

A woman with long dark hair, wearing a white t-shirt and jeans, stands in a meeting room, pointing her right hand towards a wall covered in numerous yellow and pink sticky notes. Several people are seated around a long white table in the foreground, looking towards the woman. Laptops and notebooks are open on the table. The room has large windows on the left. The image is overlaid with a semi-transparent dark grey filter. There are decorative white diamond shapes: one in the top right corner, one in the bottom right corner, and a red triangle with a white diamond inside on the left side.

Governare con i numeri: ricerca, elaborazione e presentazione dei dati

Bari – 9 aprile 2024

Vi parleremo di:

1. Che cos'è la statistica
2. Il linguaggio e gli strumenti della statistica
3. La rappresentazione dei fenomeni statistici
4. Le medie
5. La variabilità
6. Il confronto dei dati nel tempo e nello spazio





Istat**Learning**
AULA E ONLINE

CHE COS'È LA STATISTICA

La statistica è...

- La statistica è il miglior antidoto contro il pregiudizio
- Ci sono cose imprevedibili.
Per tutto il resto c'è la statistica
- Ciò che non può essere rappresentato statisticamente è un'opinione, non è scienza
- Statistica: dal caos all'ordine

Fonte: https://www.istat.it/it/files//2011/02/volume_Twitter.pdf



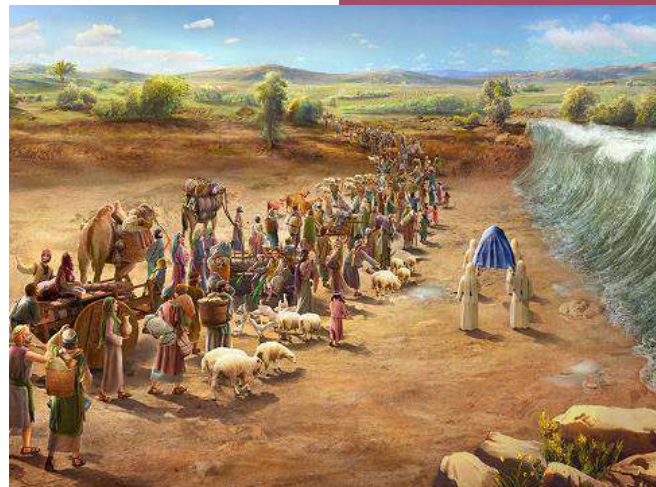
La statistica è...

- la scienza che ha per oggetto lo studio dei fenomeni collettivi suscettibili di misurazione e di descrizione quantitativa (Vocabolario Treccani)
- la scienza che ha come fine lo studio quantitativo e qualitativo di un collettivo, per descrivere un fenomeno (Wikipedia/Portale Statistica)
- **una disciplina di sintesi e di analisi che ha come obiettivo la rappresentazione quantitativa della realtà** (Antolini F. - Truglia F., 2009)

Quando nasce la statistica

«Fate il computo di tutta la comunità ..., secondo le loro famiglie, secondo i loro casati paterni, contando i nomi di tutti i maschi, testa per testa, dai vent'anni in su, quanti ... possono andare in guerra; ... li censirete, schiera per schiera. Sarà con voi un uomo per tribù, un uomo che sia capo del casato dei suoi padri»

(Libro dei Numeri, I, 2-4)



**Inventari della popolazione e dei suoi beni,
finalizzati a scopi amministrativi e militari**

Quando nasce la statistica

- L'utilizzo del termine statistica - inteso come «*descrizione delle qualità che caratterizzano e degli elementi che compongono uno Stato*» è attribuito a Girolamo Ghislini
- Alla base della moderna concezione di statistica possiamo considerare la "aritmetica politica" di John Graunt e William Petty, definita come «*l'arte di ragionare mediante le cifre sulle cose che riguardano il governo*»

Quando nasce la statistica



«La politica insegna
come gli stati devono
essere; la statistica
descrive come essi sono
realmente»

(Goffredo Achenwall)



Istat**Learning**
AULA E ONLINE

IL LINGUAGGIO E GLI STRUMENTI DELLA STATISTICA

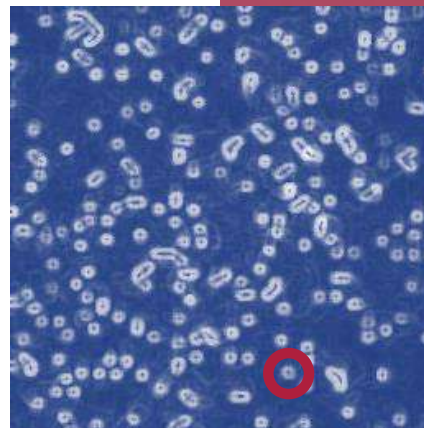
Sentirete parlare di

- **Caratteri**: sono le caratteristiche che è possibile osservare (età, altezza, reddito, stato civile, titolo di studio, condizione professionale, ecc.)
- Un carattere può assumere **modalità**
 - numerica → carattere quantitativo o **variabile**
 - non-numerica → carattere qualitativo o **mutabile**
- Le modalità possono essere tanto numerose da rendere necessario il loro accorpamento in **classi**

Il processo di classificazione organizza
le possibili modalità in gruppi distinti,
che costituiscono a loro volta macro-modalità

Sentirete parlare di

- **Unità statistica**: è l'unità elementare oggetto di osservazione
 - le più comuni unità statistiche sono individui, famiglie, imprese, aziende agricole
- A volte occorre distinguere fra unità statistica
 - **di rilevazione** → quella presso la quale si rilevano le informazioni
 - **di analisi** → quella di cui si studiano le caratteristiche
- **Collettivo statistico**: è l'insieme delle unità statistiche omogenee relativamente a uno o più caratteri



Sentirete parlare di

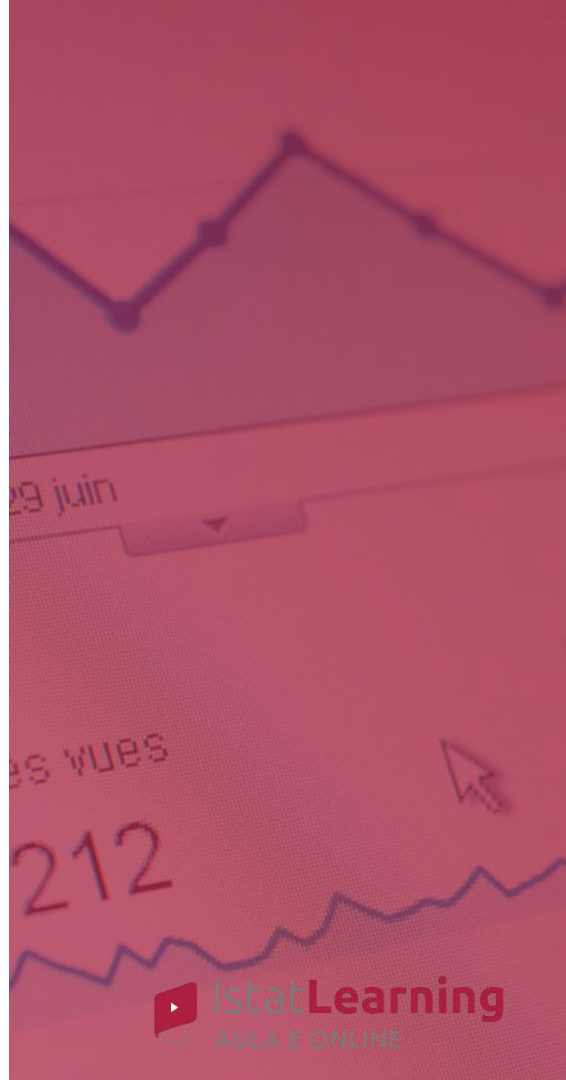
- **Serie:** successione di dati, descrive il modo in cui uno o più caratteri si manifestano (distribuiscono) nel collettivo

Unità	Caratteri				
	Età	Genere	Titolo di studio	Condizione professionale	Stato civile
0000	28	F	Laurea	Occupata	Nubile
0001	32	M	Diploma	Disoccupato	Celibe
.....
9999	31	F	Laurea	Occupata	Coniugata

Sentirete parlare di

- **Distribuzione statistica di frequenza:** descrive quante volte si manifesta ciascuna delle modalità possibili del carattere osservato

Titolo di studio	Frequenza Assoluta	Frequenza relativa	Frequenza percentuale
Nessuno Licenza elementare	800	0,08	8%
Licenza media	3.000	0,30	30%
Diploma	3.700	0,37	37%
Laurea o superiore	2.500	0,25	25%
TOTALE	10.000	1,00	100%



Sentirete parlare di

- **Distribuzione statistica per classi:** descrive quante volte si manifesta ciascuna delle classi in cui sono state accorpate le possibili modalità del carattere

Imprese per classe di addetti		
Categoria	Numero di addetti	Frequenza percentuale
Grandi imprese	250 e oltre	0,08%
Medie imprese	50-249	0,47%
Piccole imprese	10-49	4,23%
Microimprese	0-9	95,23%

Grandi imprese

Medie imprese

Piccole imprese

Microimprese

La Tabella statistica

- La tabella costituisce la prima sintesi dei dati ed è il punto di partenza per la rappresentazione, la lettura, l'interpretazione dei dati e le elaborazioni successive
- Essa deve essere autoesplicativa e di chiara e facile comprensione
- La tabella può contenere una distribuzione semplice, doppia, multipla

Tavola 1.4 Superficie territoriale e popolazione residente per zona altimetrica dei Comuni e regione (a)
Anno 2021, superficie territoriale in km²

ANNI REGIONI	Valori assoluti						Valori percentuali					
	Montagna		Collina		Pianura		Montagna		Collina		Pianura	
	Superficie (a)	Popolazione (b)	Superficie (a)	Popolazione (b)	Superficie (a)	Popolazione (b)	Superficie (a)	Popolazione (b)	Superficie (a)	Popolazione (b)	Superficie (a)	Popolazione (b)
2020	106.308	7.191.729	125.771	22.914.894	69.989	29.129.780	35,2	12,2	41,6	38,8	23,2	49,4
2021 - PER REGIONE												
Piemonte	10.977	470.088	7.669	1.303.121	6.721	2.479.070	43,2	11,1	30,3	30,6	28,5	58,3
Valle d'Aosta/Valle d'Aosta	3.261	123.337	-	-	-	-	100,0	100,0	-	-	-	-
Liguria	3.523	751.262	1.803	756.176	-	-	66,1	40,8	34,9	60,2	-	-
Lombardia	9.835	1.012.625	2.663	2.354.804	11.245	6.697.617	40,4	10,2	12,5	20,8	47,1	69,2
Trentino-Alto Adige/South Tyrol	13.805	1.077.832	-	-	-	-	100,0	100,0	-	-	-	-
Bolzano/South Tyrol	7.398	535.774	-	-	-	-	100,0	100,0	-	-	-	-
Trento	6.207	542.058	-	-	-	-	100,0	100,0	-	-	-	-
Veneto	5.284	323.008	2.649	799.588	10.413	3.726.987	28,6	6,8	14,4	16,5	59,8	76,8
Friuli-Venezia Giulia	3.414	50.735	1.520	423.959	2.999	712.591	43,0	5,1	19,2	35,4	37,8	59,5
Emilia-Romagna	6.780	184.063	6.214	1.236.335	10.528	3.013.816	26,6	4,2	27,8	27,0	48,8	68,0
Toscana	5.767	479.580	15.262	2.436.696	1.028	757.016	25,1	13,0	66,5	66,4	8,4	20,6
Umbria	2.480	134.522	5.864	725.050	-	-	29,3	15,6	70,7	84,4	-	-
Marche	2.878	96.718	6.466	1.390.674	-	-	30,8	6,5	69,2	93,5	-	-
Lazio	4.495	291.600	0.302	1.068.806	3.434	3.454.886	26,1	5,1	54,0	34,4	19,0	60,5
Abruzzo	7.050	347.267	3.761	926.393	-	-	65,1	27,3	34,9	72,7	-	-
Molise	2.496	143.018	1.958	147.751	-	-	55,3	49,2	44,7	50,8	-	-
Campania	4.727	335.555	8.658	3.122.999	2.005	2.132.127	34,8	6,0	50,8	55,9	14,7	38,1
Puglia	288	9.100	8.638	1.066.978	10.417	2.840.010	1,5	0,2	45,2	27,0	53,3	72,7
Basilicata	4.721	242.806	4.540	227.953	811	70.140	40,9	45,0	45,1	42,0	8,1	13,0
Calabria	6.379	413.044	7.470	1.151.284	1.372	280.258	41,9	22,4	49,1	62,4	9,0	15,2
Sicilia	6.316	588.804	15.862	2.332.110	3.664	1.880.464	24,5	12,3	61,4	48,6	14,2	30,2
Sardegna	3.281	53.467	10.368	743.255	4.451	762.409	13,6	3,4	67,9	47,1	19,5	49,5
Nord-ovest	27.396	2.357.312	12.564	4.114.191	17.967	9.376.687	47,3	14,9	21,7	26,0	31,0	59,2
Nord-est	26.363	1.659.798	10.382	2.459.872	23.939	7.483.463	46,0	14,3	16,6	21,3	38,4	64,6
Centro	15.828	1.000.429	37.645	6.524.819	1.362	4.211.860	26,9	8,4	65,8	45,6	9,2	35,0
Sud	25.932	1.490.659	35.600	6.532.458	14.008	5.328.544	34,7	11,1	45,3	49,3	19,0	39,8
Isole	9.587	642.381	32.220	3.375.195	8.115	2.862.873	19,2	10,1	64,5	48,2	15,3	41,7
ITALIA	106.308	7.143.779	125.771	22.806.936	69.989	29.033.467	35,2	12,1	41,6	38,7	23,2	49,2

Fonte: Istat, Movimento e calcolo della popolazione residente annuale (R); Variazioni territoriali, disseminazione dei comuni, calcolo della superficie comunale (C).

(a) I Comuni compresi in più di una zona altimetrica sono classificati in un'unica zona, sulla base del criterio della prevalenza della superficie. Cfr. Glossario.
(b) Il dato della popolazione per l'anno 2021 è provvisorio.

La struttura standard di una tabella

Contiene la descrizione dei dati, chiara e concisa

Identifica i
dati
contenuti in
ciascuna
riga della
tabella

Titolo	
Fiancata	Testata
Dati	
Note	
Fonte	

Contiene la
descrizione
del
contenuto
delle
colonne

Per fornire ogni ulteriore indicazione
utile alla comprensione e all'utilizzo
dei dati

Per indicare l'organizzazione che
ha prodotto i dati e la tipologia di
rilevazione

TAVOLA 1. POPOLAZIONE RESIDENTE DI 9 ANNI E OLTRE PER GRADO DI ISTRUZIONE.
Valori assoluti e composizione percentuale

GRADO DI ISTRUZIONE	Anno 2020		Anno 2019	
	v.a.	%	v.a.	%
Analfabeti	51.829	1,0	57.557	1,1
Alfabeti privi di titolo di studio	222.513	4,3	236.311	4,5
Licenza di scuola elementare	813.932	15,7	844.649	16,1
Licenza di scuola media	1.649.656	31,9	1.704.036	32,4
Secondaria	1.741.508	33,7	1.753.602	33,4
Terziaria I livello	160.106	3,1	161.508	3,1
Terziaria II livello	519.942	10,0	484.999	9,2
Dottorato di ricerca/Alta formazione	14.808	0,3	14.444	0,3
CAMPANIA	5.174.294	100,0	5.257.106	100,0

Fonte: Censimento permanente della popolazione – Anno 2020

Frequenze relative e frequenze percentuali

- In alcuni casi può risultare opportuno trasformare le distribuzioni in modo da prescindere dall'ampiezza del collettivo
- Tale obiettivo si raggiunge con le **frequenze relative** (f_i) che si ottengono dividendo ciascuna frequenza assoluta (n_i) per l'ampiezza del collettivo (N):

$$f_i = \frac{n_i}{N} \quad \text{con } i = 1, 2, \dots, u, \dots$$

- Se le frequenze così ottenute vengono moltiplicate per 100 otteniamo le **frequenze percentuali** (p_i)

$$p_i = f_i \times 100 = \frac{n_i}{N} \times 100$$

con $i = 1, 2, \dots, u, \dots$

Le frequenze f_i e p_i esprimono la frazione relativa o percentuale posseduta dalla modalità in questione rispetto al totale

Collettivo di unità statistiche

Carattere

TAVOLA 1. POPOLAZIONE RESIDENTE DI 9 ANNI E OLTRE PER GRADO DI ISTRUZIONE.
Valori assoluti e composizione percentuale

GRADO DI ISTRUZIONE	Anno 2020		Anno 2019	
	v.a.	%	v.a.	%
Analfabeti	51.829	1,0	57.557	1,1
Alfabeti privi di titolo di studio	222.513	4,3	236.311	4,5
Licenza di scuola elementare	813.932	15,7	844.649	16,1
Licenza di scuola media	1.649.656	31,9	1.704.036	32,4
Secondaria	1.741.508	33,7	1.753.602	33,4
Terziaria I livello	160.106	3,1	161.508	3,1
Terziaria II livello	519.942	10,0	484.999	9,2
Dottorato di ricerca/Alta formazione	14.808	0,3	14.444	0,3
CAMPANIA	5.174.294	100,0	5.257.106	100,0

Fonte: Censimento permanente della popolazione – Anno 2020

$$(1.741.508/5.174.294) \times 100 = 33,7$$

Tabella 1 - Popolazione residente per cittadinanza e regione - valori assoluti e percentuali - Anno 2011

Territorio	Popolazione residente		Popolazione straniera	Popolazione straniera
	straniera	italiana	residente (% su totale stranieri)	residente (% su popolazione italiana)
Piemonte	411.095	3.845.255	8,17	10,69
Valle d'Aosta / Vallée d'Aoste	8.090	115.270	0,16	7,02
Liguria	145.465	1.363.762	2,89	10,67
Lombardia	1.155.393	8.787.611	22,97	13,15
Trentino Alto Adige / Südtirol	97.390	976.184	1,94	9,98
Veneto	493.119	4.354.626	9,80	11,32
Friuli-Venezia Giulia	113.151	1.081.496	2,25	10,46
Emilia-Romagna	549.820	3.875.546	10,93	14,19
Toscana	406.508	3.256.683	8,08	12,48
Umbria	89.663	769.149	1,78	11,66
Marche	126.820	1.360.330	2,52	9,32
Lazio	618.142	5.096.740	12,29	12,13
Abruzzo	80.988	1.194.962	1,61	6,78
Molise	11.463	280.687	0,23	4,08
Campania	239.990	5.384.430	4,77	4,46
Puglia	135.173	3.787.768	2,69	3,57
Basilicata	22.184	518.984	0,44	4,27
Calabria	93.257	1.762.197	1,85	5,29
Sicilia	184.605	4.648.724	3,67	3,97
Sardegna	48.400	1.539.013	0,96	3,14
Italia	5.030.716	53.999.417	100,0	9,3

se cambia il **denominatore**
cambia il significato della %
calcolata

stranieri residenti
nella regione «X»

totale stranieri
residenti in Italia

stranieri residenti
nella regione «X»

italiani residenti
nella regione «X»

Fonte: Censimento permanente della popolazione e della popolazione - Anno 2020

Tabella 1 - Popolazione straniera residente per regione - valori assoluti e percentuali - Anno 2011

Territorio	Popolazione straniera	Territorio	Popolazione straniera residente (% su popolazione italiana)
Lombardia	1.155.393	Emilia-Romagna	14,19
Lazio	618.142	Lombardia	13,15
Emilia-Romagna	549.820	Toscana	12,48
Veneto	493.119	Lazio	12,13
Piemonte	411.095	Umbria	11,66
Toscana	406.508	Veneto	11,32
Campania	239.990	Piemonte	10,69
Sicilia	184.605	Liguria	10,67
Liguria	145.465	Friuli-Venezia Giulia	10,46
Puglia	135.173	Trentino Alto Adige	9,98
Marche	126.820	Marche	9,32
Friuli-Venezia Giulia	113.151	Valle d'Aosta	7,02
Trentino Alto Adige	97.390	Abruzzo	6,78
Calabria	93.257	Calabria	5,29
Umbria	89.663	Campania	4,46
Abruzzo	80.988	Basilicata	4,27
Sardegna	48.400	Molise	4,08
Basilicata	22.184	Sicilia	3,97
Molise	11.463	Puglia	3,57
Valle d'Aosta	8.090	Sardegna	3,14
Italia	5.030.716	Italia	9,3

Fonte: Censimento permanente della popolazione e della popolazione - Anno 2020



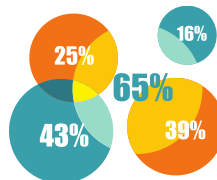
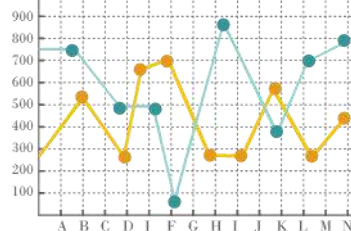
Istat**Learning**
AULA E ONLINE

LA RAPPRESENTAZIONE DEI FENOMENI STATISTICI

I grafici statistici

Oltre che in tabelle e prospetti numerici, i dati statistici possono essere rappresentati in **forma grafica**, attraverso figure costruite in modo che le misure di alcuni loro elementi (segmenti, aree, volumi ...) rappresentino le frequenze o le intensità di un carattere

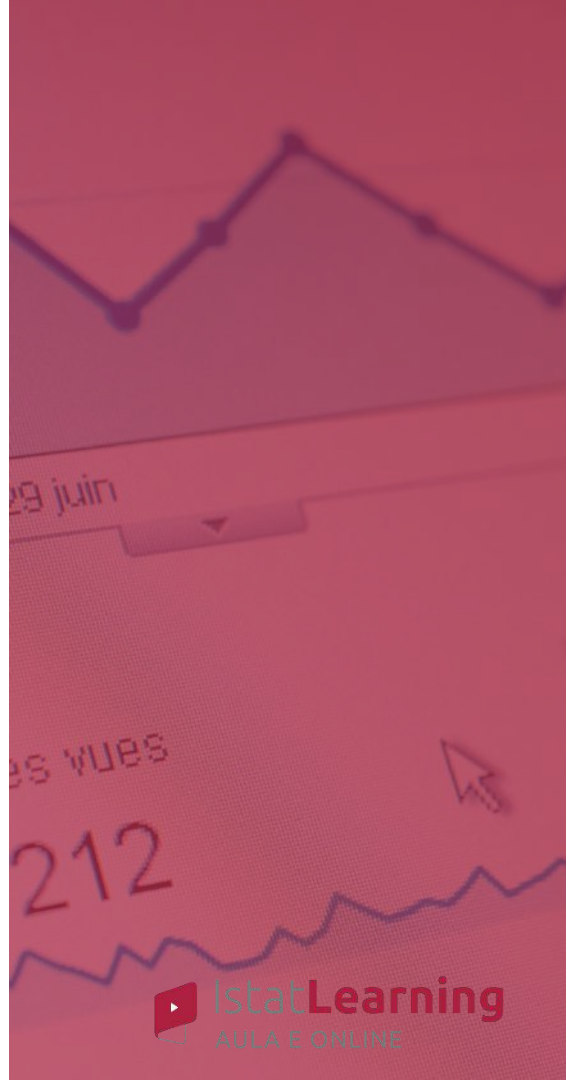
Rispetto alle tabelle, le rappresentazioni grafiche hanno il pregio di porre in evidenza immediata le caratteristiche di un fenomeno



I grafici statistici

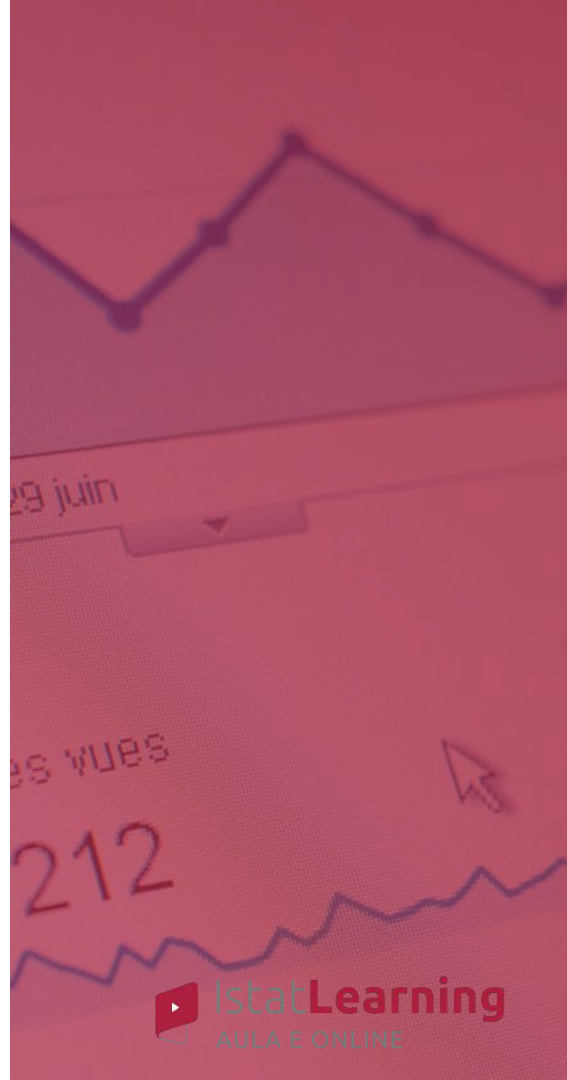
Vantaggi delle rappresentazioni grafiche:

- **Visualizzazione immediata** dell'andamento del fenomeno
- **Sintesi** e quindi possibilità, in poco spazio, di **confrontare più distribuzioni** (curve, spezzate, ecc.)
- **Forma più divulgativa** per i dati statistici rispetto a quanto consentito dalla forma tabellare
- **Potenzialità investigative**, un grafico può
 - mettere in risalto i casi “anomali” (particolari picchi grafici)
 - evidenziare le correlazioni tra caratteri aventi tra loro un legame logico
 - permettere di individuare andamenti di fondo (trend) interpolabili con funzioni matematiche



Le componenti di un grafico

- I **dati**: sono rappresentati in barre, linee, aree o punti
- Le **componenti di supporto**:
 - Il titolo del grafico
 - I titoli degli assi
 - Le etichette degli assi
 - L'unità di misura dei dati
 - La griglia
 - La legenda
 - Le etichette dei dati
 - Le note
 - La fonte dei daticonsentono la comprensione dei dati
- Gli **elementi decorativi**: non sono legati ai dati



Le rappresentazioni grafiche

- Grafici a nastri
- Grafici a barre
- Istogrammi
- Diagrammi circolari
- Grafici radar
- Cartogrammi
- Ideogrammi
- Diagrammi cartesiani

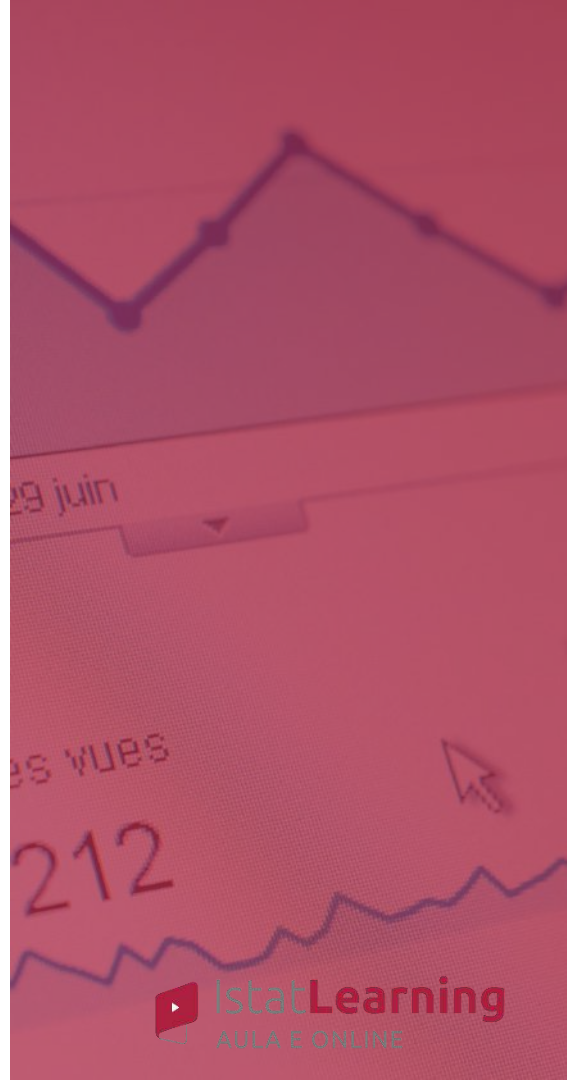
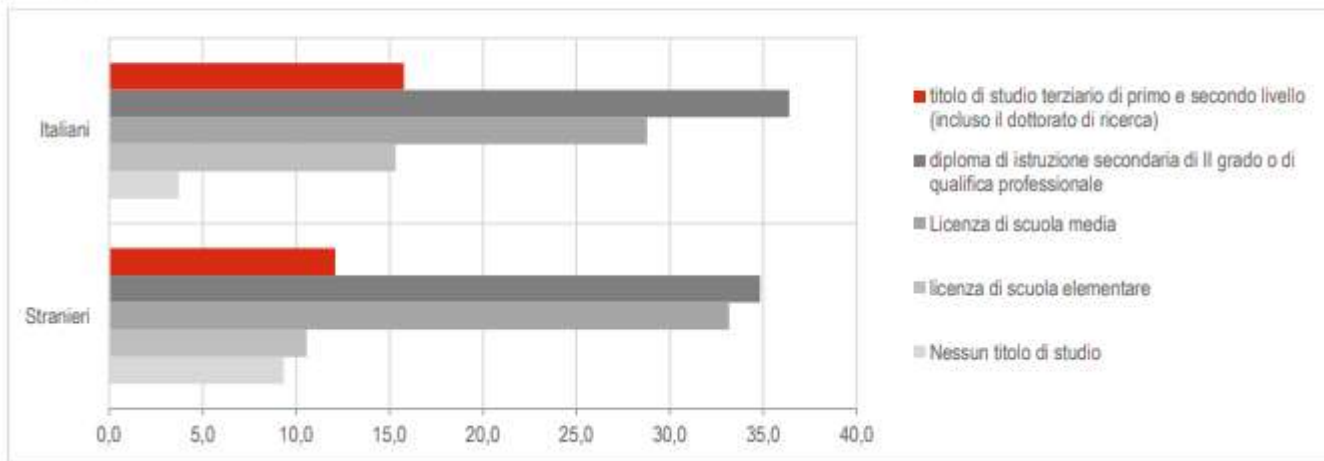


Grafico a nastri

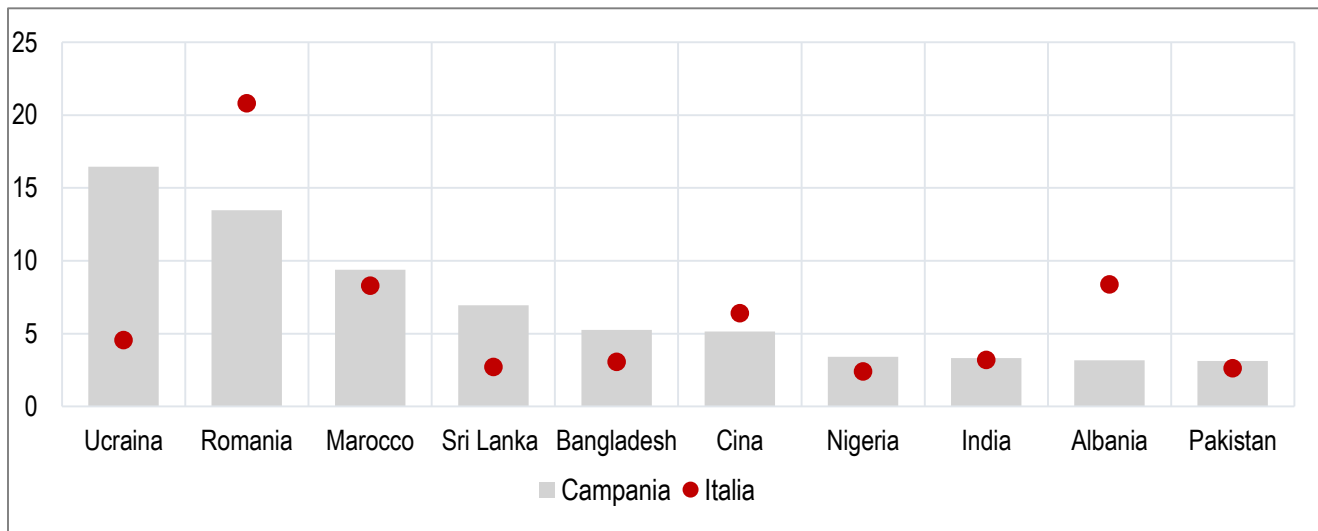
FIGURA 12. POPOLAZIONE DI 9 ANNI E PIÙ PER GRADO DI ISTRUZIONE E CITTADINANZA. Anno 2021, valori percentuali



Fonte: Censimento permanente della popolazione e delle abitazioni - Anno 2021

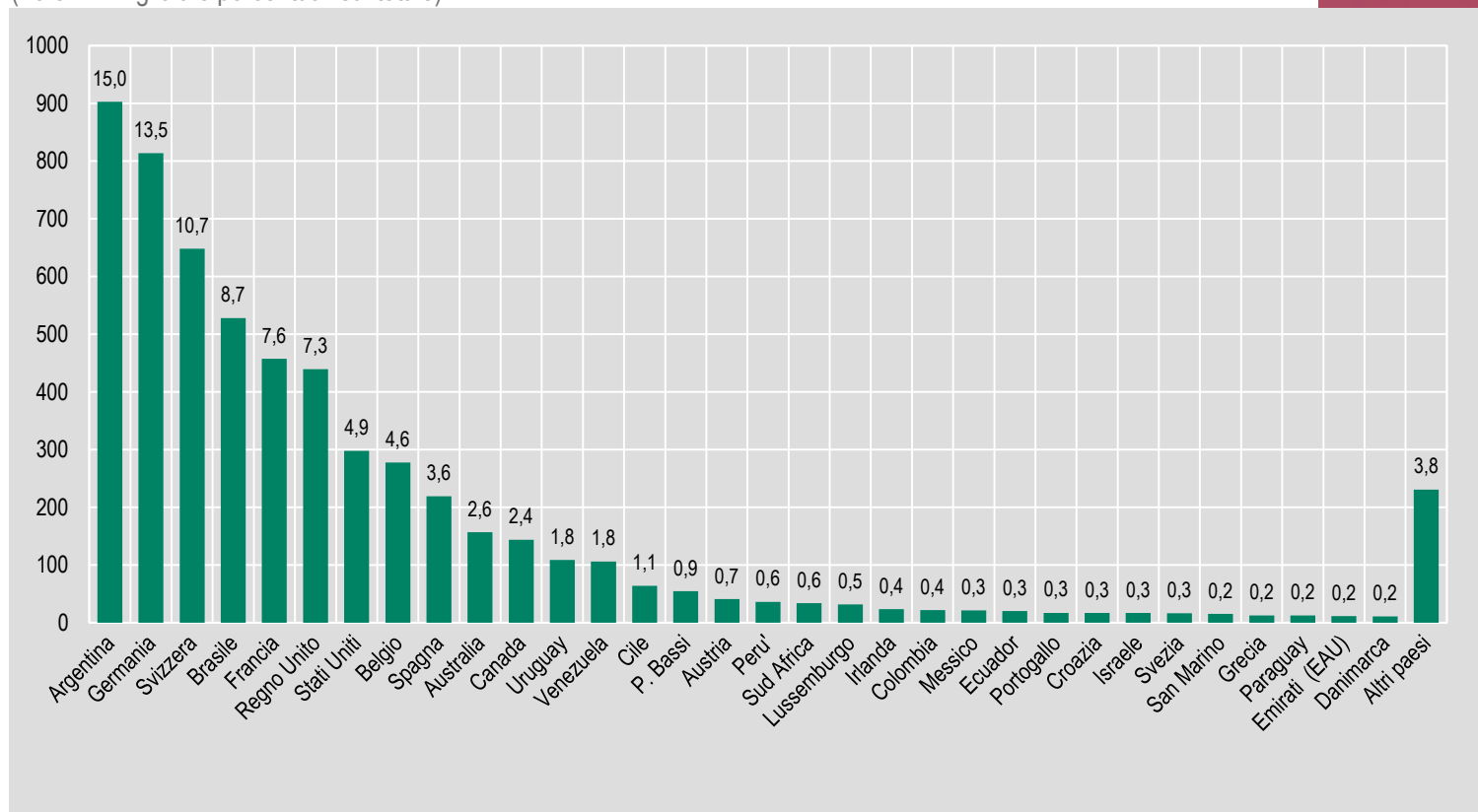
Grafico a barre

POPOLAZIONE STRANIERA PER CITTADINANZA, CAMPANIA e ITALIA. Censimento 2020.
(Valori percentuali per le prime dieci cittadinanze)



Fonte: Censimento permanente della popolazione e delle abitazioni - Anno 2020

Figura 12 - Cittadini italiani residenti all'estero iscritti all'AIRE, per Paese di residenza al 31 dicembre 2021.
(Valori in migliaia e percentuali sul totale)



Fonte: Ministero dell'Interno anagrafe degli Italiani residenti all'estero - Anno 2021

Istogramma

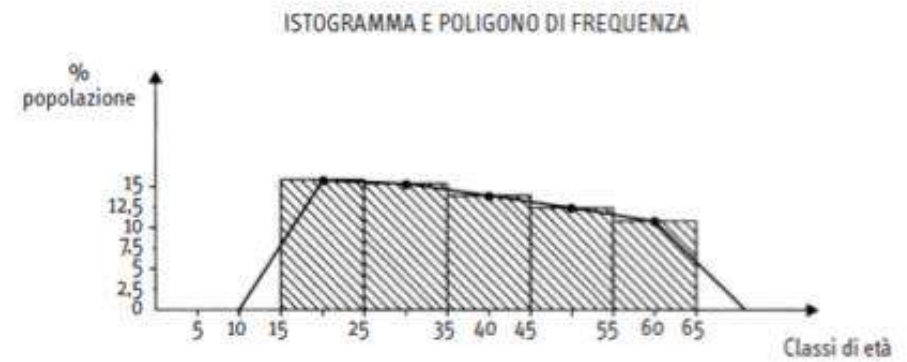
Gli istogrammi si impiegano per rappresentare graficamente distribuzioni di frequenza di caratteri quantitativi le cui modalità sono costituite da classi di valori

A tal fine occorre distinguere due casi:

1. Classi di valori di uguale ampiezza

In questo caso avremo tanti rettangoli contigui, ciascuno avente base uguale all'ampiezza della classe e altezza uguale o proporzionale alla frequenza (assoluta o relativa) assunta nell'insieme delle unità della classe

FIGURA 9 Popolazione residente italiana da 15 a 65 anni per classi decennali di età. Censimento 1991. Valori percentuali sul totale della popolazione italiana residente



Fonte: nostra elaborazione su dati ISTAT.

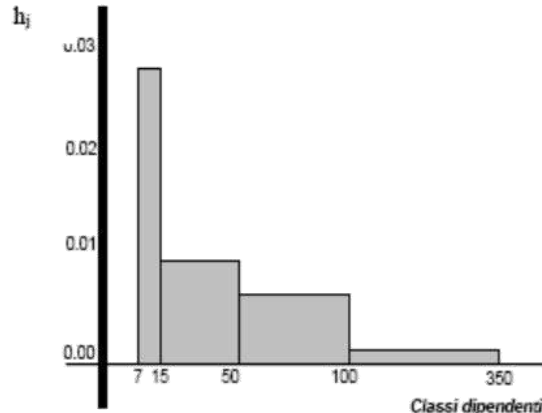
Istogramma

2. Classi di valori di ampiezza diversa

In questo caso avremo una serie di rettangoli aventi basi diverse uguali all'ampiezza delle classi e altezze da calcolarsi, in modo che le frequenze siano proporzionali alle aree dei rispettivi rettangoli

In ordinata, pertanto, avremo le cosiddette densità di frequenza date dal rapporto tra la frequenza (assoluta o relativa) di ciascuna classe e la relativa ampiezza

Classi dipendenti	n_j	a_j	f_j	$h_j = \frac{f_j}{a_j}$
7 — 15	6	9	0,24	0,027
15 — 50	7	35	0,28	0,008
50 — 100	7	50	0,28	0,0056
100 — 350	5	250	0,20	0,0008
Totale	25		1,00	

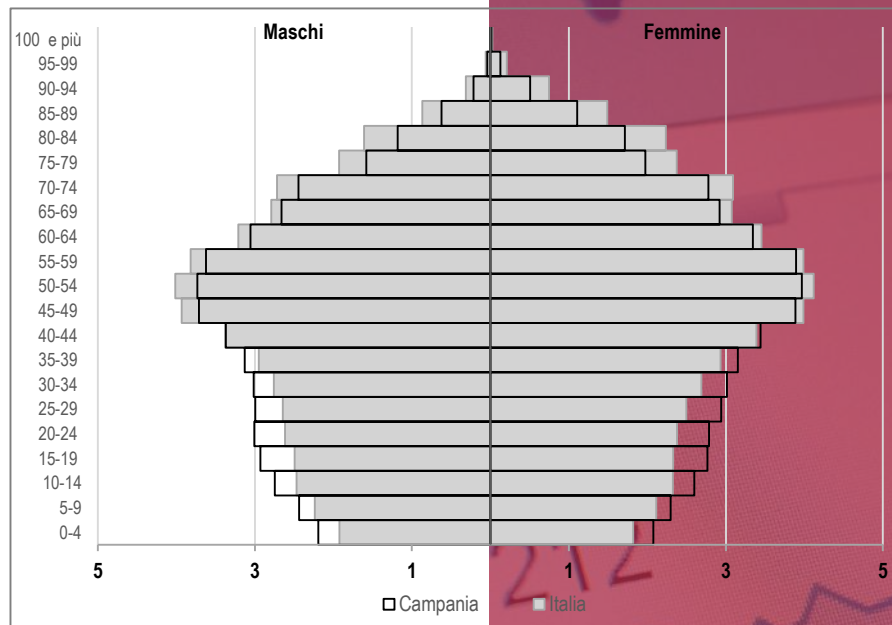


Piramide delle età

La **piramide delle età** è una particolare ed efficace rappresentazione grafica della **struttura per età e sesso** di un dato collettivo o popolazione

È rappresentata da **due istogrammi rovesciati**, corrispondenti ai due sessi, sul cui asse verticale comune sono indicate le modalità del carattere età e sul cui asse orizzontale sono riportate le frequenze (assolute o relative) sia dei maschi che delle femmine, corrispondenti a ciascuna classe di età considerata

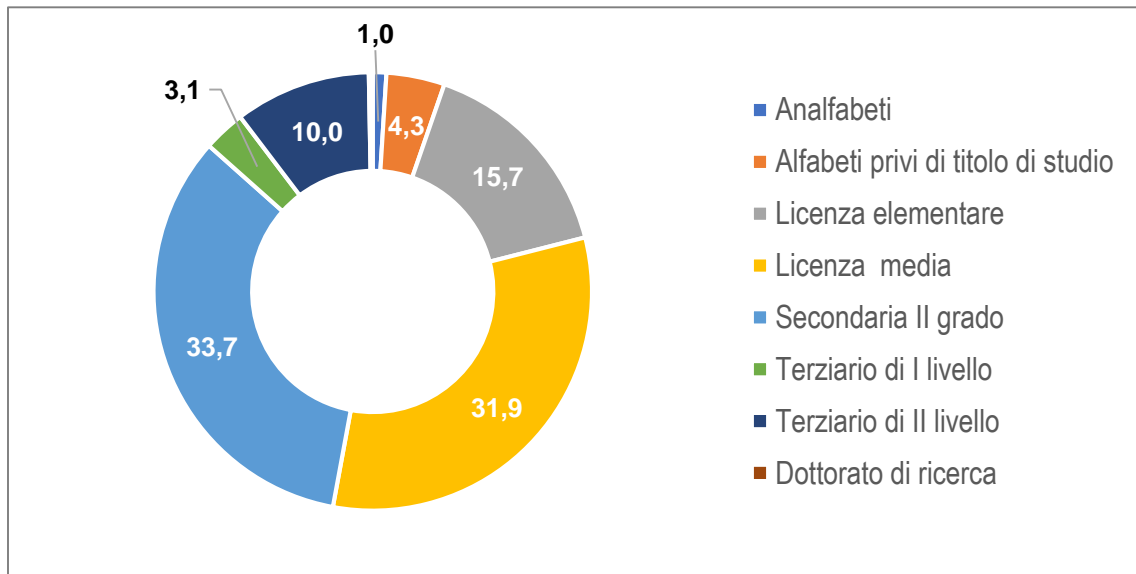
PIRAMIDE DELLE ETÀ DELLA
POPOLAZIONE RESIDENTE,
CAMPANIA E ITALIA.
Valori percentuali



Fonte: Censimento permanente della
popolazione e delle abitazioni - Anno 2020

Diagramma circolare

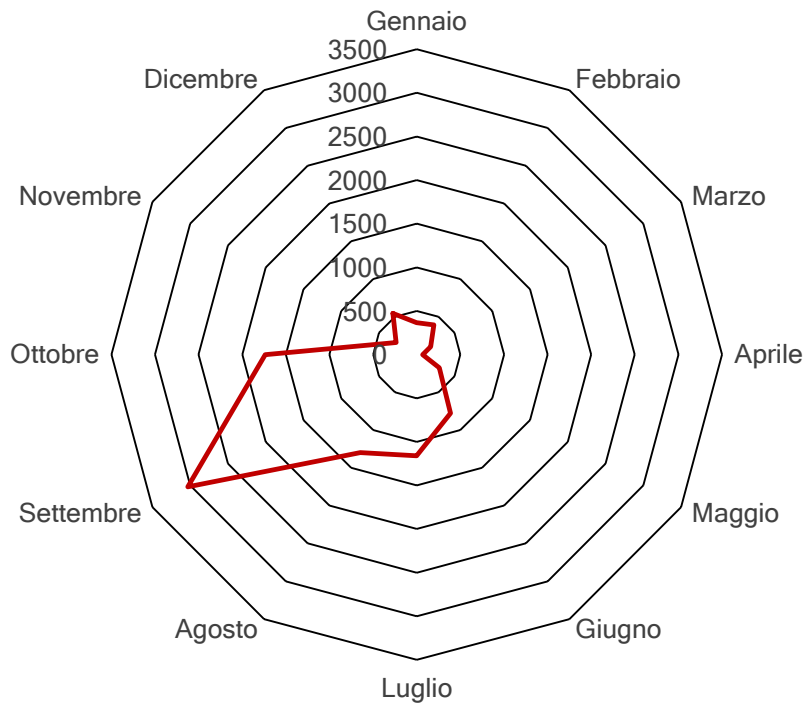
POPOLAZIONE DI 9 ANNI E PIU' PER GRADO DI ISTRUZIONE - CAMPANIA. Censimento 2020



Fonte: Censimento permanente della popolazione e delle abitazioni – Anno 2021

Grafico radar

Matrimoni per mese di celebrazione. Campania - Anno 2020

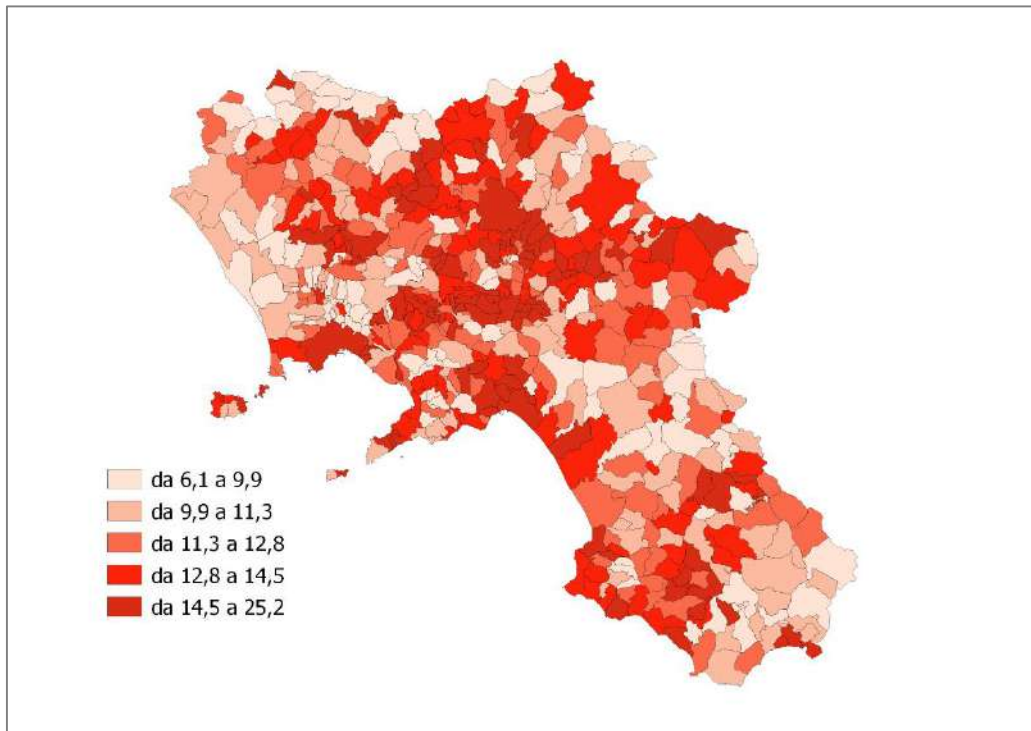


Fonte: Istat

Cartogramma

POPOLAZIONE CON ISTRUZIONE TERZIARIA E SUPERIORE: RAPPORTO TRA LE PERSONE CON ISTRUZIONE TERZIARIA E SUPERIORE E LA POPOLAZIONE DI 9 ANNI E PIU'.

Censimento 2020, valori percentuali, quintili



Ideogrammi

Gli ideogrammi, o pittogrammi, sono forme di rappresentazioni grafiche di carattere divulgativo che si avvalgono di figure, simboli, generalmente tutti simili tra loro, aventi un'immediata attinenza con il carattere considerato (figure umane, oggetti ecc.) e di grandezza o numero variabile per indicare l'entità della frequenza o dell'intensità del carattere rappresentato.

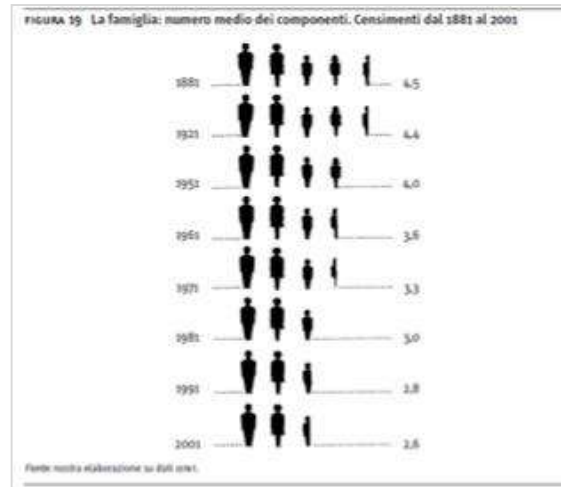
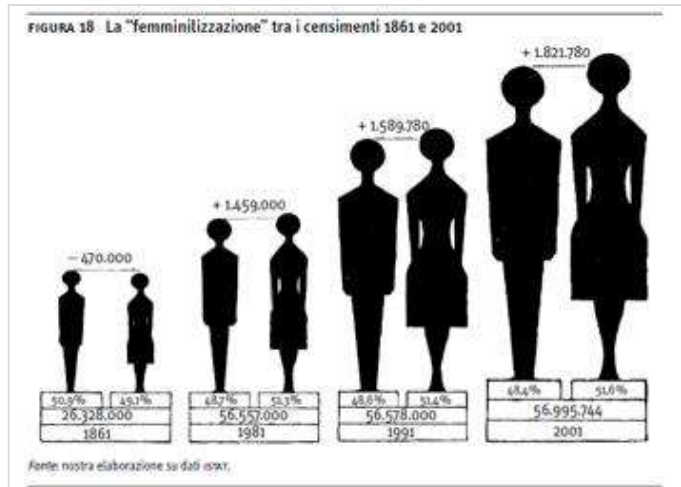
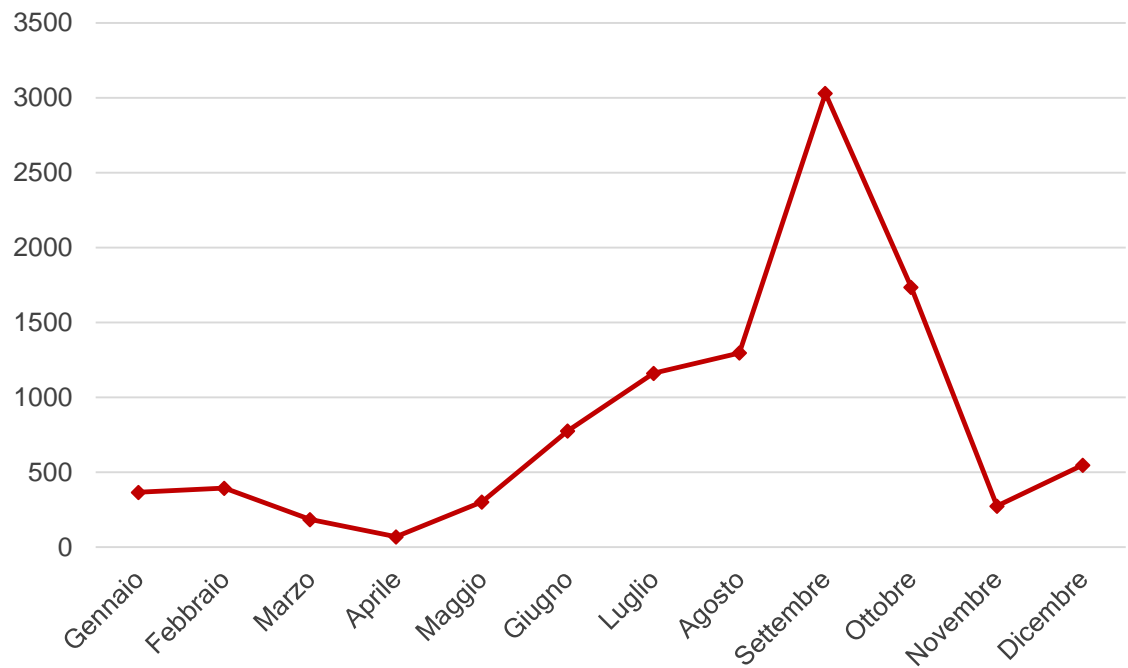


Diagramma cartesiano

Matrimoni per mese di celebrazione. Campania - Anno 2020



Fonte: Istat

Avvertenze per l'uso

Fare riferimento alle tabelle e non solo ai grafici

...inoltre

- Leggere sempre l'intestazione
- Individuare la fonte
- [Leggere i metadati](#)

Tabella 1- Popolazione residente per ripartizione geografica- anno2011

Ripartizioni geografiche	Popolazione residente
Nord	27.283.974
Centro	11.657.380
Mezzogiorno	20.629.227
Italia	59.570.581

Fonte: 15° Censimento generale della Popolazione e delle abitazioni

Spesso un solo indicatore può non essere sufficiente per descrivere un fenomeno complesso

Il "dilemma" di TRILUSSA

"Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:
e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perché c'è un antro che ne magna due"

$$\left[\text{pollo} + 0 \right] / 2 = \text{pollo} (?)$$

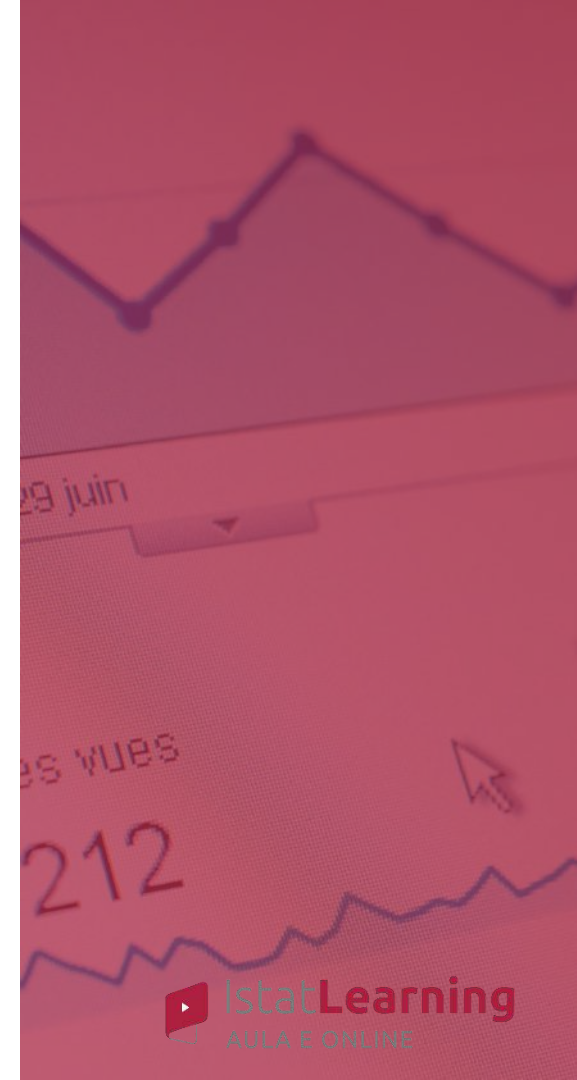
Considerazioni aggiuntive: grafici

Utili quando si richiede di:

- ✓ Effettuare dei confronti tra dati
- ✓ Evidenziare le variazioni nel tempo
- ✓ Evidenziare differenze nelle distribuzioni dei caratteri
- ✓ Evidenziare la correlazione tra variabili
- ✓ Sottolineare il contributo relativo di un item rispetto al totale

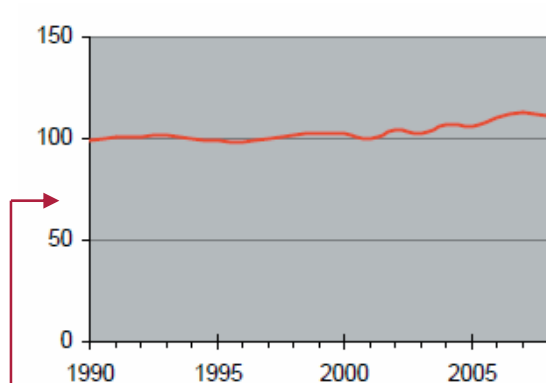
Sconsigliati quando:

- ✗ I dati sono molto dispersi
- ✗ È disponibile un set molto limitato di valori
- ✗ Riportano troppi valori
- ✗ Non mostrano variazioni evidenti nei dati



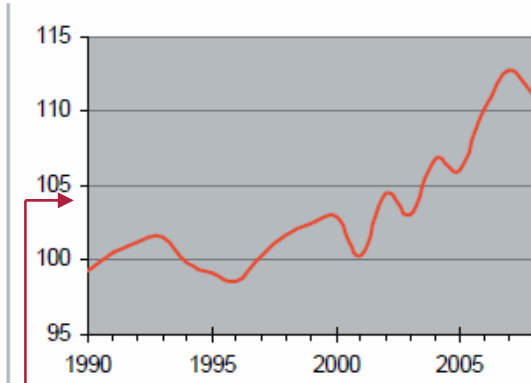
Considerazioni aggiuntive: distorsione del messaggio

Problemi di scala: questi grafici rappresentano la stessa distribuzione ma l'andamento della serie viene percepito in maniera diversa



Fonte: UNECE, Making Data Meaningful Part 2: A guide to presenting statistics, United Nations, Geneva 2009.

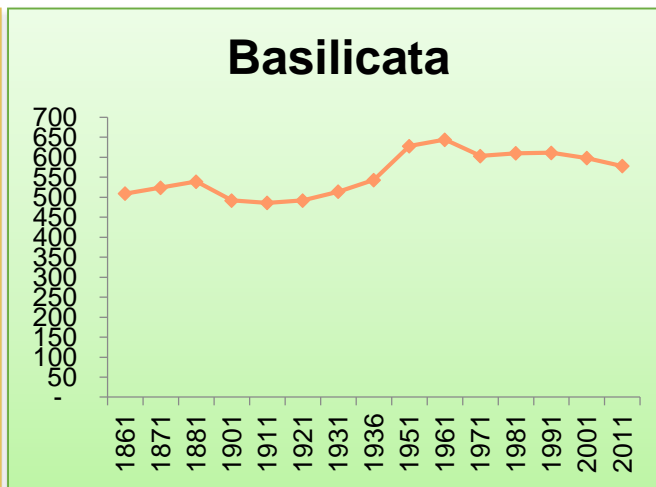
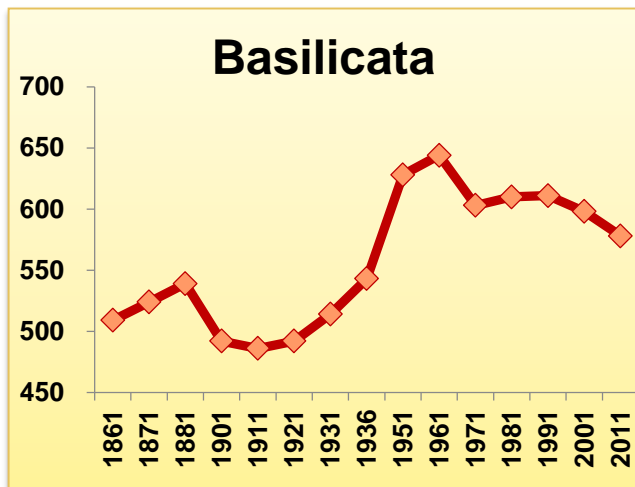
Una scala molto ampia mette in evidenza un trend stabile per circa 10 anni, seguito da una moderata crescita



Riducendo la scala risultano più evidenti le variazioni dei primi 10 anni e sembra più marcata la crescita dell'ultimo periodo

Considerazioni aggiuntive: distorsione del messaggio

Attenzione alla **scala dell'asse Y**: a seconda delle scelte di scala effettuate, possono accentuarsi o perdersi le oscillazioni e i "trend"





LE MEDIE

Le medie

- Le medie sintetizzano l'intera distribuzione di un carattere in un solo valore o in una singola modalità
 - **medie analitiche**: sono il risultato di operazioni algebriche sui valori assunti dal carattere
 - media aritmetica
 - media geometrica
 - **medie di posizione**: si ottengono scegliendo particolari elementi della distribuzione
 - mediana
 - moda
 - quartili

Individuare la media

- La **media aritmetica** è data dal rapporto fra l'ammontare del carattere e il numero dei casi osservati

$$M = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

- La media aritmetica dipende da tutti i valori osservati
→ risente dei valori estremi (c.d. valori anomali)
- A una stessa media aritmetica possono corrispondere distribuzioni molto diverse

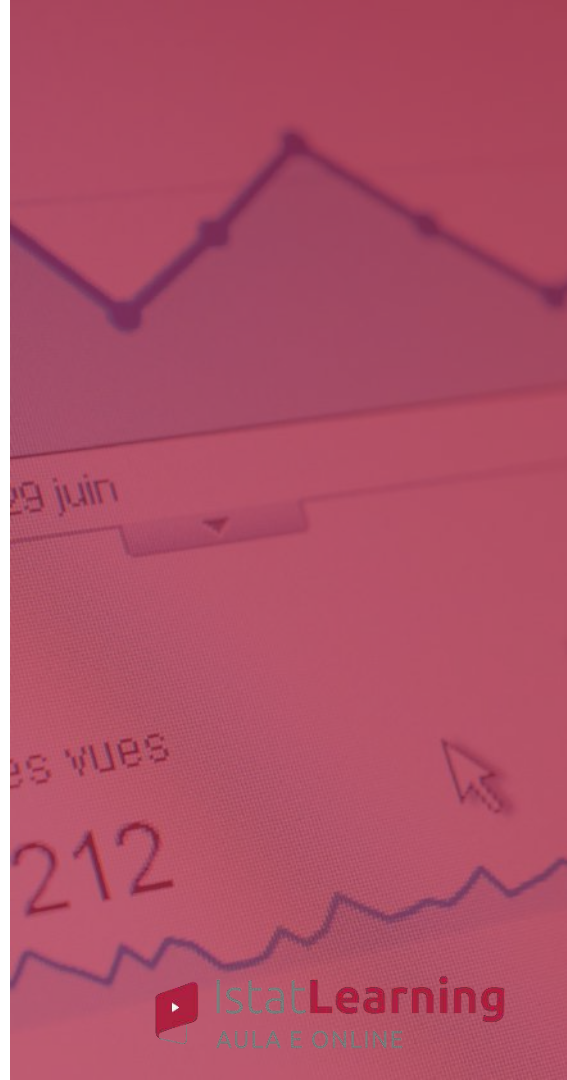
Individuare la media

- La **media geometrica** è una media analitica utilizzata principalmente nel caso in cui l'insieme dei dati osservati è costituito da valori positivi generati da rapporti

$$M_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Individuare la media

- Le medie analitiche sono indici molto sensibili ai valori estremi; esse, inoltre, possono essere calcolate solo se il carattere è quantitativo
- Un indice più robusto (nel senso che è meno sensibile alle oscillazioni dei valori estremi) peraltro calcolabile anche sui caratteri qualitativi ordinabili è la **mediana**
- La mediana è da preferire alla media per tutti quei casi in cui si possono presentare valori estremi molto piccoli o molto grandi (per esempio distribuzioni relative al reddito o ai consumi)



Individuare la media

- La **mediana** di un insieme di unità statistiche ordinate secondo un carattere ordinabile è la modalità presentata dall'unità centrale, dove per unità centrale si intende quella che divide il collettivo in due parti di uguale numerosità
 - una parte formata dalle unità del collettivo che presentano una modalità precedente o uguale a quella dell'unità centrale
 - una parte formata dalle unità del collettivo che presentano una modalità successiva o uguale a quella dell'unità centrale

Individuare la media

- La **moda** è la modalità della distribuzione che si presenta con la frequenza più elevata
 - fornisce informazioni solo su una modalità del carattere
 - dipende solo dalle frequenze
 - acquista validità solo se vi è una netta prevalenza di una modalità/intensità
 - La moda può essere calcolata per qualsiasi tipo di carattere, anche per i caratteri qualitativi sconnessi

Individuare la media

- I **quartili** sono quelle modalità che suddividono la distribuzione in due parti
 - il **primo quartile** è preceduto dal 25% delle unità osservate, aventi modalità uguale o minore



- il **terzo quartile** è seguito dal 25% delle unità osservate, aventi modalità uguale o maggiore



- Il secondo e il terzo quartile individuano un intervallo centrale che contiene circa il 50% delle unità statistiche





LA VARIABILITÀ

La variabilità

- La variabilità di una distribuzione esprime la tendenza delle unità statistiche del collettivo oggetto di studio a manifestare modalità diverse del carattere osservato
 - in presenza di bassa variabilità, una distribuzione statistica può essere descritta adeguatamente dalla media
 - di contro, la media è tanto meno rappresentativa della distribuzione quanto più questa è caratterizzata da dispersione, cioè dalla tendenza delle unità statistiche a discostarsi dal valore centrale



Misurare la variabilità

- **Campo di variazione:** dati "n" valori ordinati in senso crescente $x_1 \leq x_2 \dots \leq x_n$ il campo di variazione è dato dalla differenza fra il massimo e il minimo

$$R = x_n - x_1$$

- **Differenza interquartile**, definita come

$$W = Q_3 - Q_1$$

Poiché il primo e il terzo quartile individuano un intervallo centrale che contiene circa il 50% delle unità statistiche, la differenza interquartile rappresenta il campo di variazione per il 50% delle unità centrali, quelle prossime alla mediana; pur essendo fondato soltanto su due valori, W ha il pregio di non essere influenzato da eventuali dati anomali

Misurare la variabilità

- **Varianza**: calcolata come media aritmetica dei quadrati degli scarti fra ciascun valore e la media della di

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - M)^2$$

- Alla varianza si preferisce la sua radice quadrata, detta **deviazione standard** e indicata con σ , in quanto è espressa nella stessa unità di misura del carattere osservato

Misurare la variabilità

- Sia la varianza sia la deviazione standard sono indici di variabilità assoluti, risentono quindi
 - dell'unità di misura
 - dell'ordine di grandezza dei dati
- Tali indici non consentono di confrontare il grado di variabilità
 - di fenomeni che presentano unità di misura differenti
 - di caratteri che, anche se misurati con la stessa unità di misura, sono riferiti a collettivi molto diversi (pertanto aventi valori medi altrettanto diversi)
- Il **Coefficiente di Variazione**, in quanto indice relativizzato, consente di mettere a confronto la variabilità di due o più distribuzioni

$$CV = \frac{\sigma}{M} 100$$

Misurare la variabilità

- Si parla di **concentrazione** quando la maggior parte dell'ammontare complessivo di un carattere si distribuisce su poche unità statistiche
 - se a ogni unità statistica è riferita la medesima quantità del carattere di ogni altra unità del collettivo, si verifica l'**equidistribuzione**
 - viceversa, quando l'intero ammontare del carattere è posseduto da una sola unità, siamo in presenza della **massima concentrazione**
- I due casi estremi sono da considerare situazioni di riferimento, dalle quali la realtà osservata si allontana con vari gradi di scostamento

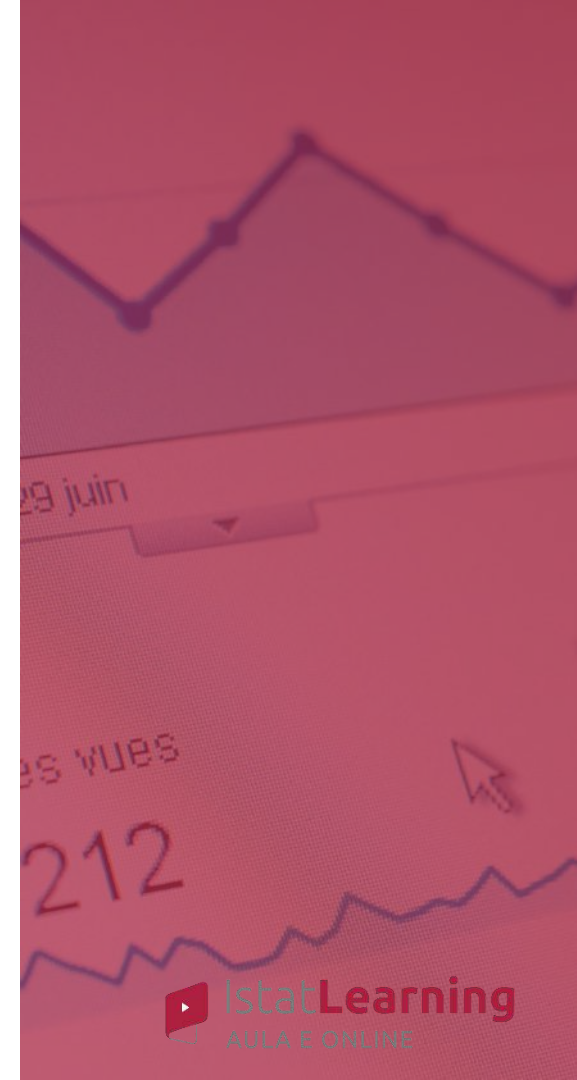


Istat**Learning**
AULA E ONLINE

IL CONFRONTO DEI DATI NEL TEMPO E NELLO SPAZIO

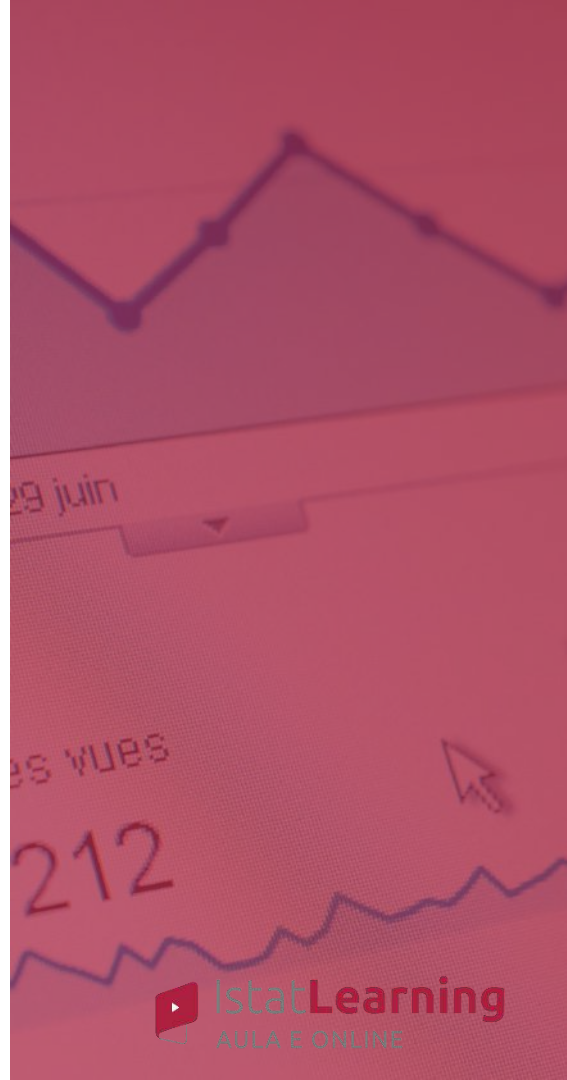
Serie storiche e territoriali

- Si utilizzano per studiare l'evoluzione di un fenomeno nel tempo o confrontare il suo comportamento nello spazio
- La **serie storica** è costituita dalla sequenza cronologica dei valori assunti da una variabile osservata sempre sul medesimo collettivo
- La **serie territoriale** rappresenta le osservazioni effettuate nello stesso momento su collettivi dislocati in luoghi diversi



I rapporti statistici

- I rapporti statistici hanno la funzione di rendere confrontabili grandezze altrimenti incomparabili
- Cause di incomparabilità possono essere
 - Disomogeneità delle grandezze da confrontare
 - Diversa unità di misura
 - Riferimento dei dati a tempi diversi
 - Rilevazione dei dati in luoghi diversi
- Il Rapporto statistico è determinato dal quoziente:
- Tra intensità o frequenze di uno stesso fenomeno riferite a
 - Tempi diversi
 - Luoghi diversi
 - Circostanze diverse
- Tra intensità o frequenze di due diversi fenomeni, almeno uno dei quali collettivo



I rapporti statistici

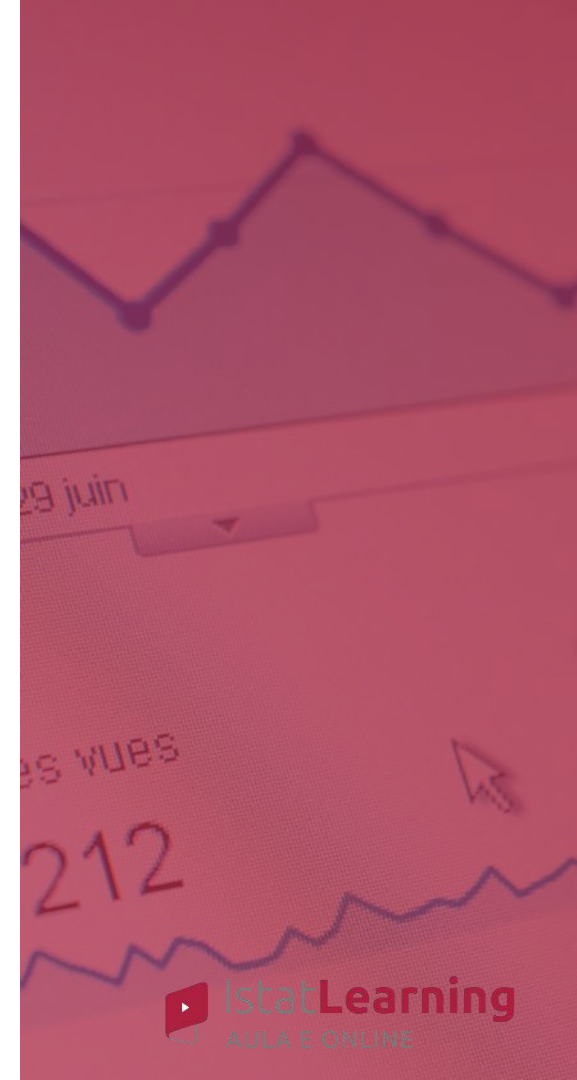
- Quasi sempre un rapporto statistico dà origine a un concetto nuovo, rispetto ai concetti insiti nei due termini del quoziente, per esempio:
numero di nati/popolazione = capacità riproduttiva
- I rapporti statistici costituiscono un metodo di eliminazione, poiché annullano l'effetto del fenomeno che figura al denominatore, per esempio:
PIL Italia **vs** PIL Austria ha scarso significato
ben più rilevante il confronto
PIL pro-capite Italia **vs** PIL pro-capite Austria

I rapporti statistici

- I **rapporti di composizione**: si determinano dividendo l'intensità o la frequenza, concernenti una data modalità di un fenomeno, per l'intensità o la frequenza complessiva dello stesso fenomeno; essi si configurano come quozienti di "parte al tutto", che mettono in evidenza l'importanza relativa dell'intensità o della frequenza considerata
- I **rapporti di coesistenza**: si determinano calcolando il quoziente tra le intensità di due fenomeni diversi nello stesso luogo (o periodo), o tra le intensità di uno stesso fenomeno in due luoghi diversi

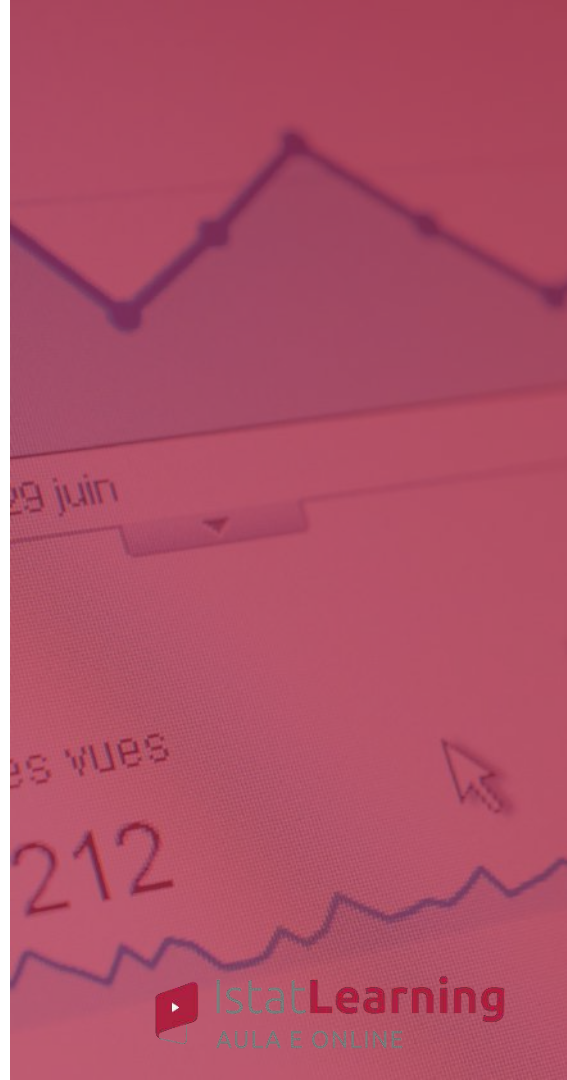
I rapporti statistici

- I **rapporti di derivazione** si determinano dividendo l'intensità o la frequenza di un fenomeno per l'intensità o la frequenza di un altro fenomeno, che ne costituisce il necessario e logico presupposto, e che viene perciò denominato "fenomeno generante"
- I **rapporti di densità** si istituiscono ragguagliando la frequenza complessiva di un collettivo o l'intensità globale di un determinato carattere a una dimensione di spazio o di tempo



Indicatori ambientali e... sostenibilità

- Per indicatori ambientali e di sostenibilità ambientale si intendono valori statistici e parametri utili alla valutazione qualitativa e quantitativa delle condizioni ambientali di un sistema in esame.
- Gli indicatori ambientali sono parametri rilevabili che mostrano le condizioni di un sistema ambientale.
- Gli indicatori di sostenibilità sono indicatori ambientali particolari che aiutano a comprendere se le condizioni ambientali rientrano in determinate aspettative e soddisfano gli obiettivi di sviluppo sostenibile.

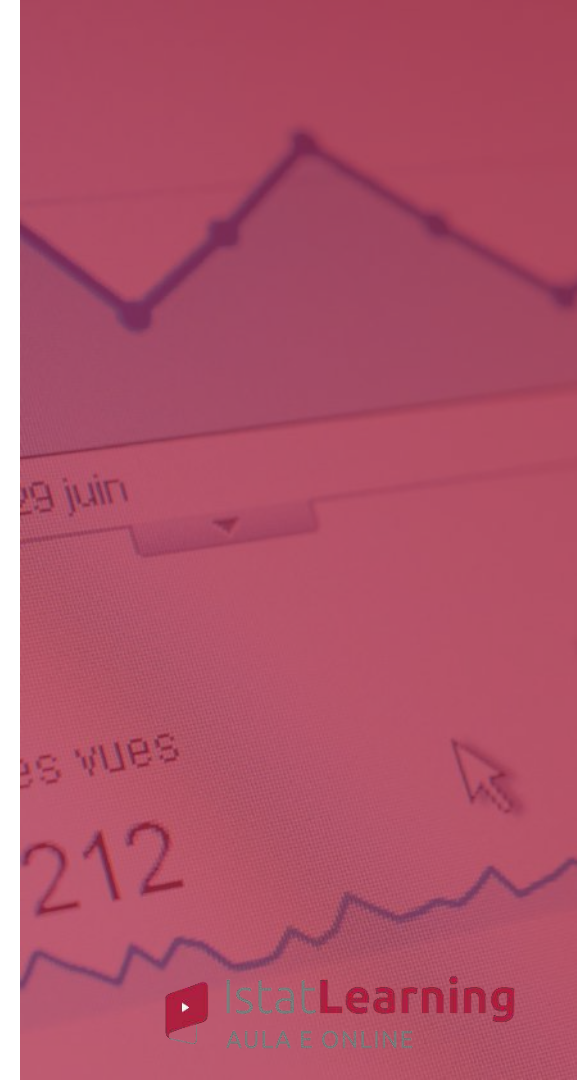


Indicatori elementari

- Input
- Output
- Outcome

Oppure

- Oggettivi
- Soggettivi



Il caso del Covid



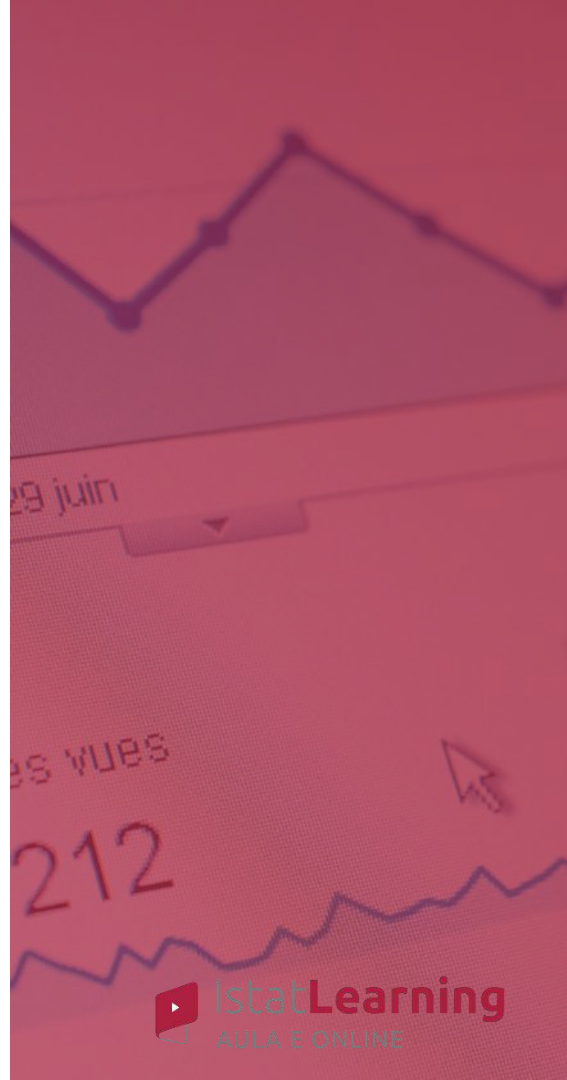
← Questa è una variabile

Questo è un indicatore →



Fenomeni multidimensionali

- Molti fenomeni (positivi e negativi), quali lo sviluppo, il progresso, la qualità della vita, l'innovazione e la competitività, oppure la povertà, la disuguaglianza, la malnutrizione sono molto difficili da misurare poiché sono caratterizzati da una molteplicità di aspetti o dimensioni
- Si dice che sono fattori latenti ossia costrutti teorici inosservabili sottostanti ad una serie di indicatori manifesti e solo stimabili a partire da essi
- In sostanza, al fine di misurarli abbiamo bisogno di una serie di indicatori che debbano essere considerati contestualmente poiché ciascuno è parte costituente del fattore latente stesso

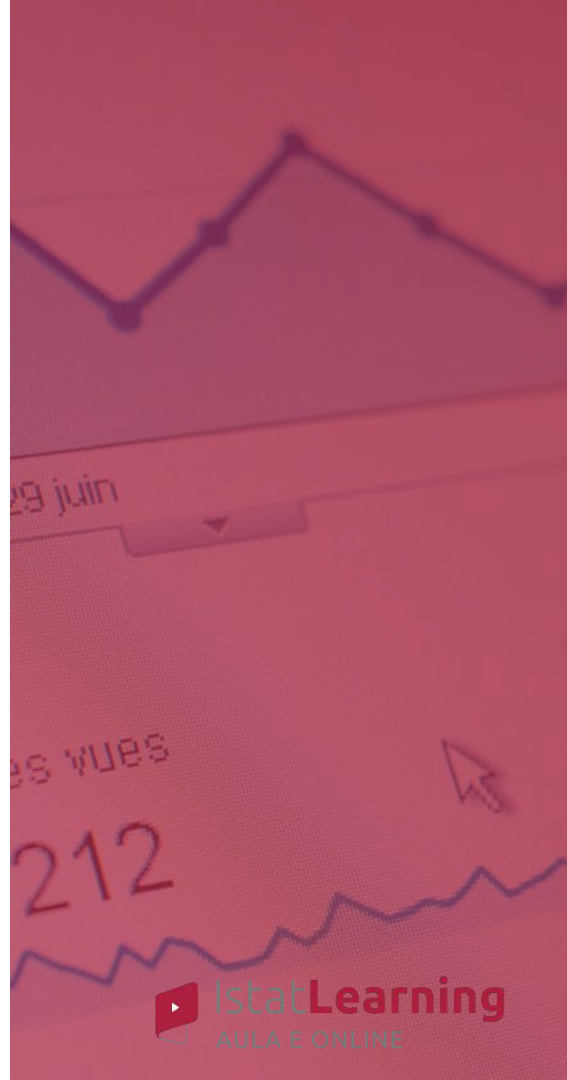


Il Benessere equo e sostenibile

- Il benessere, la qualità della vita sono fenomeni multidimensionali che, al fine di essere misurati, richiedono il coinvolgimento di diversi indicatori elementari, meglio se raccolti in domini espressione del fenomeno stesso
- Il Benessere Equo e Sostenibile misurato dall'Istat (BES) è costituito da 12 domini all'interno dei quali sono raccolti, in totale, circa 130 indicatori socio-economici-ambientali
- Il fenomeno benessere può essere misurato considerando sia il cruscotto di tutti gli indicatori sia una misura sintetica che ne renda più agevole la lettura e ne faciliti la comunicazione ad un pubblico vasto
- Le due misurazioni non si escludono, ma rispondono ad obiettivi differenti

Indici sintetici

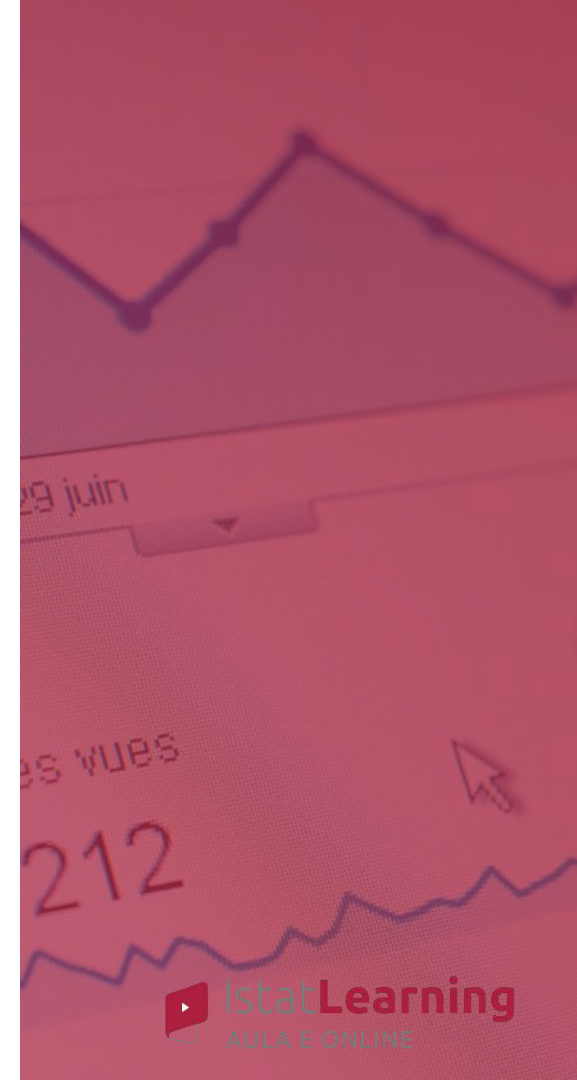
- Un indice sintetico è il composto di singoli indicatori in un'unica misura sulla base di un modello sottostante
- Infatti con un indice composito si vuole fornire un'unica misura che gestisca la complessità interpretativa di un fattore latente
- È molto utile per effettuare comparazioni territoriali e temporali, per fare graduatorie tra aree geografiche, per misurare *performance*
- L'Indice di Sviluppo Umano (ISU o *Human Development Index* - HDI) è un indice sintetico



Il ruolo dell'Istat

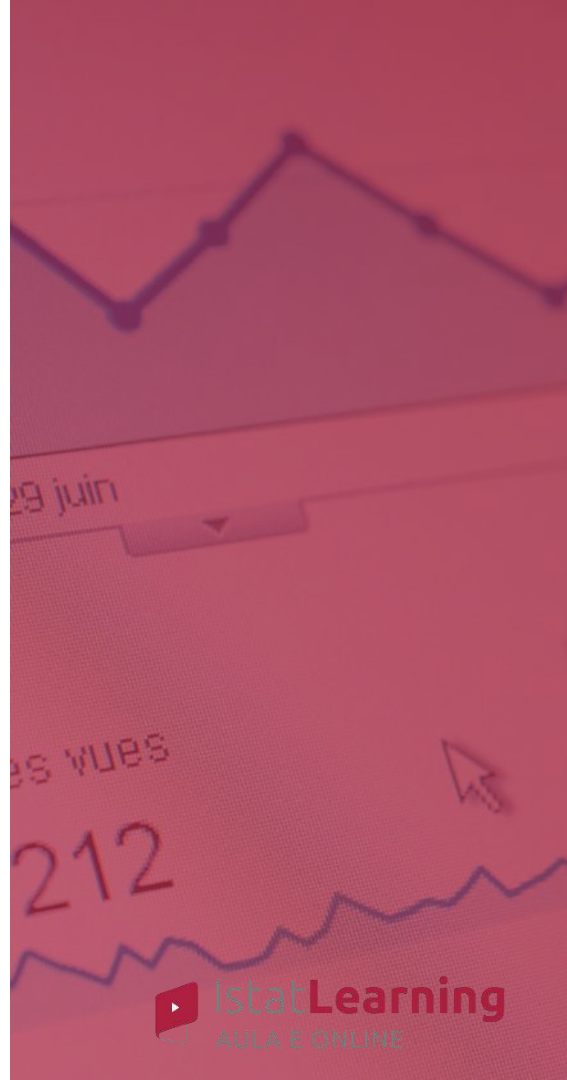
L'Istituto Nazionale di Statistica, all'interno del SISTAN, ricopre un ruolo fondamentale nella costruzione e pubblicazione di indicatori. Tra i tanti si ricordano:

- Indicatori territoriali per le politiche di sviluppo (316 indicatori disponibili a livello regionale e sub regionale, per macroarea e per le aree obiettivo dei diversi cicli delle politiche di sviluppo rilasciati nell'ambito della rete Sistan. Le serie storiche, nella maggior parte dei casi, partono dal 1995 e arrivano fino all'ultimo anno disponibile);
- Indicatori del BES (anche con indici sintetici a livello regionale);
- Indicatori per gli obiettivi di sviluppo sostenibile delle Nazioni Unite;
- Atalante Statistico dei Comuni (database di dati comunali provenienti da fonti ufficiali);
- <http://amisuradicomune.istat.it/aMisuraDiComune/>
- Ruolo delle «nuove» fonti di dati e modalità di interazione/integrazione con le indagini campionarie.



Indicatori e indici per lo sviluppo

- Statistica ufficiale, bene pubblico: uno strumento imparziale per comprendere e decidere
- Misurare per comprendere le differenze economiche, sociali, territoriali, anche in termini di disuguaglianze.
- Il valore della variabilità nella conoscenza delle realtà
- Nella coesione sociale, il ruolo di indicatori e indici per misurare, monitorare, valutare è divenuto fondamentale: strumenti preziosi e insostituibili per le politiche di un paese



Link utili

<https://www.istat.it/it/dati-alla-mano>

<https://www.istat.it/it/metodi-e-strumenti/glossario>

<https://www.istat.it/it/metodi-e-strumenti/classificazioni>

<https://www.istat.it/it/metodi-e-strumenti/strumenti-per-la-qualit%C3%A0/linee-guida>



—
GRAZIE
PER L'ATTENZIONE