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

6. LA CONCENTRAZIONE

Prof. Maurizio Pertichetti

Statistica sociale

6. LA CONCENTRAZIONE

La **concentrazione** è un aspetto della variabilità che concerne esclusivamente alcune tipologie di caratteri cosiddetti trasferibili.
Un carattere è detto trasferibile se ha senso immaginare che una unità statistica possa cedere tutto o parte del carattere posseduto ad un'altra unità statistica. E' trasferibile la variabile la cui intensità globale o una sua parte sia attribuibile (anche solo idealmente) ad una sola o a poche unità. Sono variabili trasferibili: Reddito, Diritti di possesso, Quote di mercato, Stato di salute; Variabili non trasferibili sono quelle che riguardano aspetti intrinseci delle unità: altezza, peso, età. 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;
- il diagramma di Lorenz.

Sia dato un carattere X, esso si dice che ha un alto grado di concentrazione se molta parte della sua complessiva intensità è distribuita tra un numero ridotto di unità statistiche. Si parla di:

- **concentrazione nulla** (o **equidistribuzione**) quando tutte le unità possiedono il carattere nella stessa misura;
- **concentrazione massima** quando una sola unità possiede l'intero ammontare del carattere.

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 poca ne posseggono, 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.
Percontro, 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.
Si supponga di disporre di n quantità non negative (dette intensità) poste in ordine non decrescente:

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(i)} \leq \ldots \leq x_{(n)} \quad \left(\text{per le quali l'intensità totale } \mathbf{q} = \sum_{i=1}^{n} x_i \right)$$

ottenute dall'osservazione di un carattere X (reddito, patrimonio, etc.) che goda della proprietà della trasferibilità, tale per cui si possano presentare le due situazioni limite:

- l'**equidistribuzione**, ovvero le n intensità sono uguali tra loro : $x1 = x2 =\ldots= xi=\ldots= xn = q/n$;
- la **concentrazione massima**, ovvero tutta l'intensità q è addensata su una singola osservazione, cosicché: : $x1 = x2 =\ldots= xi=\ldots= xn-1 = 0$ e $xn = q$.

Si definiscono **frazioni cumulate delle frequenze** delle prime i unità statistiche le quantità:

$$p_i = \frac{i}{n} \quad \text{ovvero} \quad p_1 = \frac{1}{n} \quad p_2 = \frac{2}{n} \quad \ldots \quad p_n = \frac{n}{n} = 1$$

Si definiscono inoltre, **frazioni cumulate dell'intensità relativa** posseduta dalle prime i unità statistiche le seguenti quantità:

$$q_i = \frac{x_1+x_2+..+x_i}{q} = \frac{\sum_{j=1}^{i} x_{(j)}}{q} \quad \text{ovvero} \quad q_1 = \frac{x_{(1)}}{q} \quad q_2 = \frac{\sum_{j=1}^{2} x_{(j)}}{q} \quad \ldots. \quad q_n = \frac{\sum_{j=1}^{n} x_{(j)}}{q} = 1$$

La somma $R = \sum_{i=1}^{n-1} (p_i - q_i)$ è una misura **assoluta** della concentrazione e il suo valore:

- è = 0 nel caso di **equidistribuzione**;
- è > 0 nel caso di **concentrazione**;
- è = a $\Sigma pi = \frac{n-1}{2}$ nel caso di **massima concentrazione** R_{max}

Rapportando **R** al suo massimo si ha l'indice normalizzato $\mathbf{R_N}$, ovvero:

$$R_N = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i} \quad \text{oppure} \quad R_N = \frac{2}{n-1} \sum_{i=1}^{n-1} (p_i - q_i)$$

detto **rapporto di concentrazione di Gini**.

Si tratta di un indice variabile tra 0 e 1 :
- è = 0 nel caso di equidistribuzione;
- è = 1 nel caso di massima concentrazione;

che fornisce una misura sintetica della concentrazione di un carattere trasferibile.

In presenza di distribuzioni con modalità raggruppate in classi, per la misura del grado di concentrazione si usano le seguenti formule:

$R = \frac{\Delta}{2\mu}$ ovvero il rapporto tra la differenza semplice media e il suo massimo, dove Δ è uguale a:

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

e l'altra che utilizza il metodo detto dei trapezzi la cui espressione è:

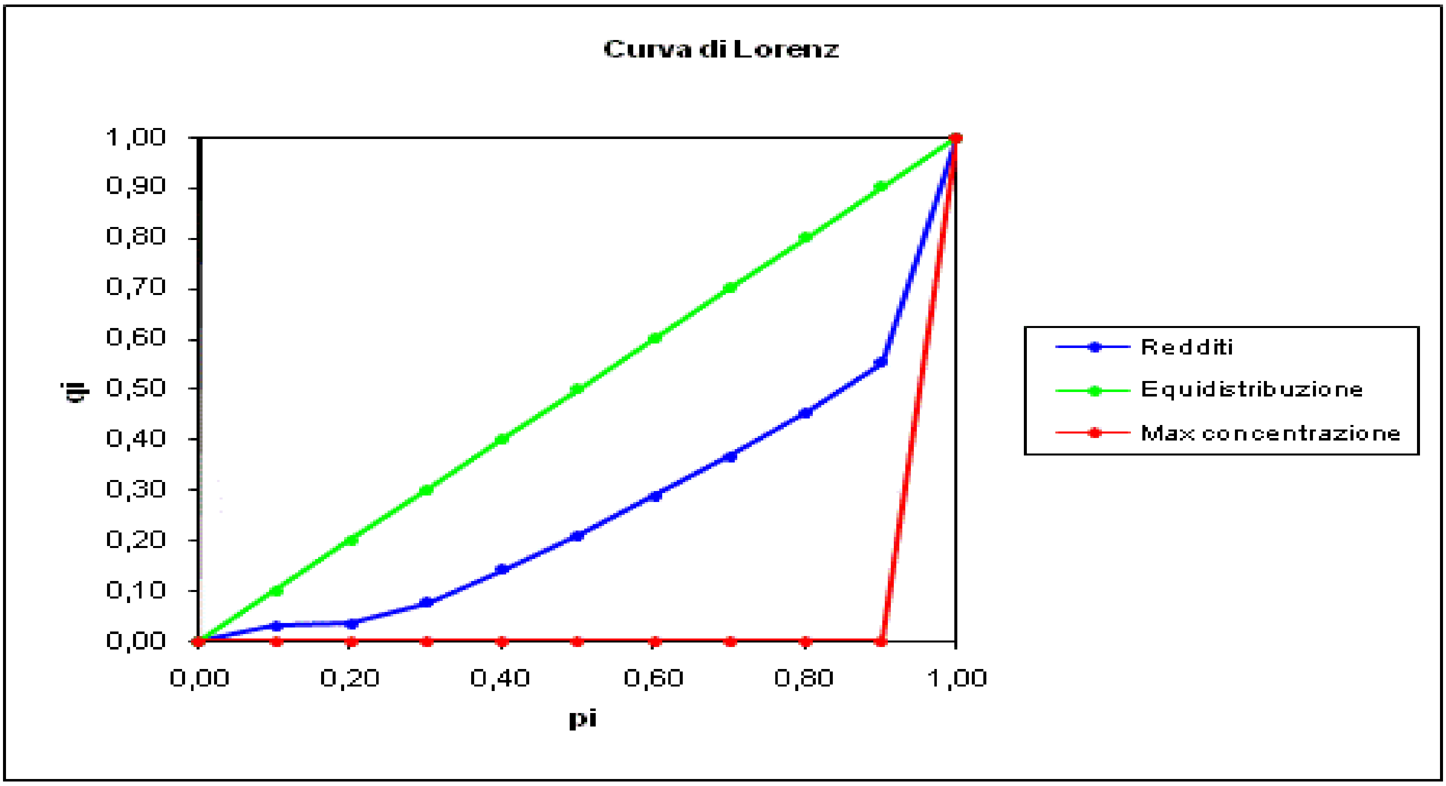
$$R = 1 - \sum_{i=0}^{k-1} (p_{i+1}-p_i)(q_i + q_{i+1})$$

Dalla definizione di equidistribuzione e di concentrazione di un carattere quantitativo (trasferibile) emerge in modo evidente l'esistenza di un legame fra questi concetti e la variabilità. **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 (varianza, differenza semplice media, etc.). Si può dimostrare facilmente che tale distribuzione è proprio quella di massima concentrazione, cioè quella in cui n-1 individui posseggono 0 e un solo individuo possiede l'ammontare complessivo A. Il concetto statistico di omogeneità è per certi versi analogo a quello di equidistribuzione o concentrazione nulla: difatti, un collettivo è detto omogeneo rispetto ad un dato carattere se tutte le sue unità presentano la medesima modalità del carattere.

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, nonché di calcolare alcune misure sintetiche della concentrazione.

Tale diagramma, che trova forma mediante un sistema di assi cartesiani in cui sulle ascisse sono riportate le somme progressive delle frequenze e sulle ordinate sono riportate le somme progressive delle corrispondenti intensità del fenomeno, viene ad essere rappresentato da una curva crescente, avente la concavità rivolta verso l'alto, detta curva di concentrazione.



Esempi di calcolo

Senza frequenze

Reddito mensile di 8 individui	x_i ordinati	n_i	x_i' cum	n_i' cum	$p_i = n_i$ cum $/n$	$q_i = x_i$ cum$/\Sigma x$	(p_i-q_i)
3.408,62	1.239,50	1	1.239,50	1	0,125	0,051	0,074
1.239,50	1.807,60	1	3.047,10	2	0,250	0,125	0,125
1.807,60	2.014,18	1	5.061,28	3	0,375	0,208	0,167
2.014,18	2.685,58	1	7.746,86	4	0,500	0,318	0,182
2.685,58	3.408,62	1	11.155,48	5	0,625	0,458	0,167
5.009,63	3.718,49	1	14.873,97	6	0,750	0,610	0,140
4.493,18	4.493,18	1	19.367,15	7	0,875	0,794	0,081
3.718,49	5.009,63	1	24.376,78	8	1,000	1,000	0
Σ	24.376,78	8			3,500		0,936

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

Il valore di $R = \sum_{i=1}^{n-1} (p_i - q_i)$:

è = 0 nel caso di **equidistribuzione**;

è > 0 nel caso di **concentrazione**;

è = $\Sigma pi = \frac{n-1}{2}$ nel caso di **massima concentrazione** R_{max}

$$R_N = \frac{R}{R_{max}} = \frac{\Sigma(p_i-q_i)}{\Sigma pi} = \frac{0,936}{3,50} = 0,268 \quad \text{o anche} \quad \frac{2}{n-1}\Sigma(p_i-q_i) = \frac{2}{7}\, 0,936 = 0,268$$

Con frequenze e classi di modalità

classi di reddito	aziende
1 - 15	14
16 - 30	23
31 - 45	10
46 - 60	12
61 - 75	6
76 - 90	7
91 e oltre	28
totale	100

x_i	n_i	$(x_{i+1}-x_i)$	n_i'	$(n-n_i')$	$n_i'(n-n_i')$	$n_i'(n-n_i')(x_{i+1}-x_i)$
8	14	15	14	86	1.204	18.060
23	23	15	37	63	2.331	34.965
38	10	15	47	53	2.491	37.365
53	12	15	59	41	2.419	36.285
68	6	15	65	35	2.275	34.125
83	7	15	72	28	2.016	30.240
98	28	.....	100	....	....	....
...	100	.....	...	....	....	191.040

x_in_i
112
529
380
636
408
581
2.744
5.390

n 100

μ 53,90

$$R = \frac{\Delta}{2\mu} = \frac{38,6}{107,8} = 0,358$$

$$\Delta = \frac{2\sum_{i=1}^{k-1} n_i'(n-n_i')(x_{i+1}-x_i)}{n(n-1)} = \frac{2 * 191.040}{100 * 99} = \frac{382.080}{9.900} = 38,6$$

$2\mu = 107,8$

Senza frequenze con il metodo dei trapezi

x_i	n_i
0	0
1.239,50	1
1.807,60	1
2.014,18	1
2.685,58	1
3.408,62	1
3.718,49	1
4.493,18	1
5.009,63	1
24.376,78	8

q_i = x_i cum	i	p_i = n_i cum /n	q_i = x_i cum/Σx
0,00	0	0,000	0,000
1.239,50	1	0,125	0,051
3.047,10	2	0,250	0,125
5.061,28	3	0,375	0,208
7.746,86	4	0,500	0,318
11.155,48	5	0,625	0,458
14.873,97	6	0,750	0,610
19.367,15	7	0,875	0,794
24.376,78	8	1,000	1,000
.....	...	.....	....

$p_{i+1} - p_i$	$q_i + q_{i+1}$	$(p_{i+1} - p_i)(qi + q_{i+1})$
0,125	0,051	0,006
0,125	0,176	0,022
0,125	0,333	0,042
0,125	0,525	0,066
0,125	0,775	0,097
0,125	1,068	0,133
0,125	1,405	0,176
0,125	1,794	0,224
.....	.....	.....
.....		0,766

$$R = 1 - \sum_{i=0}^{k-1} (p_{i+1} - p_i)(qi + q_{i+1}) = 1 - 0,766 = 0,234$$

Si determina con R=Δ/2μ

x_i	n_i
3.500	2
4.200	4
5.800	2
8.600	5
10.250	1
12.200	1
44.550	15

$(x_{i+1}-x_i)$	n_i'
700	2
1.600	6
2.800	8
1.650	13
1.950	14
.....	15
.....	...

$(n-n_i')$	$n_i'(n-n_i')$	$n_i'(n-n_i')(x_{i+1}-x_i)$
13	26	18.200
9	54	86.400
7	56	156.800
2	26	42.900
1	14	27.300
....	....	....
....	....	331.600

x_in_i
7.000
16.800
11.600
43.000
10.250
12.200
100.850

n 15

μ 6.723,33

$$R = \frac{\Delta}{2\mu} = \frac{3.158{,}10}{13.446{,}7} = 0{,}235$$

$$\Delta = \frac{2 \sum_{i=1}^{k-1} n_i^{'} (n - n_i^{'}) (x_{i+1} - x_i)}{n(n-1)} = \frac{2 \quad * \quad 331.600}{15 \quad * \quad 14} = \frac{663.200}{210} = 3.158{,}10$$

$$2\mu = 13.446{,}7$$

Senza frequenze con il metodo dei trapezi

x_i	n_i
0	0
1.500.000	1
1.650.000	1
1.730.000	1
1.845.000	1
6.725.000	4

$q_i =$ x_i cum	i
0,00	0
1.500.000	1
3.150.000	2
4.880.000	3
6.725.000	4
…..	…

$p_i =$ n_i cum /n	$q_i =$ x_i cum/Σx
0,000	0,000
0,250	0,223
0,500	0,468
0,750	0,726
1,000	1,000
…..	….

$p_{i+1} - p_i$	$q_i + q_{i+1}$	$(p_{i+1} - p_i)(qi + q_{i+1})$
0,250	0,223	0,0558
0,250	0,691	0,1729
0,250	1,194	0,2985
0,250	1,726	0,4314
…..	…..	…..
…..		0,9586

$$R = 1 - \sum^{k-1} (p_{i+1} - p_i)(qi + q_{i+1}) = 1 - 0{,}9586 = 0{,}041$$