando i valori come percentuali) che il 20% delle famiglie della suddetta comunità e sempre nell'anno di riferimento possiede l'equivalente dello 0,5% della ricchezza, il 40% delle famiglie l'equivalente dello 0,15%, il 60% delle famiglie l'equivalente del 30%, ecc., fino al 100% delle famiglie cui corrisponde il possesso del 100% dell'intera ricchezza.

Generalizzando:

- Si ipotizzi di avere i valori di  $n$  intensità, ordinate in modo non decrescente, corrispondenti alle modalità di un fenomeno  $X$  (reddito, patrimonio, ricchezza, etc.) che goda della proprietà della trasferibilità:

$$x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n \quad \left( \text{per le quali l'intensità totale è: } q = \sum_{i=1}^n x_i \right)$$

tale per cui si possano presentare le due situazioni limite:

- l'**equidistribuzione**, ovvero le  $n$  intensità sono uguali tra loro :  $x_1 = x_2 = \dots = x_i = \dots = x_n = q/n$  ;
- la **concentrazione massima**, ovvero tutta l'intensità  $q$  è addensata su una singola osservazione, cosicché:  $x_1 = x_2 = \dots = x_i = \dots = x_{n-1} = 0$  e  $x_n = q$ .

- Si assumano le seguenti frazioni come **somme progressive delle frequenze** delle prime  $i$  unità statistiche:

$$p_i = \frac{n_i' \text{ cum}}{n} \quad \text{ovvero} \quad p_1 = \frac{n_1' \text{ cum}}{n} \quad p_2 = \frac{n_2' \text{ cum}}{n} \quad \dots \quad p_n = \frac{n_n' \text{ cum}}{n} = 1 \quad ;$$

- e si assumano le sottostanti frazioni come **somme progressive delle corrispondenti intensità** delle prime  $i$  unità statistiche:

$$q_i = \frac{x_i' \text{ cum}}{q} \quad \text{ovvero} \quad q_1 = \frac{x_1' \text{ cum}}{q} \quad q_2 = \frac{x_2' \text{ cum}}{q} \quad \dots \quad q_n = \frac{x_n' \text{ cum}}{q} = 1 \quad .$$

Ebbene, la somma  $R = \sum_{i=1}^{n-1} (p_i - q_i)$  si può considerare come una misura **assoluta** della concentrazione e il suo valore è:

- = a **0** nel caso di **equidistribuzione**;
- > di **0** nel caso di **concentrazione**;
- = a  $\sum_{i=1}^{n-1} p_i$  od anche a  $\frac{n-1}{2}$  nel caso di **massima concentrazione**  $R_{\max}$ .

Rapportando il valore di **R** al suo massimo si ha l'indice normalizzato **R<sub>N</sub>**, ovvero:

$$R_N = \frac{R}{R_{\max}} = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \quad \text{od anche} \quad \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\frac{n-1}{2}} = \frac{2}{n-1} \sum_{i=1}^{n-1} (p_i - q_i) \quad \text{detto rapporto di concentrazione di Gini.}$$

**R<sub>N</sub>** è un indice che varia tra 0 e 1:

- è = 0 nel caso di equidistribuzione;
- è = 1 nel caso di massima concentrazione.

Riprendendo l'esempio fatto in precedenza si ha:

| $x_i$<br>ordinati | $n_i$ | $x_i'$<br>cum | $n_i'$<br>cum | $p_i =$<br>$n_i' \text{ cum} / n$ | $q_i =$<br>$x_i' \text{ cum} / q$ | $(p_i - q_i)$ |
|-------------------|-------|---------------|---------------|-----------------------------------|-----------------------------------|---------------|
| 500.000           | 1     | 500.000       | 1             | 0,200                             | 0,050                             | 0,150         |
| 1.000.000         | 1     | 1.500.000     | 2             | 0,400                             | 0,150                             | 0,250         |
| 1.500.000         | 1     | 3.000.000     | 3             | 0,600                             | 0,300                             | 0,300         |
| 2.000.000         | 1     | 5.000.000     | 4             | 0,800                             | 0,500                             | 0,300         |
| 5.000.000         | 1     | 10.000.000    | 5             | 1,000                             | 1,000                             | 0,000         |
| 10.000.000        | 5     |               |               |                                   | $\Sigma$                          | 1,000         |

$R = \Sigma (p_i - q_i) = 1,000$  ed è > 0 perché c'è **concentrazione**;

$$R_N = \frac{R}{R_{\max}} = \frac{\Sigma(p_i - q_i)}{\Sigma p_i} \quad \text{oppure} \quad = \frac{2}{n-1} \Sigma(p_i - q_i) = 0,500$$

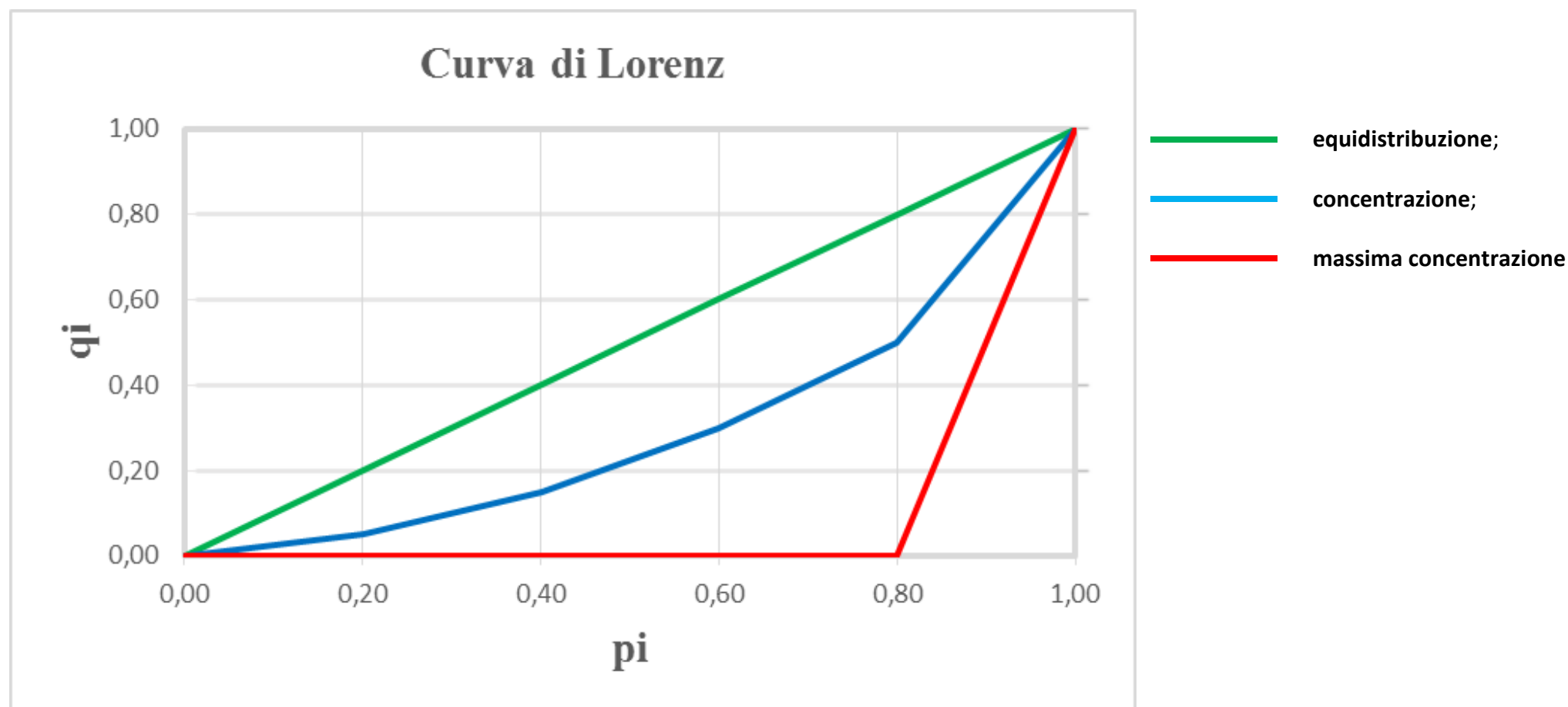
Si ricorda che in tutte le sommatorie **i** va da 1 a n-1.

Dalla definizione di equidistribuzione e di concentrazione di un carattere quantitativo trasferibile emerge in modo evidente l'esistenza del legame fra questi concetti e la variabilità. Nel senso che la **situazione di equidistribuzione coincide chiaramente con una situazione di variabilità nulla**. Mentre all'**aumentare della concentrazione aumenta la variabilità e viceversa**.

Fissato l'ammontare complessivo  $A$  di un carattere trasferibile, è possibile determinare la distribuzione di frequenze che ha la massima variabilità rispetto ad una data misura (come ad esempio la varianza). Si può dimostrare facilmente che tale distribuzione è proprio quella di massima concentrazione, cioè quella in cui  $n-1$  unità posseggono 0 e una sola possiede l'ammontare complessivo  $A$ .

Graficamente la concentrazione si rappresenta come si è detto attraverso il **diagramma di Lorenz** che consente di mettere rapidamente a confronto una situazione di concentrazione realmente osservata con la situazione ideale di equidistribuzione.

Tale diagramma, che trova forma utilizzando un sistema di assi cartesiani in cui sulle ascisse sono riportati i valori progressivi  $p_i$  delle frequenze e sulle ordinate i valori progressivi  $q_i$  delle corrispondenti intensità del fenomeno, entrambi espressi nella forma di frequenze cumulate relative, assume la forma di una retta, detta retta di equidistribuzione, di fatto bisettrice dell'angolo del primo quadrante, quando l'intensità del fenomeno è ripartita in parti uguali tra tutti i componenti della collettività, o di una curva crescente, avente la concavità rivolta verso l'alto, detta curva di concentrazione, che nella ipotesi di massima concentrazione coincide con l'asse delle ascisse fino a  $n-1$  casi, ovvero quelli di intensità pari a zero, per innalzarsi poi sino a  $q_i = 1,00$  in corrispondenza dell'unità che da sola dispone dell'ammontare complessivo del fenomeno.



## Alcuni esempi

| $x_i$<br>ordinati | $n_i$ | $x_i'$<br>cum | $n_i'$<br>cum | $p_i =$<br>$n_i \text{ cum} / n$ | $q_i =$<br>$x_i \text{ cum} / \Sigma x$ | $(p_i - q_i)$ |
|-------------------|-------|---------------|---------------|----------------------------------|-----------------------------------------|---------------|
| 2.000.000         | 1     | 2.000.000     | 1             | 0,200                            | 0,200                                   | 0,000         |
| 2.000.000         | 1     | 4.000.000     | 2             | 0,400                            | 0,400                                   | 0,000         |
| 2.000.000         | 1     | 6.000.000     | 3             | 0,600                            | 0,600                                   | 0,000         |
| 2.000.000         | 1     | 8.000.000     | 4             | 0,800                            | 0,800                                   | 0,000         |
| 2.000.000         | 1     | 10.000.000    | 5             | 1,000                            | 1,000                                   | 0             |
| 10.000.000        | 5     |               |               | 2,000                            |                                         | 0,000         |

Il valore di  $R = \sum_{i=1}^{n-1} (p_i - q_i)$  è = 0 perché c'è **equidistribuzione**;

$$R_N = \frac{R}{R_{\max}} = \frac{\sum (p_i - q_i)}{\sum p_i} = \frac{0,000}{2,000} = 0,000$$

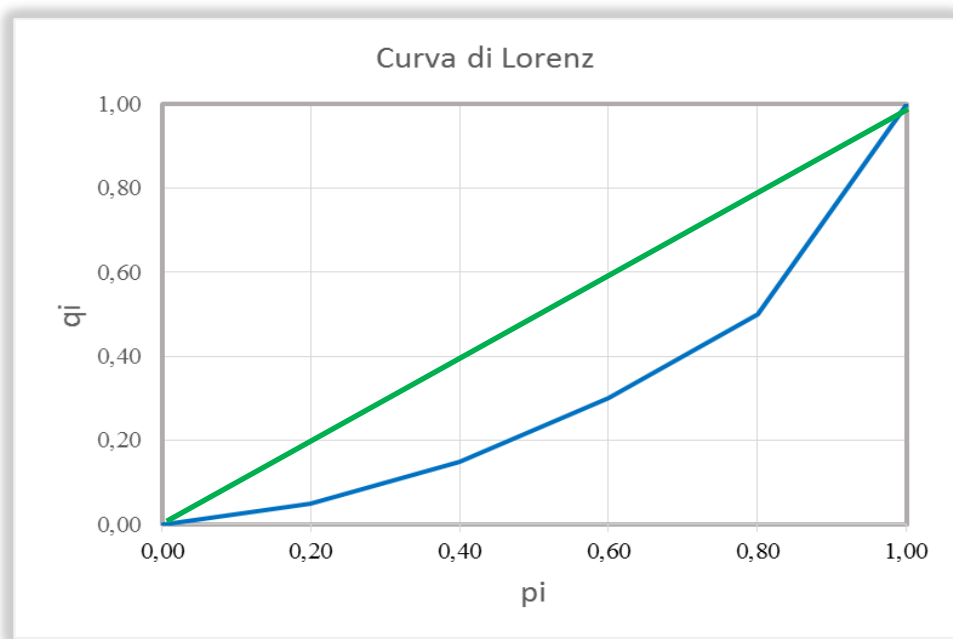
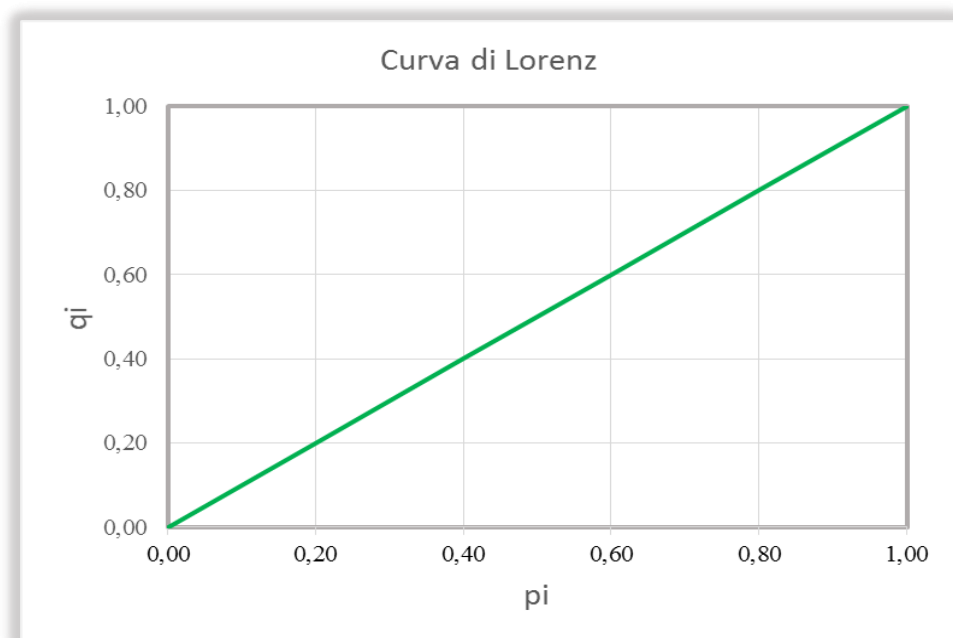
$$\text{o anche} = \frac{2}{n-1} \sum (p_i - q_i) = \frac{2}{4} 0,000 = 0,000$$

| $x_i$<br>ordinati | $n_i$ | $x_i'$<br>cum | $n_i'$<br>cum | $p_i =$<br>$n_i \text{ cum} / n$ | $q_i =$<br>$x_i \text{ cum} / \Sigma x$ | $(p_i - q_i)$ |
|-------------------|-------|---------------|---------------|----------------------------------|-----------------------------------------|---------------|
| 500.000           | 1     | 500.000       | 1             | 0,200                            | 0,050                                   | 0,150         |
| 1.000.000         | 1     | 1.500.000     | 2             | 0,400                            | 0,150                                   | 0,250         |
| 1.500.000         | 1     | 3.000.000     | 3             | 0,600                            | 0,300                                   | 0,300         |
| 2.000.000         | 1     | 5.000.000     | 4             | 0,800                            | 0,500                                   | 0,300         |
| 5.000.000         | 1     | 10.000.000    | 5             | 1,000                            | 1,000                                   | 0             |
| 10.000.000        | 5     |               |               | 2,000                            |                                         | 1,000         |

Il valore di  $R = \sum_{i=1}^{n-1} (p_i - q_i)$  è > 0 perché c'è **concentrazione**;

$$R_N = \frac{R}{R_{\max}} = \frac{\sum (p_i - q_i)}{\sum p_i} = \frac{1,000}{2,000} = 0,500$$

$$\text{o anche} = \frac{2}{n-1} \sum (p_i - q_i) = \frac{2}{4} 1,000 = 0,500$$

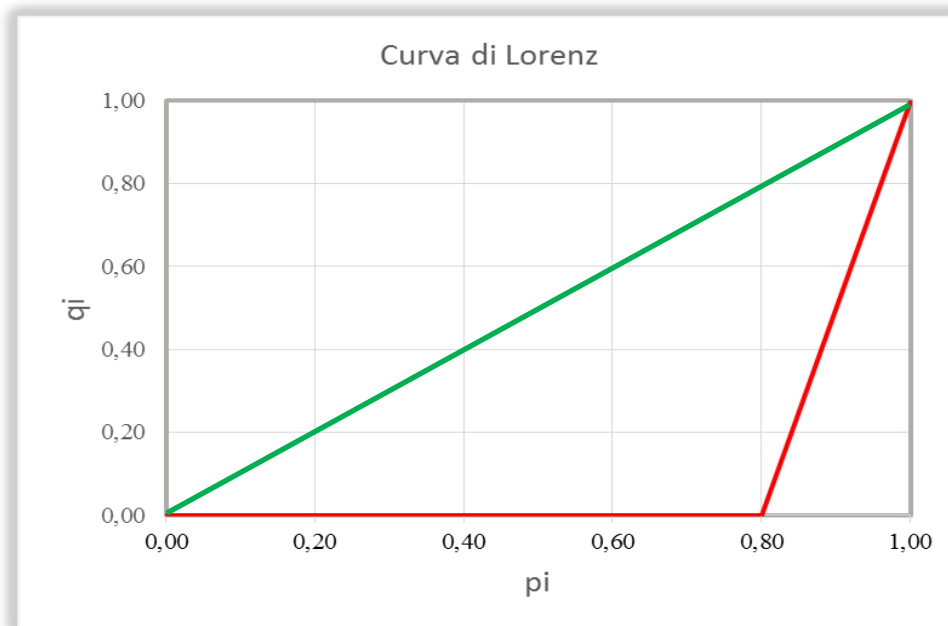


| $x_i$<br>ordinati | $n_i$ | $x_i'$<br>cum | $n_i'$<br>cum | $p_i =$<br>$n_i \text{ cum} / n$ | $q_i =$<br>$x_i \text{ cum} / \Sigma x$ | $(p_i - q_i)$ |
|-------------------|-------|---------------|---------------|----------------------------------|-----------------------------------------|---------------|
| 0                 | 1     | 0             | 1             | 0,200                            | 0,000                                   | 0,200         |
| 0                 | 1     | 0             | 2             | 0,400                            | 0,000                                   | 0,400         |
| 0                 | 1     | 0             | 3             | 0,600                            | 0,000                                   | 0,600         |
| 0                 | 1     | 0             | 4             | 0,800                            | 0,000                                   | 0,800         |
| 10.000.000        | 1     | 10.000.000    | 5             | 1,000                            | 1,000                                   | 0             |
| 10.000.000        | 5     |               |               | 2,000                            |                                         | 2,000         |

Il valore di  $R = \sum_{i=1}^{n-1} (p_i - q_i)$  è  $= \Sigma p_i = \frac{n-1}{2}$  perché c'è massima concentrazione =  $R_{\max}$

$$R_N = \frac{R}{R_{\max}} = \frac{\Sigma(p_i - q_i)}{\Sigma p_i} = \frac{2,000}{2,000} = 1,000$$

$$\text{o anche} = \frac{2}{n-1} \Sigma(p_i - q_i) = \frac{2}{4} \cdot 2,000 = 1,000$$



Altri esercizi

| $x_i$<br>ordinati | $n_i$ | $x_i$<br>cum | $n_i$ cum | $p_i =$<br>cum / n | $q_i =$<br>xi cum / $\Sigma x$ | $(p_i - q_i)$ |
|-------------------|-------|--------------|-----------|--------------------|--------------------------------|---------------|
| 3.500             | 1     | 3.500        | 1         | 0,167              | 0,079                          | 0,088         |
| 4.200             | 1     | 7.700        | 2         | 0,333              | 0,173                          | 0,160         |
| 5.800             | 1     | 13.500       | 3         | 0,500              | 0,303                          | 0,197         |
| 8.600             | 1     | 22.100       | 4         | 0,667              | 0,496                          | 0,171         |
| 10.250            | 1     | 32.350       | 5         | 0,833              | 0,726                          | 0,107         |
| 12.200            | 1     | 44.550       | 6         | 1,000              | 1,000                          | 0,000         |
| 44.550            | 6     |              |           | 2,500              |                                | 0,723         |

$$R = \sum_{i=1}^{n-1} (p_i - q_i) = 0,723 \quad R_N = \frac{R}{R_{\max}} = \frac{0,723}{2,500} = 0,289$$

$$\text{o anche} \quad R_N = \frac{R}{R_{\max}} = \frac{2}{5} * 0,723 = 0,289$$

| $x_i$<br>ordinat | $n_i$ | $x_i$<br>cum | $n_i$ cum | $p_i =$<br>cum / n | $q_i =$<br>xi cum / $\Sigma x$ | $(p_i - q_i)$ |
|------------------|-------|--------------|-----------|--------------------|--------------------------------|---------------|
| 600              | 1     | 600          | 1         | 0,200              | 0,080                          | 0,120         |
| 900              | 1     | 1.500        | 2         | 0,400              | 0,200                          | 0,200         |
| 1.200            | 1     | 2.700        | 3         | 0,600              | 0,360                          | 0,240         |
| 1.900            | 1     | 4.600        | 4         | 0,800              | 0,613                          | 0,187         |
| 2.900            | 1     | 7.500        | 5         | 1,000              | 1,000                          | 0,000         |
| 7.500            | 5     |              |           | 2,000              |                                | 0,747         |

$$R = \sum_{i=1}^{n-1} (p_i - q_i) = 0,747 \quad R_N = \frac{R}{R_{\max}} = \frac{0,747}{2,000} = 0,373$$

o anche  $R_N = \frac{R}{R_{\max}} = \frac{2}{4} * 0,747 = 0,373$

| Redd €<br>$x_i$<br>ordinati | Individui<br>$n_i$ | $x_i n_i$ | $x_i n_i$ cum | $n_i$ cum | $p_i =$<br>$n_i$ cum / n | $q_i =$<br>$x_i n_i$ cum / $\Sigma x_i n_i$ | $(p_i - q_i)$ |
|-----------------------------|--------------------|-----------|---------------|-----------|--------------------------|---------------------------------------------|---------------|
| 200                         | 30                 | 6.000     | 6.000         | 30        | 0,300                    | 0,162                                       | 0,138         |
| 250                         | 25                 | 6.250     | 12.250        | 55        | 0,550                    | 0,331                                       | 0,219         |
| 400                         | 20                 | 8.000     | 20.250        | 75        | 0,750                    | 0,547                                       | 0,203         |
| 550                         | 10                 | 5.500     | 25.750        | 85        | 0,850                    | 0,696                                       | 0,154         |
| 750                         | 15                 | 11.250    | 37.000        | 100       | 1,000                    | 1,000                                       | 0,000         |
|                             | 100                | 37.000    |               |           | 2,450                    |                                             | 0,714         |

$$R = \sum_{i=1}^{n-1} (p_i - q_i) = 0,714 \quad R_N = \frac{R}{R_{\max}} = \frac{0,714}{2,450} = 0,291$$



| Amp dem<br>in .000 | Residenti | val centr<br>$x_i$ | $x_i n_i$   | $x_i n_i$ cum | $n_i$ cum | $p_i =$<br>$n_i$ cum /n | $q_i =$<br>$x_i n_i$ cum/ $\Sigma x_i n_i$ | $(p_i - q_i)$ |
|--------------------|-----------|--------------------|-------------|---------------|-----------|-------------------------|--------------------------------------------|---------------|
| 0 - 5              | 8.411     | 2,5                | 21.027,5    | 21.027,5      | 8.411,0   | 0,059                   | 0,002                                      | 0,057         |
| 5 - 20             | 33.719    | 12,5               | 421.487,5   | 442.515,0     | 42.130,0  | 0,297                   | 0,048                                      | 0,249         |
| 20 - 50            | 43.199    | 35,0               | 1.511.965,0 | 1.954.480,0   | 85.329,0  | 0,602                   | 0,212                                      | 0,390         |
| 50 - 100           | 20.800    | 75,0               | 1.560.000,0 | 3.514.480,0   | 106.129,0 | 0,748                   | 0,381                                      | 0,367         |
| 100 - 220          | 35.727    | 160,0              | 5.716.320,0 | 9.230.800,0   | 141.856,0 | 1,000                   | 1,000                                      | 0,000         |
|                    | 141.856   |                    | 9.230.800,0 |               |           | 1,706                   |                                            | 1,063         |

$$R = \sum_{i=1}^{n-1} (p_i - q_i) = 1,063 \quad R_N = \frac{R}{R_{\max}} = \frac{1,063}{1,706} = 0,623$$

| Classi di<br>Investim | Imprese | val centr<br>$x_i$ | $x_i n_i$ | $x_i n_i$ cum | $n_i$ cum | $p_i =$<br>$n_i$ cum /n | $q_i =$<br>$x_i n_i$ cum/ $\Sigma x_i n_i$ | $(p_i - q_i)$ |
|-----------------------|---------|--------------------|-----------|---------------|-----------|-------------------------|--------------------------------------------|---------------|
| 1 - 15                | 14      | 8,0                | 112       | 112,0         | 14,0      | 0,14                    | 0,021                                      | 0,119         |
| 16 - 30               | 23      | 23,0               | 529       | 641,0         | 37,0      | 0,37                    | 0,119                                      | 0,251         |
| 31 - 45               | 10      | 38,0               | 380       | 1.021,0       | 47,0      | 0,47                    | 0,189                                      | 0,281         |
| 46 - 60               | 12      | 53,0               | 636       | 1.657,0       | 59,0      | 0,59                    | 0,307                                      | 0,283         |
| 61 - 75               | 6       | 68,0               | 408       | 2.065,0       | 65,0      | 0,65                    | 0,383                                      | 0,267         |
| 76 - 90               | 7       | 83,0               | 581       | 2.646,0       | 72,0      | 0,72                    | 0,491                                      | 0,229         |
| 91 - 105              | 28      | 98,0               | 2.744     | 5.390,0       | 100,0     | 1,00                    | 1,000                                      | 0,000         |
|                       | 100     |                    | 5.390     |               |           | 2,94                    |                                            | 1,429         |

$$R = \sum_{i=1}^{n-1} (p_i - q_i) = 1,429 \quad R_N = \frac{R}{R_{\max}} = \frac{1,429}{2,940} = 0,486$$

| $x_i$<br>ordinati | $n_i$ | $x_i$<br>cum | $n_i$ cum | $p_i =$<br>$n_i \text{ cum} / n$ | $q_i =$<br>$x_i \text{ cum} / \Sigma x$ | $(p_i - q_i)$ |
|-------------------|-------|--------------|-----------|----------------------------------|-----------------------------------------|---------------|
| 10                | 1     | 10           | 1         | 0,143                            | 0,071                                   | 0,071         |
| 12                | 1     | 22           | 2         | 0,286                            | 0,157                                   | 0,129         |
| 15                | 1     | 37           | 3         | 0,429                            | 0,264                                   | 0,164         |
| 18                | 1     | 55           | 4         | 0,571                            | 0,393                                   | 0,179         |
| 20                | 1     | 75           | 5         | 0,714                            | 0,536                                   | 0,179         |
| 30                | 1     | 105          | 6         | 0,857                            | 0,750                                   | 0,107         |
| 35                | 1     | 140          | 7         | 1,000                            | 1,000                                   | 0,000         |
| 140               | 7     |              |           | 3,000                            |                                         | 0,829         |

$$R = \sum_{i=1}^{n-1} (p_i - q_i) = 0,829$$

$$R_N = \frac{R}{R_{\max}} = \frac{0,829}{3,000} = 0,276$$

$$\text{o anche } \frac{2}{6} \cdot 0,829 = 0,276$$