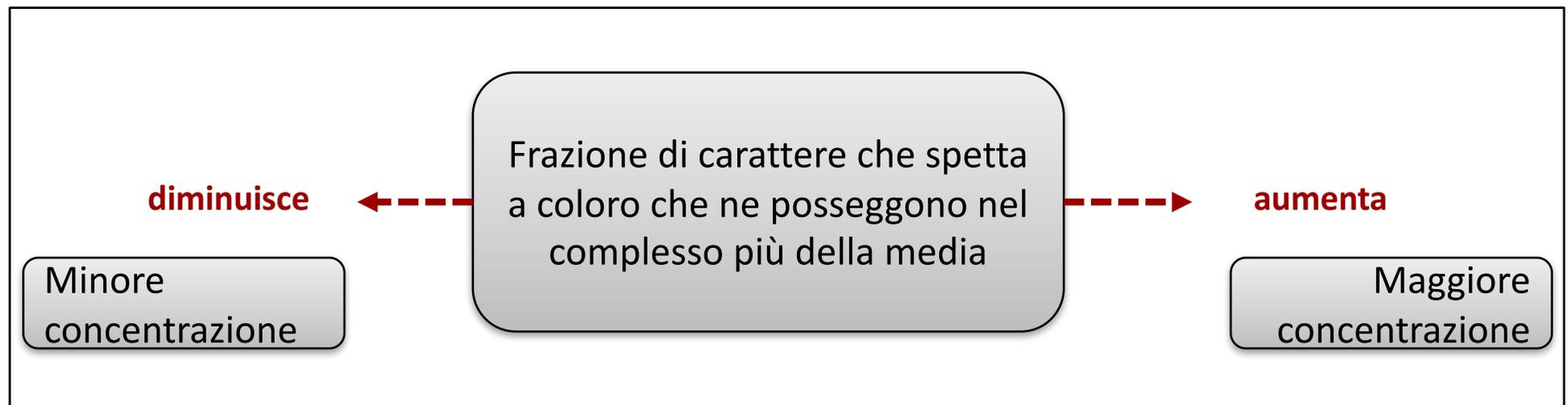


01 – La concentrazione

Unità n° 05

Nello studio dei fenomeni economici e sociali descritti attraverso caratteri quantitativi di tipo *trasferibile* può essere interessante analizzare la cosiddetta **concentrazione**. Riguarda il modo in cui un fenomeno trasferibile si ripartisce tra le unità, in particolare l'attitudine ad accentrarsi in un numero di unità ridotto



Si parla di **disuguaglianza distributiva** e si considera la concentrazione come un eccesso di tale particolare fenomeno

Esistono ovvie relazioni fra le nozioni di **concentrazione**, **disuguaglianza** e **variabilità**

02 – Trasferibilità

Unità n° 05

Un carattere (quantitativo) è *trasferibile* se:

- 1) ***assume soltanto valori non negativi***
- 2) ***ammette, sotto l'aspetto logico, che l'ammontare del carattere o parte di esso possa essere trasferito (anche solo idealmente) da un'unità all'altra del collettivo***

I caratteri non trasferibili riguardano aspetti intrinseci delle unità e non possono essere trasferiti senza trasferire, in solido, l'unità stessa

Se i valori del carattere sono livelli raggiungibili da qualsiasi unità ed ha un senso la loro somma o aggregazione allora lo studio di concentrazione è plausibile

Caratteri trasferibili	Caratteri non trasferibili
Reddito	Età
Quote di mercato	Altezza
Quote di produzione	Peso
Rischio commerciale (fatturato)	Componenti del nucleo familiare
...	...

03 – Definizioni preliminari (1)

Unità n° 05

Data una distribuzione unitaria di n osservazioni di un carattere X , ordinati in senso crescente

$$x_1 \leq x_2 \leq \dots \leq x_n$$

si è interessati a studiare come l'ammontare totale del carattere

$$A = \sum_{i=1}^n x_i$$

sia ripartito fra le diverse unità statistiche che compongono il collettivo:

(1) equidistribuzione (assenza di concentrazione)

ciascuna delle n unità possiede $1/n$ dell'ammontare complessivo A del carattere, ossia:

$$x_i = A/n = \bar{x}$$

(2) massima concentrazione

l'intero ammontare del carattere è posseduto da una unità:

$$x_1 = x_2 = \dots = x_{n-1} = 0 \qquad x_n = A = n\bar{x}$$

04 – Definizioni preliminari (2)**Unità n° 05**

Si consideri un carattere (trasferibile) X osservato su un collettivo di N unità statistiche e siano queste ordinate secondo l'ammontare posseduto

Indichiamo con p_i la frequenza relativa cumulata delle prime i unità

$$p_i = \frac{i}{n}$$

Allo stesso modo, dato l'ammontare A , indichiamo con q_i la frazione cumulata del carattere posseduto dalle prime i unità

$$q_i = \frac{A_i}{A} \quad \text{con} \quad A_i = A_1 + A_2 + A_3 + \dots + A_i$$

Si dimostra che per ogni unità del collettivo vale la relazione

$$p_i \geq q_i$$

05 – Misurare la concentrazione

Unità n° 05

Per misurare la concentrazione di un carattere X in un collettivo è necessario considerare gli scarti (differenze) tra frequenze relative cumulate delle unità statistiche e frazione cumulata dell'ammontare di carattere posseduto:

$$c_i = (p_i - q_i)$$

La concentrazione totale nel collettivo è ottenuta sommando gli scarti

$$\sum_{i=1}^{N-1} (p_i - q_i)$$

(1) quando si ha minima concentrazione (EQUIDISTRIBUZIONE) la somma è pari a 0

(2) quando si ha massima concentrazione allora la somma è pari a $\sum_{i=1}^{N-1} p_i$

06 – Costruire un indice normalizzato

Unità n° 05

In statistica spesso abbiamo l'esigenza di costruire indici che consentano un confronto tra le modalità di una distribuzione o tra distribuzioni diverse prescindendo dall'unità di misura del carattere studiato

Per ottenere un indice che varia tra 0 e 1 e può essere espresso in percentuale si usa dividere il valore dell'indice per il valore massimo che l'indice stesso può assumere:

ESEMPIO

n_i = numero di unità statistiche che presentano la i -esima modalità

f_i = numero di unità statistiche che presentano la i -esima modalità in rapporto alla dimensione del collettivo

In generale l'indice ottenuto come rapporto tra il valore osservato e il massimo valore che si poteva osservare rispetto al fenomeno studiato è detto **INDICE NORMALIZZATO**

N.B.: la misura è influenzata dalle ipotesi iniziali assunte per calcolare il massimo dell'indice

07 – Rapporto di concentrazione

Unità n° 05

Per misurare la concentrazione di un carattere in un collettivo è possibile utilizzare un indice normalizzato noto come **rapporto di concentrazione**

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

se $R=0$ \longrightarrow **equidistribuzione**

se $R=1$ \longrightarrow **max concentrazione**

Si esprime in percentuale: un valore k compreso tra 0 e 1 indica un livello di concentrazione pari al $k\%$ di quella massima osservabile nella distribuzione del carattere studiato

08 – Regole di lettura dell'indice

Unità n° 05

La lettura e l'interpretazione di un indice è semplice ma spesso risulta difficile esprimerlo in modo chiaro e univocamente comprensibile

In generale abbiamo visto come per $R=0$ si possa dire che *“non c'è concentrazione”* oppure che *“c'è equidistribuzione”* del carattere studiato nel collettivo; allo stesso modo per $R=1$ si può dire che *“c'è massima concentrazione”*

Quando l'indice R è compreso tra 0 e 1 allora la quantità ottenuta può essere letta in termini percentuali rispetto a quella massima osservabile in quel collettivo per quel carattere:

- *da 0 a 0.25 la concentrazione è bassa*

(es. $R=0.18 \rightarrow 18\%$ della max concentrazione osservabile, quindi si ha una bassa concentrazione)

- *da 0.25 a 0.5 la concentrazione è medio-bassa*

(es. $R=0.36 \rightarrow 36\%$ della max concentrazione osservabile, quindi si ha una concentrazione medio-bassa)

- *per $R=0.5$ si ha una media concentrazione*

- *da 0.5 a 0.75 la concentrazione è medio-alta*

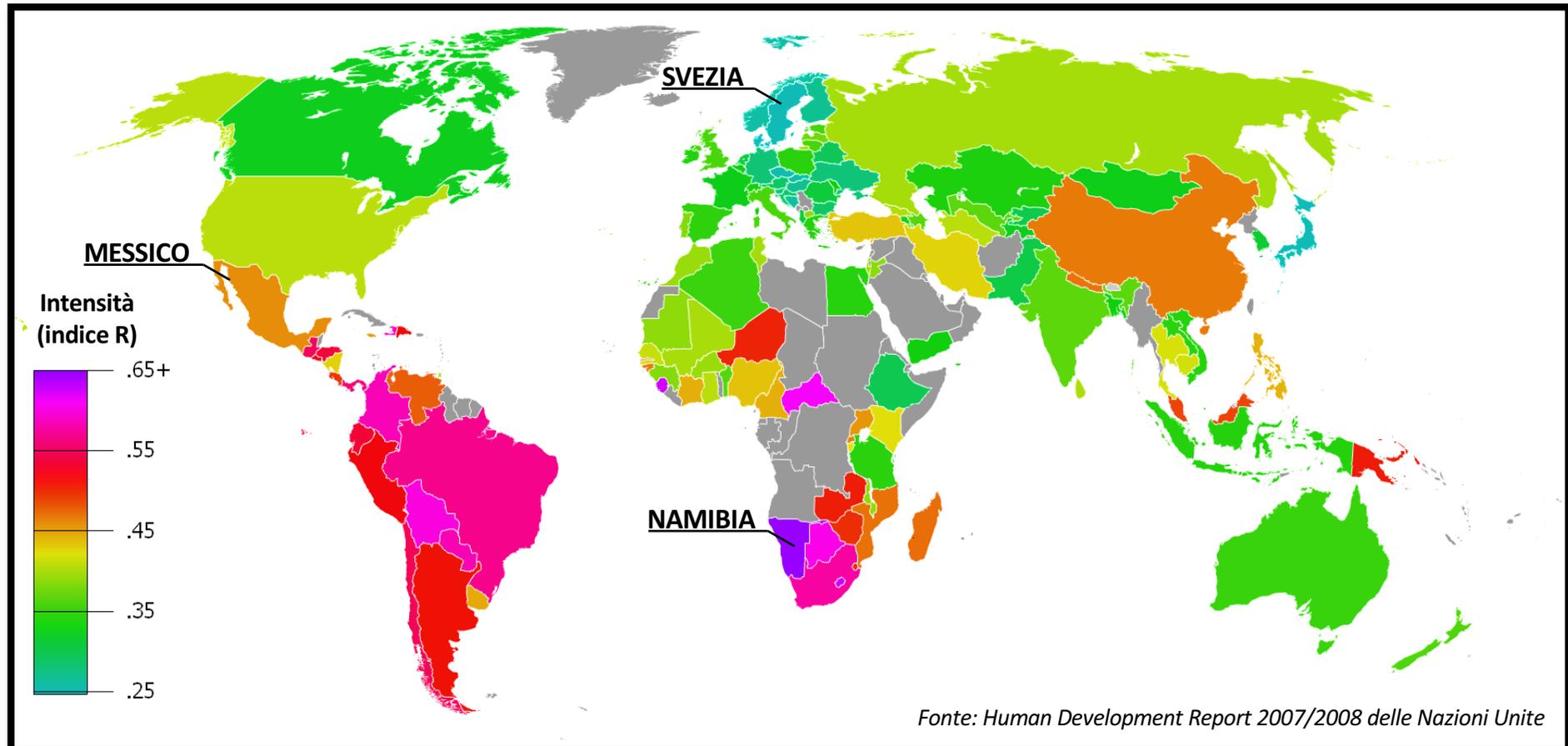
(es. $R=0.69 \rightarrow 69\%$ della max concentrazione osservabile, quindi si ha una concentrazione medio-alta)

- *da 0.75 a 1 la concentrazione è alta*

(es. $R=0.83 \rightarrow 83\%$ della max concentrazione osservabile, quindi si ha una alta concentrazione)

09 – Esempio: disuguaglianze di reddito nel mondo

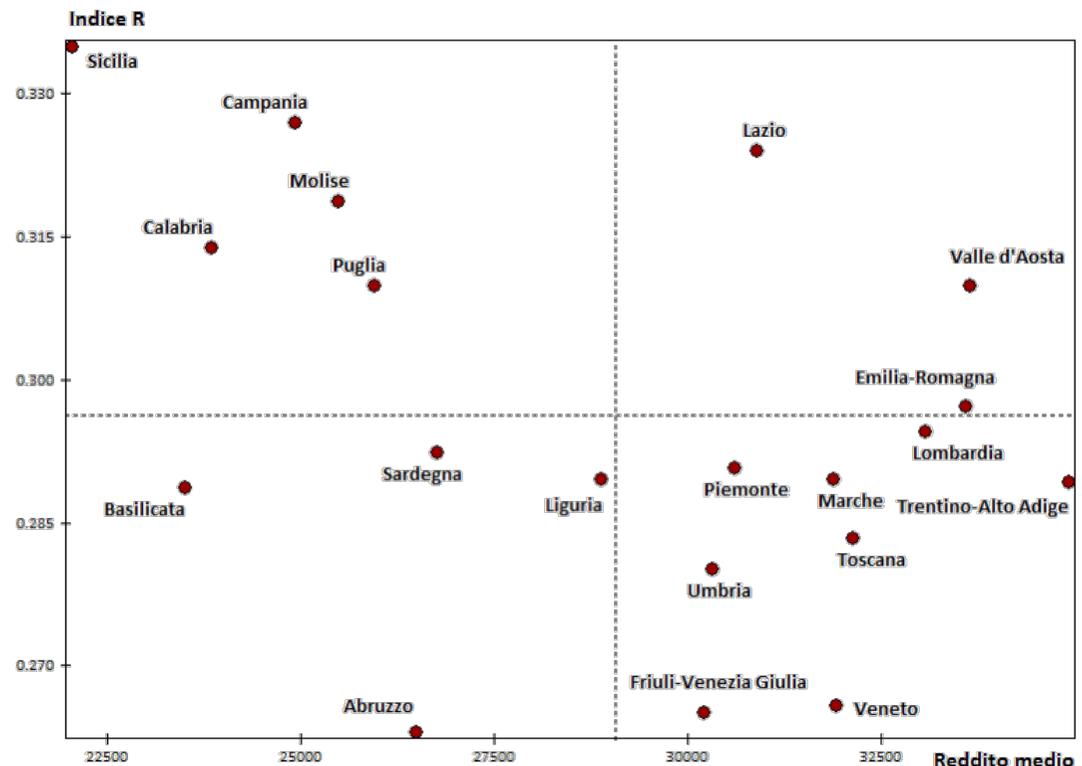
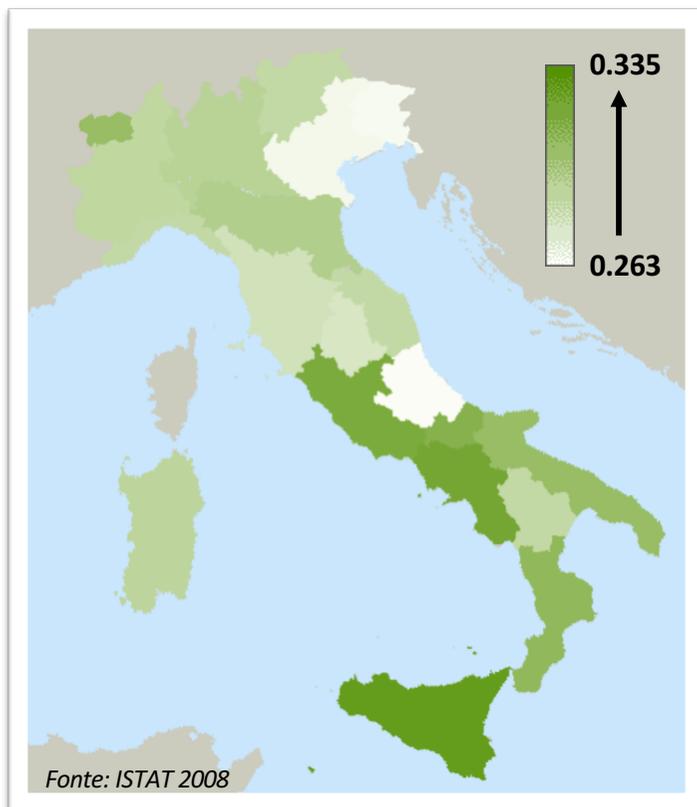
Unità n° 05



Dalla mappa è possibile vedere come la Namibia abbia un alto livello di concentrazione della ricchezza (R = 70.7%, maggiore disuguaglianza), mentre la Svezia ha un livello di concentrazione basso (R = 23%, minore disuguaglianza). Il Messico ha un livello di concentrazione medio (R = 48.2%)

10 – Esempio: concentrazione dei redditi familiari in Italia 2008

Unità n° 05



Piemonte	0.291	Marche	0.289
Valle d'Aosta	0.310	Lazio	0.324
Lombardia	0.295	Abruzzo	0.263
Liguria	0.290	Molise	0.319
Trentino-Alto Adige	0.289	Campania	0.327
Veneto	0.266	Puglia	0.310
Friuli-Venezia Giulia	0.265	Basilicata	0.289
Emilia-Romagna	0.297	Calabria	0.314
Toscana	0.283	Sicilia	0.335
Umbria	0.280	Sardegna	0.292

Dalla lettura dei dati si evince come in Sicilia ci sia più disuguaglianza rispetto, ad esempio al Trentino. Allo stesso modo possiamo fare un confronto tra Calabria e Basilicata tenendo conto anche del reddito medio: la disuguaglianza è maggiore nella prima regione rispetto alla seconda anche se mediamente il reddito è ad un livello abbastanza vicino... Ciò può essere interpretato col fatto che il reddito medio in Calabria subisce una maggior influenza dei redditi molto bassi o alti

11 – Uno schema per il calcolo

Unità n° 05

Ordinate
in senso
crescente

unità	X	p_i	A_i	q_i	$p_i - q_i$
1	x_1	$1/n$	A_1	A_1/A_n	$p_1 - q_1$
2	x_2	$2/n$	A_2	A_2/A_n	$p_2 - q_2$
3	x_3	$3/n$	A_3	A_3/A_n	$p_3 - q_3$
...
i	x_i	i/n	A_i	A_i/A_n	$p_i - q_i$
...
n	x_n	1	A_n	1	0

Per procedere da un punto di vista operativo allo studio della concentrazione è conveniente organizzare i dati secondo la tabella riportata di fianco

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

\longrightarrow Totale dell'ultima colonna fino alla penultima riga
 \longrightarrow Totale della terza colonna fino alla penultima riga

12 – Esempio

Unità n° 05

Vogliamo studiare la concentrazione dei ricavi derivanti dalla pubblicità in un collettivo di emittenti private operanti in una certa Regione

Emittenti televisive	Introiti pubblicitari	i	p_i	A_i	q_i
Tele Noi	3390	1	0.111	3390	0.029
TV9	4610	2	0.222	8000	0.067
Rete Beta	6970	3	0.333	14970	0.126
Tele Super	13200	4	0.444	28170	0.237
Canale 20	15240	5	0.556	43410	0.365
Noia TV	17980	6	0.667	61390	0.517
Tele Bella	18570	7	0.778	79960	0.673
Onda Sud	18890	8	0.889	98850	0.832
TV Maxi	19940	9	1	118790	1
Totale	118790				

$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = 0.288 \rightarrow \boxed{29\%}$$

La concentrazione degli introiti pubblicitari è pari al 29% di quella massima osservabile

C'è un basso livello di concentrazione

13 – Alcune considerazioni (1)

Unità n° 05

La tabella rappresentata in basso riporta i redditi di una società in cui vivono cinque individui. A causa della rapida crescita economica, i redditi di tutti gli individui raddoppiano. In questa società la disuguaglianza è aumentata, diminuita o invariata?

u.s.	x_i	A_i	p_i	q_i	$p_i - q_i$		u.s.	x_i	A_i	p_i	q_i	$p_i - q_i$	
1	200	200	0.200	0.068	0.132	R= 0.322	1	400	400	0.200	0.068	0.132	R= 0.322
2	450	650	0.400	0.220	0.180		2	900	1300	0.400	0.220	0.180	
3	550	1200	0.600	0.407	0.193		3	1100	2400	0.600	0.407	0.193	
4	750	1950	0.800	0.661	0.139		4	1500	3900	0.800	0.661	0.139	
5	1000	2950	1.000	1.000			5	2000	5900	1.000	1.000		

Quindi se applichiamo una trasformazione del tipo aX al carattere studiato manteniamo lo stesso livello di concentrazione, come dimostrato dall'esempio

Cosa accade se invece applichiamo una trasformazione del tipo $X+b$ al carattere?

Proviamo ad aggiungere a tutte le unità 500, 1000 e 2500€: cosa accade al livello di concentrazione del reddito?

14 – Alcune considerazioni (2)

Unità n° 05

u.s.	x_i	A_i	p_i	q_i	$p_i - q_i$
1	200	200	0.200	0.068	0.132
2	450	650	0.400	0.220	0.180
3	550	1200	0.600	0.407	0.193
4	750	1950	0.800	0.661	0.139
5	1000	2950	1.000	1.000	

R = 0.322
(32%)

+500



1	700	700	0.200	0.128	0.072
2	950	1650	0.400	0.303	0.097
3	1050	2700	0.600	0.495	0.105
4	1250	3950	0.800	0.725	0.075
5	1500	5450	1.000	1.000	

R = 0.174
(17%)

+1000



1	1200	1200	0.200	0.151	0.049
2	1450	2650	0.400	0.333	0.067
3	1550	4200	0.600	0.528	0.072
4	1750	5950	0.800	0.748	0.052
5	2000	7950	1.000	1.000	

R = 0.119
(12%)

+2500



1	2700	2700	0.200	0.175	0.025
2	2950	5650	0.400	0.366	0.034
3	3050	8700	0.600	0.563	0.037
4	3250	11950	0.800	0.773	0.027
5	3500	15450	1.000	1.000	

R = 0.061
(6%)

La concentrazione non è invariante per trasformazioni del tipo **X+b**

Se il reddito di ciascun individuo aumenta in modo proporzionale la concentrazione non cambia, ma se il reddito cresce della stessa quantità allora la concentrazione risulterà inferiore, perché è come se il reddito di ogni unità “si avvicinasse” a quello degli altri...

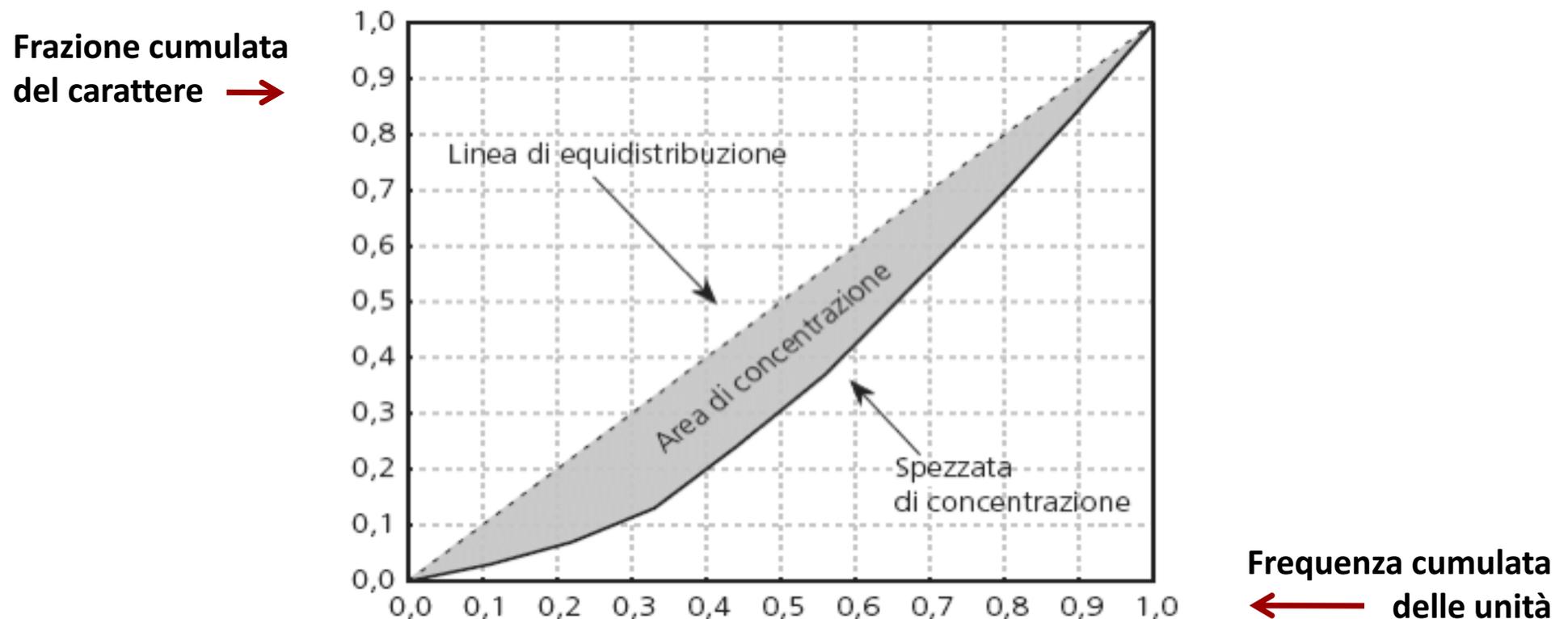
Infatti la frazione di reddito posseduta dalle diverse unità risulta essere:

u.s.		+500	+1000	+2500
1	0.068	0.128	0.151	0.175
2	0.153	0.174	0.182	0.191
3	0.186	0.193	0.195	0.197
4	0.254	0.229	0.220	0.210
5	0.339	0.275	0.252	0.227

15 – Rappresentazione grafica della concentrazione

Unità n° 05

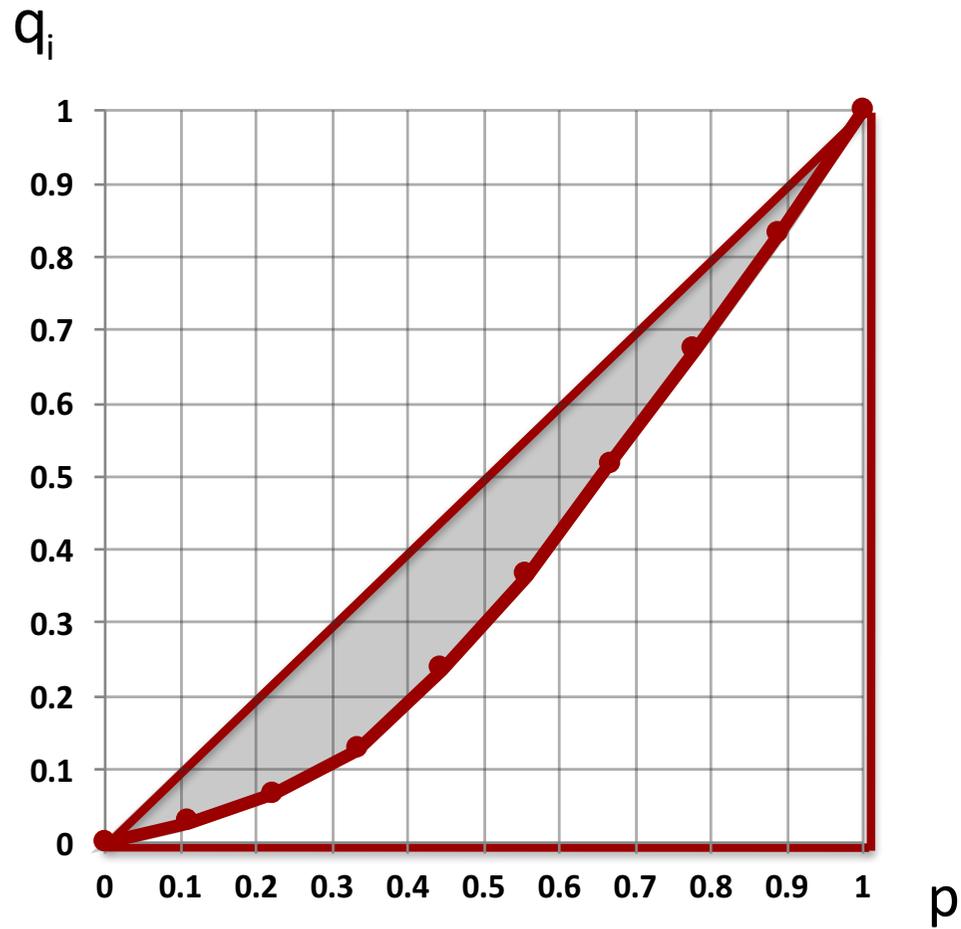
Con le coppie (p_i, q_i) è possibile realizzare una rappresentazione grafica della concentrazione detto **Curva di Lorenz**



Maggiore è l'area tra la bisettrice e la curva maggiore è il livello della concentrazione

16 – Esempio

Unità n° 05



Il valore di R esprime l'area compresa tra la spezzata di concentrazione e la linea di equidistribuzione: più piccolo è R (fino a 0) più la spezzata si avvicina alla linea, più grande è R (fino a 1) più la spezzata coincide con i cateti del triangolo (max concentrazione)

$$R = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = 0.288$$

Riprendendo l'esempio delle emittenti televisive si vede anche graficamente come ci sia una bassa concentrazione degli introiti pubblicitari

17 – Concentrazione per distribuzioni di frequenza

Unità n° 05

Consideriamo un carattere discreto con k modalità, e supponiamo di avere a disposizione il numero di unità statistiche sulle quali abbiamo osservato le diverse modalità

In questo caso per calcolare la concentrazione è più comodo utilizzare le seguenti espressioni:

Frequenza cumulata
delle unità



$$p_h = \frac{n_h}{n}$$

$$q_h = \frac{\sum_{i=1}^h x_i n_i}{\sum_{i=1}^k x_i n_i}$$



Frazione cumulata
del carattere

Per misurare la concentrazione utilizziamo ancora una volta l'indice di Gini

18 – Concentrazione per distribuzioni in classi

Unità n° 05

Nelle distribuzioni di frequenza e nelle distribuzioni in classi abbiamo una informazione in più: possiamo rilevare quante unità statistiche posseggono un certo ammontare di carattere

Nel caso di distribuzioni in classi per studiare la concentrazione dobbiamo considerare alcune ipotesi iniziali:

CASO A: se conosciamo l'ammontare di carattere posseduto e il numero di unità si assume che ci sia equidistribuzione (ogni unità della classe possiede lo stesso ammontare di carattere)

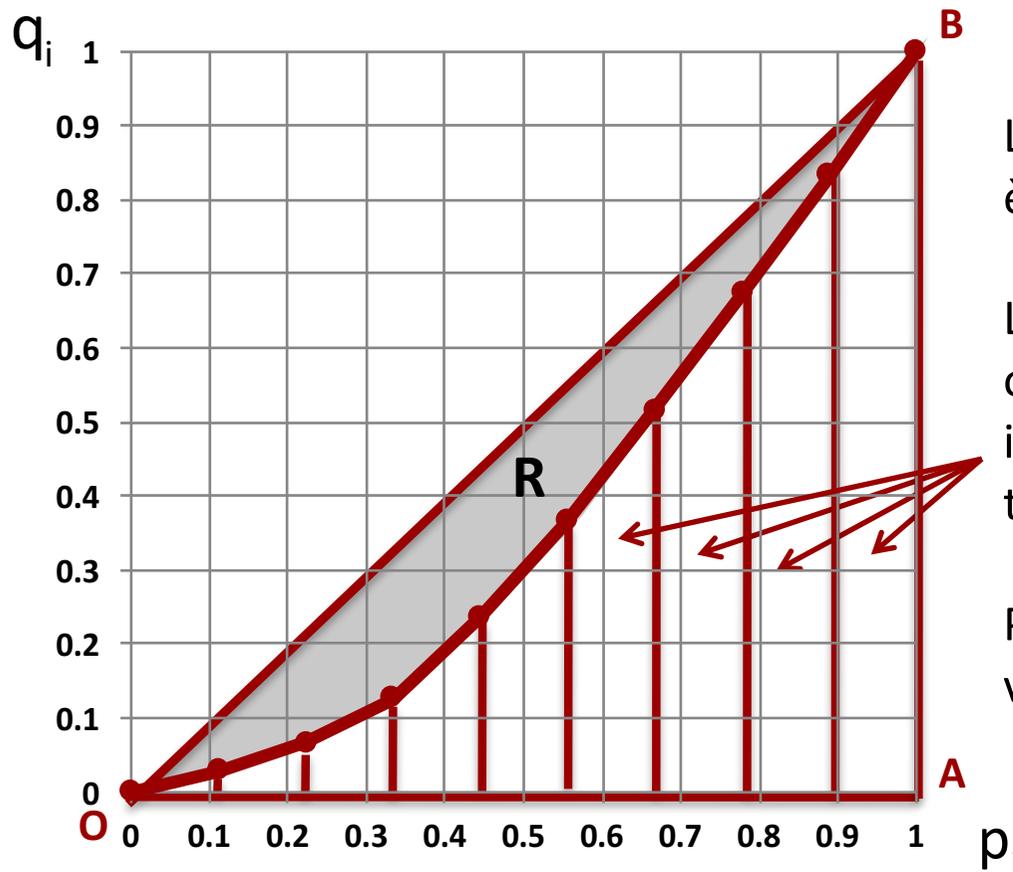
CASO B: se non conosciamo l'ammontare di carattere posseduto dalle unità della classe allora possiamo stimarlo moltiplicando il valore centrale per il numero di unità statistiche della classe

Spesso in questo caso per calcolare la concentrazione si preferisce utilizzare un altro metodo che deriva direttamente dall'osservazione della curva di Lorenz

Tale formulazione alternativa porta a risultati simili e quindi le regole di lettura rimangono quelle già viste in precedenza

19 – Metodo alternativo: i trapezi

Unità n° 05



L'area di max concentrazione (il triangolo OAB) è sempre pari a $1/2$

L'area di concentrazione (indicata con R) si ottiene sottraendo alla max concentrazione i trapezi e il triangolo (il primo da sinistra) tracciati sotto la spezzata

Per approssimazione otteniamo la formula già vista in precedenza

$$R = \frac{1}{2} - \left[\frac{p_1 q_1}{2} + \frac{q_1 + q_2 (p_2 - p_1)}{2} + \dots + \frac{q_{n-1} + q_n (p_n - p_{n-1})}{2} \right] \cong 1 - \sum_{i=0}^{n-1} (q_{i+1} + q_i) (p_{i+1} - p_i)$$

20 – Esempio **Unità n° 05**

classi	imprese	addetti
0-2	2043,0	2718,3
3-9	636,0	2845,6
10-19	103,2	1352,0
20-49	43,4	1281,2
50-99	11,8	808,7
100-499	8,3	1588,3
500-999	0,8	529,4
	2846,5	11123,5

Vogliamo studiare la concentrazione degli addetti nelle imprese di un certo settore: in questo caso si assume l'equidistribuzione dell'ammontare di carattere per ogni classe (addetti per impresa)

classi	imprese	f_i	p_i	addetti	q_i
0-2	2043,0	0,7177	0,7177	2718,3	0,2444
3-9	636,0	0,2234	0,9412	2845,6	0,2558
10-19	103,2	0,0363	0,9774	1352,0	0,1215
20-49	43,4	0,0152	0,9927	1281,2	0,1152
50-99	11,8	0,0041	0,9968	808,7	0,0727
100-499	8,3	0,0029	0,9997	1588,3	0,1428
500-999	0,8	0,0003	1,0000	529,4	0,0476
	2846,5	1		11123,5	1

Frequenza relativa delle imprese

Frazione relativa degli addetti

$$(q_{i+1} + q_i)(p_{i+1} - p_i)$$

0,1754
0,1664
0,0407
0,0207
0,0064
0,0051
0,0005
0,4152

$$R \cong 1 - 0,4152 = 0,5848$$

La concentrazione di addetti per impresa osservata, pari al 58% circa, è medio-alta

21 – Alcune considerazioni sull'interpretazione

Unità n° 05

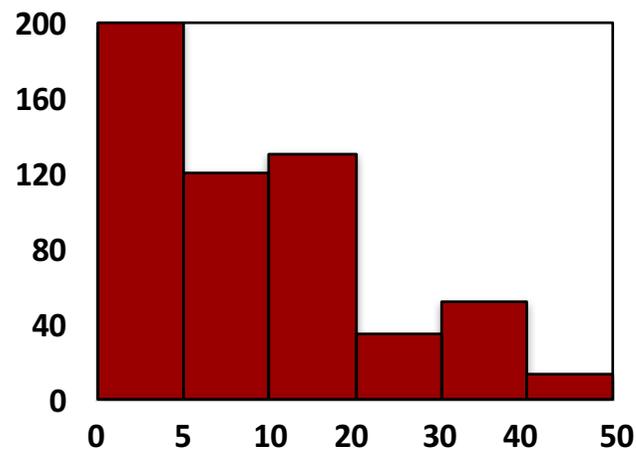
Consideriamo il Reddito Annuale delle 550 famiglie di un Comune della Provincia di Cosenza

Reddito Annuale	Famiglie
0 - 5000	200
5000 - 10000	120
10000 - 20000	130
20000 - 30000	34
30000 - 40000	52
40000 - 50000	14
Totale	550

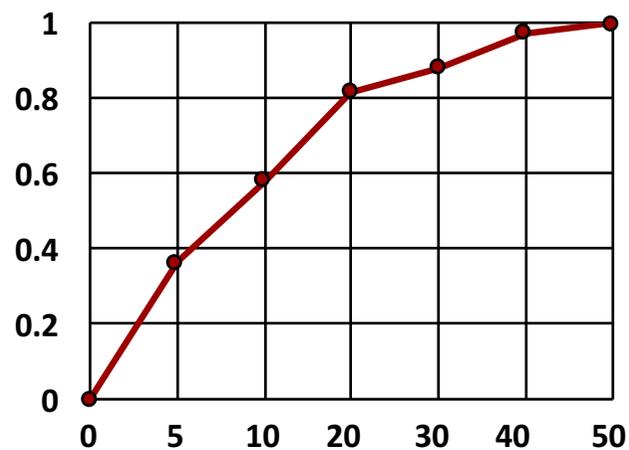
la **distribuzione** fornisce informazioni su come l'intensità del fenomeno reddito si è manifestata tra le diverse famiglie residenti nel Comune

la **ripartizione** dà indicazioni su quali sono le soglie di reddito corrispondenti alle diverse frazioni cumulate di famiglie residenti nel Comune

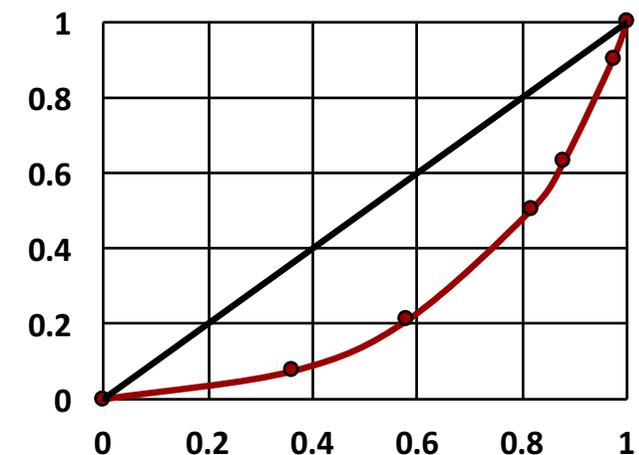
la **concentrazione** ci dice rispetto alle diverse frazioni cumulate di famiglie qual è il peso delle corrispondenti frazioni cumulate di reddito



distribuzione



ripartizione

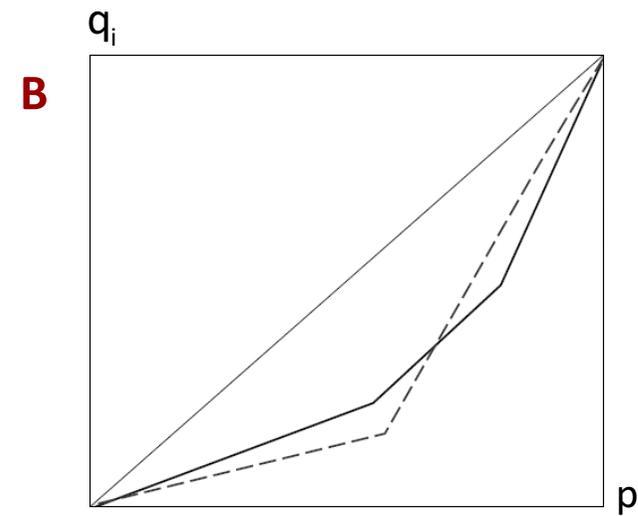
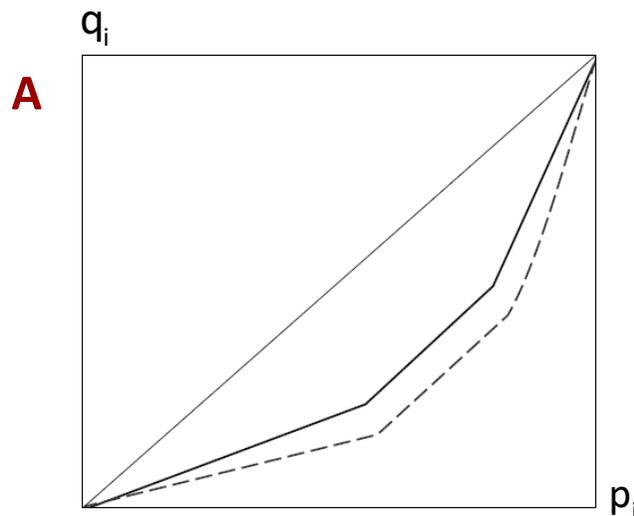


concentrazione

22 – Rapporto di Concentrazione e Curva di Lorenz

Unità n° 05

Consideriamo il livello di concentrazione di un fenomeno in due collettivi (il primo è indicato con la linea intera, il secondo con la linea tratteggiata)



Nel caso A possiamo dire che nel collettivo indicato con la linea intera il fenomeno è meno concentrato rispetto a quello indicato con la linea tratteggiata

Nel caso B non riusciamo invece dall'analisi della Curva di Lorenz a dare una risposta precisa: l'interpretazione del fenomeno diventa più complessa, soprattutto se i livelli dell'indici R sono identici

23 – Esercizi

Unità n° 05

- (1) Nella tabella di seguito sono riportati i redditi dichiarati (in €) per l'anno 2003 da alcuni parlamentari italiani

Parlamentare	Reddito
Berlusconi	12 762 234
Viale	2 158 908
Fiori	987 991
Paniz	974 054
Provera	37284

Studiare la concentrazione del reddito

- (2) Nella tabella sono riportati i dati relativi agli spettatori dei primi 10 film della stagione 2003

Titolo	Spettatori (in migliaia)
1) <i>A beautiful mind</i>	2473
2) <i>La vera storia di J. lo squartatore</i>	1584
3) <i>Il favoloso mondo di Amelie</i>	1447
4) <i>Parla con lei</i>	996
5) <i>Un viaggio chiamato amore</i>	747
6) <i>Mi chiamo Sam</i>	685
7) <i>Gosford Park</i>	642
8) <i>Casomai</i>	540
9) <i>Magdalene</i>	517
10) <i>I Tenenbaums</i>	480

Studiare la concentrazione degli spettatori

24 – Esempio

Unità n° 05

- (3) Di seguito è riportata la distribuzione in Italia nel 2000 del *reddito familiare annuo* (al netto delle imposte sul reddito e dei contributi previdenziali ed assistenziali)



<i>Classe di reddito (in migliaia di euro)</i>	<i>famiglie</i>	<i>reddito medio (in migliaia di euro)</i>
fino a 10	983	6564
da 10 a 20	2478	14985
da 20 a 30	1878	24613
da 30 a 40	1265	34718
oltre 40	1397	59931

Fonte: Banca d'Italia, Indagine sui bilanci delle famiglie italiane 2000

Si calcoli il livello della concentrazione del reddito

(Sol. -> $R = 0,342$ con il metodo dei trapezi)

25 – Esercizi

Unità n° 05

- (4) È stata analizzata la quantità di ferro (in mg) contenuta in 84 campioni di terreno A e 72 campioni di terreno B. I risultati sono riportati nella seguente tabella in cui sono riportate le distribuzioni di frequenza e l'ammontare di ferro per ogni classe:



quantità ferro (mg)	TERRENO A		TERRENO B	
	n° campioni	totale ferro (mg)	n° campioni	totale ferro (mg)
0 - 5	15	20	8	40
5 - 15	18	108	8	112
15 - 40	19	304	10	400
40 - 60	12	600	25	1025
60 - 70	6	390	15	975
70 - 100	14	1260	6	426
Totale	84	2682	72	2978

Verificare in quale terreno il ferro è più concentrato