

Ufficio Scolastico Regionale per l'Umbria Sistema Nazionale di Valutazione – ciclo di incontri di formazione

La statistica a supporto dell'autovalutazione

Francesca Paradisi

24 febbraio 2015

Indice

Sintesi ed elaborazioni di informazioni statistiche

Rapporti statistici:

- definizione e significato
- classificazione
- rapporti di composizione
- rapporti di coesistenza
- rapporti di derivazione
- rapporti indici (o numeri indici)

Indici statistici di posizione: media, mediana, moda

Indici statistici di dispersione: varianza, deviazione standard, coefficiente di variazione, concentrazione

Correlazione, scomposizione della varianza, inferenza statistica (cenni)

Sintesi ed elaborazioni delle informazioni statistiche



Nel **processo di autovalutazione** le scuole sono chiamate ad un rilevante utilizzo di **indicatori**.

L'utilizzo da parte dei gruppi di autovalutazione si traduce in un duplice compito: l'**interpretazione** degli indicatori messi a disposizione e l'**elaborazione**, interpretazione e contestualizzazione di propri indicatori di area e istituto.

Attraverso gli **indicatori** si *operativizzano* i concetti: da un concetto generale (ad esempio "benessere") si passa ad una sua semplificazione riducendone la complessità, selezionandone alcuni aspetti più significativi e attraverso gli **indici** se ne effettua una misurazione.



Sintesi ed elaborazioni delle informazioni statistiche



Nel **processo di autovalutazione** le scuole sono chiamate a **leggere** una rilevante mole di **dati**.

Nei Rapporti/risultati sono presenti svariati concetti statistici.

Gli **indici statistici di posizione** (medie) e **di dispersione** (variabilità) sono quelli più ricorrenti, alla base di ogni tipo di analisi.

L'esercitazione della giornata verterà proprio nella lettura, interpretazione di dati/risultati illustrati sotto forma di tavole e grafici (cenni di statistica inferenziale)

Sintesi ed elaborazioni delle informazioni statistiche

Tab.1 - Distribuzioni semplici di frequenze

Carattere X	Variabile 1 (frequenze assolute)	Variabile 2 (frequenze assolute)	Variabile 3 (frequenze assolute)
Modalità x_1	n_1	m_1	s_1
Modalità x_2	n_2	m_2	s_2
.	.	.	.
Modalità x_i	n_i	m_i	s_i
.	.	.	.
Modalità x_k	n_k	m_k	s_k
Totale	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k m_i = m$	$\sum_{i=1}^k s_i = s$

➔ **Rapporti statistici**
 rapporti di composizione
 rapporti di coesistenza
 rapporti di derivazione
 rapporti indici

sono costruiti
 operando sulle righe,
 e rappresentano, per
 ciascuna modalità, la
 risultante delle
 variabili coinvolte

↓

Indici statistici di posizione (media, moda, mediana, quantili)
Indici statistici di dispersione (variabilità, campo di variazione,
 scarto semplice medio, varianza, deviazione standard, concentrazione)

sono dei valori di sintesi per ciascuna colonna
 (rappresentano cioè, per ciascuna variabile, tutte le unità di riga)

Rapporti statistici - definizione e significato (1/3)

Rapporto statistico:

- è un quoziente tra due grandezze
- almeno una è di natura statistica (ossia riferita ad un fenomeno collettivo)
- tra le grandezze vi è una chiara relazione logica

Rivestono grande importanza e utilità nell'ambito delle elaborazioni statistiche:

- consentono di effettuare **comparazioni** nel tempo, nello spazio 
- sono utili **indicatori** 

Rapporti statistici - definizione e significato (2/3)

Tab.2 - Numero di scuole primarie. Anno 2013

Territorio	Numero di scuole primarie	Popolazione residente in età 6-10 anni	Scuole primarie per 1.000 abitanti in età 6-10 anni
Umbria	298	39.357	7,6
Lombardia	2.428	475.586	5,1
Italia	17.413	2.832.293	6,1

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione effettuata dal MIUR.

Nota bene

Qualora sia fissato un livello "standard" (come ad esempio nei Piani socio-sanitari regionali il numero di posti-letto per abitante), si può misurare se l'area considerata è sottodimensionata, adeguatamente dimensionata o eccessivamente dimensionata rispetto a tale valore

$$\frac{\text{numero di scuole primarie}}{\text{popolazione residente 6 - 10 anni}} * 1.000$$

Al denominatore la popolazione residente, a metà anno. Calcolata come media aritmetica fra l'ammontare della popolazione a inizio anno e quello a fine anno



Rapporti statistici - definizione e significato (3/3)

- I rapporti statistici, oltre che consentire confronti non altrimenti possibili, offrono anche elementi utili all'investigazione dei fenomeni; in altri termini sono **indicatori** di fenomeni in cui la complessità dei fattori che lei compongono non ne consente una diretta rilevazione.

Indice

Misura empirica diretta

Indicatore

Misura empirica indiretta

di fenomeni (efficienza, salute, intelligenza, qualità della vita) di cui non è possibile una rilevazione diretta tramite indici statistici

Un esempio:



Il reddito pro capite è un **indice** (misura empirica diretta) del benessere economico di un paese. Ma esiste un ampio dibattito se possa essere un buon **indicatore** della qualità della vita: il carattere duale dello sviluppo determina il paradosso che all'aumentare del reddito pro capite aumentano anche i fenomeni di degrado ambientale e sociale che allo sviluppo si accompagnano.

Rapporti statistici – classificazione (1/2)

Classificazione dei rapporti statistici secondo che siano:

- Indicatori di dotazione)



Nota bene

Per le infrastrutture puntuali, qualora il bacino di utenza non coincida con la popolazione che risiede nella provincia in cui è allocata l'infrastruttura stessa, la dotazione dovrebbe essere assegnata a tutte le potenziali aree territoriali mediante un coefficiente di attrattività e non semplicemente rapportate alla superficie territoriale o alla popolazione della provincia in cui sono localizzate.

Indicatori che misurano la dotazione di una risorsa sono costruiti ponendo a numeratore l'offerta della risorsa e a denominatore la popolazione di riferimento, quanto è al servizio della popolazione” o la superficie territoriale, quando la risorsa è al “servizio dello spazio” (Biehl, 1991). Il rapporto tra dotazione e superficie esprime le unità di risorse disponibili per kmq; il rapporto tra dotazione e popolazione, le unità di risorse disponibili per residente.

- Indicatori di funzionalità

Indicatori che misurano gli aspetti di tipo qualitativo della risorsa disponibile. Possono riferirsi al livello di utilizzazione, al grado di rendimento, alla dimensione media, alle caratteristiche organizzative e strutturali in genere.

Rapporti statistici – classificazione (2/2)

Classificazione dei rapporti statistici secondo il legame logico esistente tra numeratore e denominatore:

- Rapporti di composizione

I rapporti di composizione permettono di quantificare l'ammontare relativo ad una modalità rispetto all'ammontare totale

- Rapporti di coesistenza

I rapporti di coesistenza mettono a confronto la frequenza relativa a una modalità con la frequenza di un'altra modalità

- Rapporti di derivazione

I rapporti di derivazione mettono a numeratore un collettivo in movimento e a denominatore un collettivo di stato

- Numeri indici

I numeri indici semplici consentono di confrontare le intensità di uno stesso fenomeno in tempi diversi o in luoghi diversi o in situazioni diverse

Rapporti di composizione (1/2)

- Sono detti anche “parte del tutto”
- Mettono in relazione l'intensità o frequenza di due fenomeni uno dei quali (numeratore) può considerarsi una parte dell'altro. Mettono in relazione cioè l'intensità o la frequenza di una modalità del fenomeno con l'intensità o frequenza del fenomeno complessivo.

$$\text{Rapporto di composizione} = \frac{\text{Intensità di una modalità}}{\text{Intensità del fenomeno totale}}$$

Tra i più noti:



$$\text{Tasso di occupazione} = \frac{\text{Occupati}}{\text{Popolazione 15 +}} * 100$$

$$\text{Giovani che non lavorano e non studiano (Neet)} = \frac{\text{Pop 15 - 29 No studio/lavoro}}{\text{Popolazione 15 - 29 anni}} * 100$$

$$\text{Persone con almeno il diploma superiore} = \frac{\text{Pop 25 - 64 Istruzione sec.}}{\text{Popolazione 25 - 64 anni}} * 100$$

Rapporti di composizione (2/2)

Tab.3 – Numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Regione	Numero di iscritti totali	di cui:			Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
		femmine	ripetenti	stranieri			
Piemonte	165.890	82.510	9.044	15.692	49,7	5,5	9,5
Valle d'Aosta	5.159	2.587	433	305	50,1	8,4	5,9
Liguria	59.464	28.799	3.473	6.187	48,4	5,8	10,4
Lombardia	376.318	185.454	22.920	37.649	49,3	6,1	10,0
Provincia Autonoma Bolzano	21.010	11.403	1.603	1.562	54,3	7,6	7,4
Provincia Autonoma Trento	21.724	11.435	1.308	1.654	52,6	6,0	7,6
Veneto	200.832	98.575	11.084	17.444	49,1	5,5	8,7
Friuli-Venezia Giulia	46.332	22.492	2.881	4.203	48,5	6,2	9,1
Emilia-Romagna	173.163	84.321	9.523	21.873	48,7	5,5	12,6
Toscana	150.058	72.912	8.896	14.847	48,6	5,9	9,9
Umbria	36.555	17.607	1.676	4.131	48,2	4,6	11,3
Marche	70.264	34.145	3.384	6.795	48,6	4,8	9,7
Lazio	247.591	120.284	12.901	18.768	48,6	5,2	7,6
Abruzzo	59.190	28.549	3.259	2.931	48,2	5,5	5,0
Molise	15.197	7.367	702	417	48,5	4,6	2,7
Campania	329.426	158.012	17.959	5.694	48,0	5,5	1,7
Puglia	213.691	103.491	10.945	3.944	48,4	5,1	1,8
Basilicata	30.537	14.558	1.675	609	47,7	5,5	2,0
Calabria	101.679	48.949	4.932	4.190	48,1	4,9	4,1
Sicilia	254.359	123.050	16.878	5.063	48,4	6,6	2,0
Sardegna	74.009	35.903	7.913	1.271	48,5	10,7	1,7
Italia	2.652.448	1.292.403	153.389	175.229	48,7	5,8	6,6

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione effettuata dal MIUR

Rapporti di coesistenza (1/2)

- Mettono a confronto le frequenze o le quantità corrispondenti ad una modalità e la frequenza corrispondente ad un'altra modalità di uno stesso collettivo
- È importante il fenomeno posto al denominatore, perché è rispetto ad esso che viene misurata la variazione, espressa in termini unitari o percentuali, del fenomeno posto al numeratore.

$$\text{Rapporto di coesistenza} = \frac{\text{Intensità di una modalità}}{\text{Intensità di una diversa modalità}}$$

Tra i più noti:

misura il numero di maschi di una popolazione per 100 donne ←

$$\begin{aligned} \text{Indice di mascolinità} &= \frac{\text{Numero maschi}}{\text{Numero di femmine}} * 100 \\ \text{Rapporto dei sessi alla nascita} &= \frac{\text{Nati vivi maschi}}{\text{Nati vive femmine}} * 100 \\ \text{Indice di vecchiaia} &= \frac{\text{Popolazione 65 +}}{\text{Popolazione 0 - 14}} * 100 \end{aligned}$$

Rapporti di coesistenza (2/2)

Tab.4 - Numero di scuole primarie per tipo di gestione della scuola. Anno 2013

Regione	Numero di scuole primarie			Numero di scuole primarie private su 100 scuole primarie pubbliche
	pubbliche	private	totali	
Piemonte	1.292	80	1.372	6,2
Valle d'Aosta	82	3	85	3,7
Liguria	427	52	479	12,2
Lombardia	2.186	242	2.428	11,1
Provincia Autonoma Bolzano	324	3	327	0,9
Provincia Autonoma Trento	215	9	224	4,2
Veneto	1.399	100	1.499	7,1
Friuli-Venezia Giulia	365	23	388	6,3
Emilia-Romagna	949	75	1.024	7,9
Toscana	935	84	1.019	9,0
Umbria	288	10	298	3,5
Marche	442	17	459	3,8
Lazio	1.129	221	1.350	19,6
Abruzzo	426	19	445	4,5
Molise	139	2	141	1,4
Campania	1.538	353	1.891	23,0
Puglia	731	57	788	7,8
Basilicata	206	3	209	1,5
Calabria	859	30	889	3,5
Sicilia	1.435	132	1.567	9,2
Sardegna	509	22	531	4,3
Italia	15.876	1.537	17.413	9,7

$$\frac{\text{scuole primarie private}}{\text{scuole primarie pubbliche}} * 100$$

In Umbria nel 2013 ogni 100 scuole primarie pubbliche sono presenti 3,5 scuole primarie private

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione effettuata dal MIUR

Rapporti di derivazione (1/2)

- Mettono in relazione l'intensità o la frequenza di un carattere con quella di un altro che si ritiene la causa o il presupposto del primo
- Sono dei rapporti fra l'ammontare di un collettivo in movimento e l'ammontare di un collettivo di stato, che è del primo il presupposto
- Si distinguono in **rapporti generici** e **rapporti specifici** a seconda che il fenomeno posto al denominatore possa considerarsi un presupposto generico o un presupposto specifico. E' una maggiore specificazione del collettivo al denominatore che definisce i rapporti di derivazione specifica.

esempio:

$$\text{tasso di fecondità} = \frac{\text{Nati}}{\text{Popolazione femminile } 15 - 49}$$

numero medio di
figli per donna

$$\text{tasso di fecondità specifico} = \frac{\text{Nati da pop. femminile } 15 - 25}{\text{Popolazione femminile } 15 - 25}$$

Rapporti di derivazione (1/2)

Tab.5 - Numero di scuole e iscritti nelle scuole secondarie di primo grado per tipo di gestione. Anno 2013

Regione	Numero di scuole			Numero di iscritti			Numero medio di iscritti per scuola	Numero medio di iscritti per scuola pubblica	Numero medio di iscritti per scuola privata
	pubblica	privata	totale	pubblica	privata	totale			
Piemonte	7.442	708	8.150	1.709.406	70.352	1.779.758	218	230	99
Valle d'Aosta	502	56	558	112.938	6.166	119.104	213	225	110
Liguria	20	1	21	3.589	147	3.736	178	179	147
Lombardia	169	30	199	37.093	2.366	39.459	198	219	79
Provincia Autonoma Bolzano	1.074	194	1.268	258.020	25.696	283.716	224	240	132
Provincia Autonoma Trento	80	8	88	16.530	1.025	17.555	199	207	128
Veneto	76	8	84	15.577	1.329	16.906	201	205	166
Friuli-Venezia Giulia	588	75	663	136.883	7.089	143.972	217	233	95
Emilia-Romagna	154	12	166	30.934	1.214	32.148	194	201	101
Toscana	409	43	452	112.598	5.061	117.659	260	275	118
Umbria	400	31	431	95.414	2.226	97.640	227	239	72
Marche	106	5	111	23.517	205	23.722	214	222	41
Lazio	222	8	230	42.186	466	42.652	185	190	58
Abruzzo	575	101	676	153.425	9.162	162.587	241	267	91
Molise	218	10	228	35.996	506	36.502	160	165	51
Campania	84	..	84	8.638	..	8.638	103	103	..
Puglia	755	55	810	205.485	3.282	208.767	258	272	60
Basilicata	426	14	440	132.280	794	133.074	302	311	57
Calabria	145	..	145	17.391	..	17.391	120	120	..
Sicilia	445	11	456	60.422	504	60.926	134	136	46
Sardegna	666	40	706	166.125	2.532	168.657	239	249	63
Italia	328	6	334	44.365	582	44.947	135	135	97

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione effettuata dal MIUR

Numeri indici (1/4)

- Consentono di confrontare le intensità di uno stesso fenomeno in tempi diversi o in luoghi diversi
 - ✓ Numeri indici temporali
 - ✓ Numeri indici territoriali (o spaziali)
- Si costruiscono ponendo al denominatore un'intensità della stessa natura del fenomeno che è al numeratore, scelta come intensità base e riferita a un dato tempo o situazione. I rapporti possono essere espressi, come avviene generalmente, in termini percentuali
 - ✓ Numeri indici a base fissa (denominatore è fisso sempre uguale all'intensità scelta come base)
 - ✓ Numeri indici a base mobile (denominatore varia e corrisponde all'intensità precedente di quella posta a numeratore)

Numeri indici (2/4)

Quantificare le variazioni intervenute dal 2010 al 2013 nell'ammontare degli insegnanti (Tab.6) e negli iscritti (Tab.7) nelle scuole primarie statali in Umbria rispetto all'Italia nel suo complesso.

Tab.6 – N. di insegnanti nelle scuole primarie statali. Anno 2013

Anno	Numero di insegnanti	
	Italia	Umbria
2010	231.392	3.253
2011	253.894	3.584
2012	246.438	3.573
2013	246.874	3.588

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

Tab.7 – N. di iscritti nelle scuole primarie statali. Anno 2013

Anno	Numero di iscritti	
	Italia	Umbria
2010	2.569.375	36.747
2011	2.573.147	37.054
2012	2.564.914	37.472
2013	2.574.660	37.906

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

Numeri indici (3/4)

I Numeri indici a base fissa (base 2010=100) sono così calcolati e interpretati

Tab.6 – N. di insegnanti nelle scuole primarie statali. Anno 2013

Anno	Numero di insegnanti		Numeri indici (base 2010=100)	
	Italia	Umbria	Italia	Umbria
2010	231.392	3.253	100,0	100,0
2011	253.894	3.584	109,7	110,2
2012	246.438	3.573	106,5	109,8
2013	246.874	3.588	106,7	110,3

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

Tab.7 – N. di iscritti nelle scuole primarie statali. Anno 2013

Anno	Numero di iscritti		Numeri indici (base 2010=100)	
	Italia	Umbria	Italia	Umbria
2010	2.569.375	36.747	100,0	100,0
2011	2.573.147	37.054	100,1	100,8
2012	2.564.914	37.472	99,8	102,0
2013	2.574.660	37.906	100,2	103,2

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

$${}_{10}I_{10}^U = 3.253 / 3.253 \times 100 = 100,0$$

$${}_{10}I_{11}^U = 3.584 / 3.253 \times 100 = 110,2$$

$${}_{10}I_{12}^U = 3.573 / 3.253 \times 100 = 109,8$$

$${}_{10}I_{13}^U = 3.588 / 3.253 \times 100 = 110,3$$

In Umbria il numero di insegnanti nelle scuole primarie dal 2013 al 2010 è aumentato del 10,3 per cento, in Italia la variazione è stata del 6,7 per cento

In Umbria il numero di iscritti nelle scuole primarie dal 2013 al 2010 è aumentato del 3,2 per cento, in Italia la variazione è stata dello 0,2 per cento

Numeri indici (4/4)

I Numeri indici a base mobile sono così calcolati e interpretati

Tab.6 – N. di insegnanti nelle scuole primarie statali. Anno 2013

Anno	Numero di insegnanti		Numeri indici (base mobile per 100)	
	Italia	Umbria	Italia	Umbria
2010	231.392	3.253	-	-
2011	253.894	3.584	109,7	110,2
2012	246.438	3.573	97,1	99,7
2013	246.874	3.588	100,2	100,4

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

Tab.7 – N. di iscritti nelle scuole primarie statali. Anno 2013

Anno	Numero di iscritti		Numeri indici (base mobile per 100)	
	Italia	Umbria	Italia	Umbria
2010	2.569.375	36.747	-	-
2011	2.573.147	37.054	100,1	100,8
2012	2.564.914	37.472	99,7	101,1
2013	2.574.660	37.906	100,4	101,2

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

$${}_{09}I_{10}^u = -$$

$${}_{10}I_{11}^u = 3.584 / 3.253 \times 100 = 110,2$$

$${}_{11}I_{12}^u = 3.573 / 3.584 \times 100 = 109,8$$

$${}_{12}I_{13}^u = 3.588 / 3.573 \times 100 = 110,3$$

In Umbria il numero di insegnanti nelle scuole primarie nel 2013 rispetto all'anno precedente è aumentato dello 0,4 per cento, in Italia la variazione è stata dello 0,2 per cento

In Umbria il numero di iscritti nelle scuole primarie nel 2013 rispetto all'anno precedente è aumentato del 1,2 per cento, in Italia la variazione è stata dello 0,4 per cento

Sintesi di una distribuzione - Indici statistici di posizione e dispersione

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6

Fonte: Istat, I.Stat. Elaborazione e diffusione dei risultati della rilevazione MIUR

Indici statistici di posizione (1/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Media aritmetica	49,1	6,0	6,7
Mediana	48,5	5,5	7,6
Moda	Valle d'Aosta	Sardegna	Emilia-Romagna

In Italia nel 2013 in media ogni 100 iscritti alle scuole secondarie di secondo grado si hanno:

- 49,1 iscritti femmine
- 6,0 iscritti ripetenti
- 6,7 iscritti stranieri

Media aritmetica:
Somma dei valori assunti dalla variabile su tutti le modalità divisa per il numero delle modalità

$$\mu = \frac{9,5+5,9+10,4+\dots+1,7}{21}$$

Indici statistici di posizione (2/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Media aritmetica	49,1	6,0	6,7
Mediana	48,5	5,5	7,6
Moda	Valle d'Aosta	Sardegna	Emilia-Romagna

Mediana:

Termine che occupa il posto centrale della distribuzione di dati ordinati in modo crescente.

1,7
1,7
1,8
2,0
2,0
2,7
4,1
5,0
5,9
7,4
7,6
7,6
8,7
9,1
9,5
9,7
9,9
10,0
10,4
11,3
12,6

Essa lascia alla sua destra e alla sua sinistra un eguale numero di termini

Nel 2013 metà delle regioni italiane, ogni 100 iscritti alle scuole secondarie di secondo grado, hanno più di:

- 48,5 iscritti femmine
- 5,5 iscritti ripetenti
- 7,6 iscritti stranieri

..l'altra metà di meno

Indici statistici di posizione (3/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Media aritmetica	49,1	6,0	6,7
Mediana	48,5	5,5	7,6
Moda	Valle d'Aosta	Sardegna	Emilia-Romagna

Moda:

modalità prevalente del carattere, cioè quella a cui è associata la massima frequenza

Nel 2013 la Valle d'Aosta è la regione italiana che ha registrato un maggior numero di iscritti femmine ogni 100 iscritti alle scuole secondarie di secondo grado. Oltre la metà (50,1 per cento) degli iscritti sono di sesso femminile.

Indici statistici di posizione (4/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Media aritmetica	49,1	6,0	6,7
Mediana	48,5	5,5	7,6
Moda	Valle d'Aosta	Sardegna	Emilia-Romagna

Moda:

modalità prevalente del carattere, cioè quella a cui è associata la massima frequenza

Nel 2013 la Sardegna è la regione italiana che ha registrato un maggior numero di iscritti ripetenti ogni 100 iscritti alle scuole secondarie di secondo grado. Più del 10 per cento (10,7) degli iscritti sono ripetenti.

Indici statistici di posizione (5/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	60,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Media aritmetica	49,1	6,0	6,7
Mediana	48,5	5,5	7,6
Moda	Valle d'Aosta	Sardegna	Emilia-Romagna

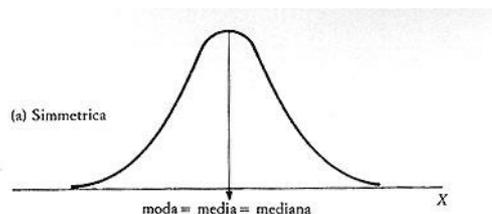
Moda:

modalità prevalente del carattere, cioè quella a cui è associata la massima frequenza

Nel 2013 l' Emilia Romagna è la regione italiana che ha registrato un maggior numero di iscritti stranieri ogni 100 iscritti alle scuole secondarie di secondo grado. Più del 12 per cento (12,6) degli iscritti sono stranieri.

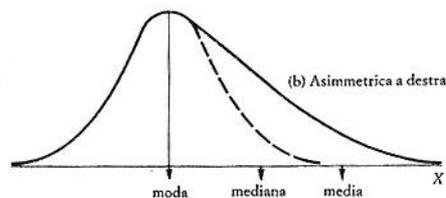
Indici statistici di posizione (6/6)

Le possibili relazioni tra i valori medi di una distribuzione di frequenza dipendono dalla sua forma:



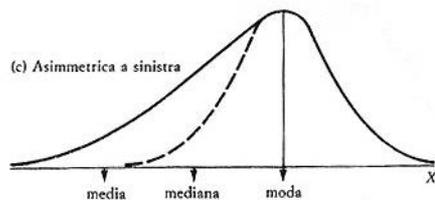
Curva normale o Gaussiana
moda = mediana = media

alta densità di valori al centro e una bassa densità alle due estremità



Curva con asimmetria positiva
moda < mediana < media

la media aritmetica è estremamente influenzata dai valori molto grandi



Curva con asimmetria negativa
media < mediana < moda

la media aritmetica è estremamente influenzata dai valori molto piccoli

Sintesi di una distribuzione – Indici statistici di dispersione (1/2)

Gli **indici di posizione** sintetizzano la realtà fornendo indicazioni su ciò che può considerarsi “tipico” o centrale in un insieme di dati, tuttavia non bastano da soli a sintetizzare i dati.

La media può assumere, infatti, lo stesso valore in distribuzioni che presentano valori anche molto diversi.

ES. Tre studenti, nel primo quadrimestre, hanno riportato le seguenti successioni di voti nelle prove scritte di matematica.

Studente	Voti				Media	Mediana
studente 1	5	6	6	7	6	6
studente 2	4	5	7	8	6	6
studente 3	3	4	8	9	6	6

Se calcoliamo le medie aritmetiche e le mediane dei voti di ogni studente, vediamo che esse coincidono

Nonostante ciò, è evidente che le tre successioni di voti sono diverse tra loro.

Per evidenziare queste differenze si deve introdurre il concetto di **variabilità**.

Sintesi di una distribuzione – Indici statistici di dispersione (2/2)

La **variabilità** è l'attitudine di un fenomeno a presentarsi con modalità diverse.

La variabilità dei dati viene misurata attraverso gli **indici di variabilità** o indici di dispersione, che misurano la dispersione dei termini tra di loro o rispetto ad una media.

Se i dati sono vicini tra loro o al loro valore medio, allora la variabilità è bassa. Se i dati si discostano fortemente tra loro o dal loro valore medio allora la variabilità è alta.

Un indice di variabilità deve soddisfare i due requisiti:

- deve essere nullo se e solo se tutte le unità presentano la stessa modalità del carattere, cioè non c'è variabilità;
- deve essere positivo e aumentare all'aumentare della diversità tra le modalità assunte dalle varie unità.

Indici statistici di dispersione (1/6)

Varianza

Indicatore di dispersione che vale zero solo nei casi in cui tutti i valori sono uguali tra di loro (e pertanto uguali alla loro media) e che cresce con il crescere delle dispersione.

La varianza viene calcolata nel seguente modo:
si calcolano i quadrati degli scarti di ciascun valore dalla media ($x_i - \mu$) e se ne calcola la media aritmetica.

Studente	Voti				Media	Mediana	Varianza
studente 1	5	6	6	7	6	6	0,5
studente 2	4	5	7	8	6	6	2,5
studente 3	3	4	8	9	6	6	6,5

$$\rightarrow \sigma^2 = \frac{(5-6)^2 + (6-6)^2 + (6-6)^2 + (7-6)^2}{4}$$

Indici statistici di dispersione (2/6)

Deviazione standard (scarto quadratico medio)

La varianza esprime meglio la distribuzione dei valori ma l'elevamento al quadrato rappresenta un'anomalia che può essere recuperata solo con un'operazione inversa, l'estrazione della radice quadrata.

Il valore che otteniamo è la deviazione standard, indicata con la lettera σ (sigma), ed è l'indice di variabilità più utilizzato per misurare la dispersione dei dati intorno al valore atteso.

Studente	Voti				Media	Mediana	Varianza	Scarto quadratico medio
studente 1	5	6	6	7	6	6	0,5	0,71
studente 2	4	5	7	8	6	6	2,5	1,58
studente 3	3	4	8	9	6	6	6,5	2,55

$$\sigma = \sqrt{\frac{(5-6)^2 + (6-6)^2 + (6-6)^2 + (7-6)^2}{4}}$$

Indici statistici di dispersione (3/6)

Coefficiente di variazione

Gli indici finora esaminati sono **indici assoluti** perché espressi nella stessa unità di misura del fenomeno in esame.

Se l'obiettivo è confrontare gruppi di dati in base alla loro variabilità si deve ricorrere a misure di variabilità relativa.

$$CV = \sigma/\mu \times 100$$

Il coefficiente di variazione è indipendente sia dall'unità di misura, che dall'ordine di grandezza. Per questo motivo esso è un numero puro.

Studente	Voti				Media	Mediana	Varianza	Scarto quadratico medio	Coefficiente di variazione
studente 1	5	6	6	7	6	6	0,5	0,71	0,12
studente 2	4	5	7	8	6	6	2,5	1,58	0,26
studente 3	3	4	8	9	6	6	6,5	2,55	0,42

→ $\frac{0,71}{6}$

Indici statistici di dispersione (4/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Varianza	2,4	1,9	12,5
Deviazione std	1,5	1,4	3,5
CV	3,1	23,2	52,8

Confronto delle distribuzioni in termini di variabilità:

Possibile mediante il **Coefficiente di Variazione**, che essendo un numero puro, non dipende dall'ordine di grandezza.

La più alta variabilità territoriale si registra per la distribuzione del numero di iscritti stranieri (CV=52,8), seguita dalla distribuzione degli iscritti ripetenti (CV=23,2). La distribuzione degli iscritti secondo il genere presenta la variabilità più ridotta (CV=3,1).

Indici statistici di dispersione (5/6)

Concentrazione

Un particolare aspetto della variabilità di un fenomeno è la **concentrazione**. Lo studio della concentrazione è utile per vedere se il fenomeno è equamente distribuito fra tutte le unità statistiche oppure è concentrato in poche unità.

Due casi limite:

Equidistribuzione: tutte le unità statistiche possiedono la stessa quantità del carattere X:

Unità statistiche: 1, 2, ..., n

Quantità posseduta: μ, μ, \dots, μ

Massima concentrazione: Tutto il carattere è posseduto da una sola unità statistica:

unità statistiche: 1, 2, ..., n-1, n

Quantità posseduta: 0, 0, ..., 0, $N\mu$

Indici statistici di dispersione (6/6)

Tab.8 - Indici sul numero di iscritti alle scuole secondarie di secondo grado. Anno 2013

Territorio	Numero di iscritti femmine per 100 iscritti	Numero di iscritti ripetenti per 100 iscritti	Numero di iscritti stranieri per 100 iscritti
Piemonte	49,7	5,5	9,5
Valle d'Aosta	50,1	8,4	5,9
Liguria	48,4	5,8	10,4
Lombardia	49,3	6,1	10,0
Provincia Autonoma Bolzano	54,3	7,6	7,4
Provincia Autonoma Trento	52,6	6,0	7,6
Veneto	49,1	5,5	8,7
Friuli-Venezia Giulia	48,5	6,2	9,1
Emilia-Romagna	48,7	5,5	12,6
Toscana	48,6	5,9	9,9
Umbria	48,2	4,6	11,3
Marche	48,6	4,8	9,7
Lazio	48,6	5,2	7,6
Abruzzo	48,2	5,5	5,0
Molise	48,5	4,6	2,7
Campania	48,0	5,5	1,7
Puglia	48,4	5,1	1,8
Basilicata	47,7	5,5	2,0
Calabria	48,1	4,9	4,1
Sicilia	48,4	6,6	2,0
Sardegna	48,5	10,7	1,7
Italia	48,7	5,8	6,6
Rapporto di concentrazione Gini	0,013	0,103	0,243

Rapporto di concentrazione del Gini

$$R_G = \frac{\sum_{i=1}^{n-1} (P_i - Q_i)}{\sum_{i=1}^{n-1} P_i}$$

Ordiniamo le quantità del carattere (intensità) possedute dalle unità statistiche in senso non decrescente:

$$0 \leq x_1 \leq x_2 \leq \dots \leq x_n$$

P_i è la frequenza cumulate delle i -unità più povere

Q_i è la quantità relativa di carattere delle i -unità più povere

Esercitazione 1 RILEVAZIONI NAZIONALI DEGLI APPRENDIMENTI 2013-14 - RAPPORTO RISULTATI

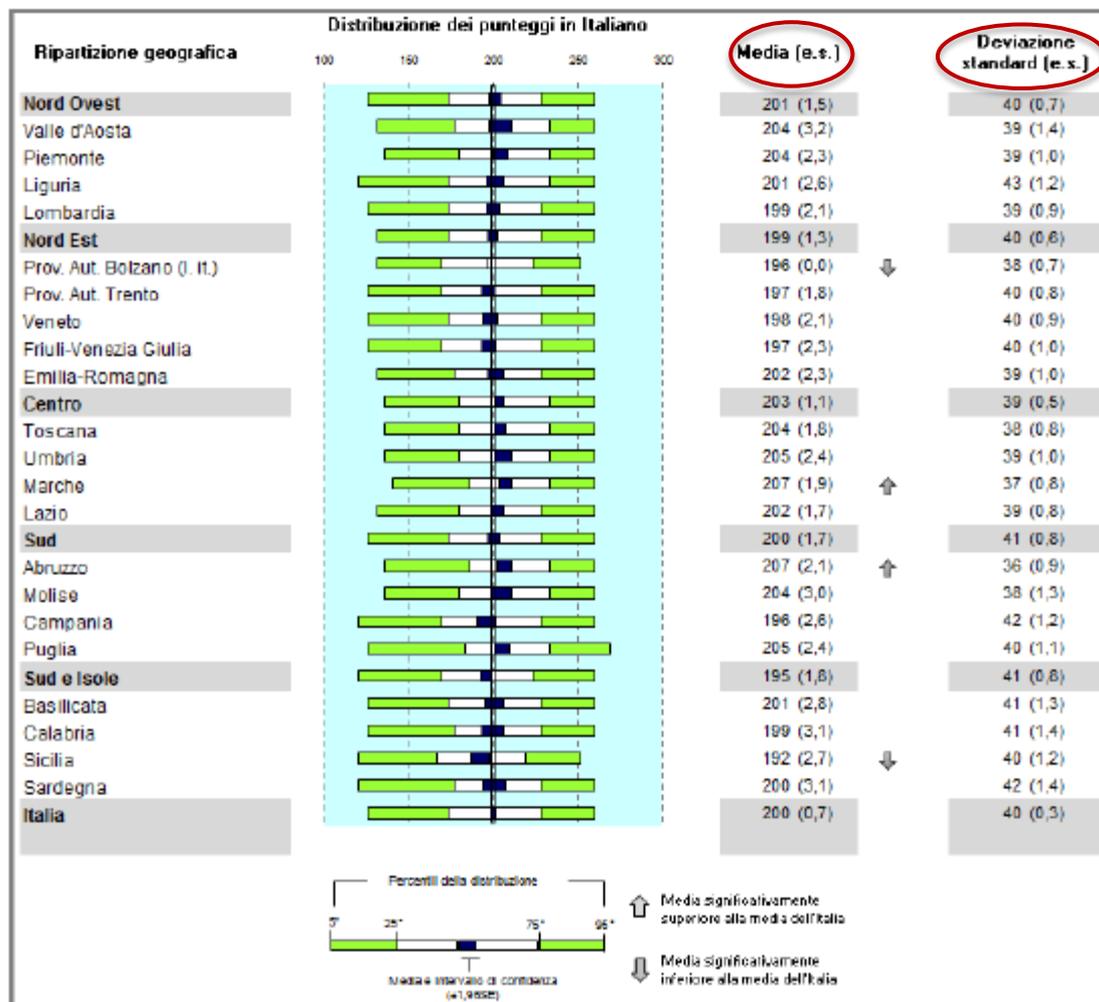


Figura 4.1: Distribuzione dei punteggi della prova di Italiano – classe II primaria

Risultati relativi ad un campione

La **media**, che ne sintetizza la tendenza centrale, e la **deviazione standard**, che esprime la variabilità dei risultati rispetto al valore medio stesso.

Ma siamo nell'ambito della **statistica inferenziale** cioè nella parte di statistica che, in base ad informazioni ricavate da un campione, trae conclusioni sulla popolazione.

E' per questo che, accanto ai valori della media e della deviazione standard, calcolati non sull'intera popolazione ma su di un campione casuale da essa estratto, è posto l'**errore standard (e.s.)**.

L'errore standard fornisce una misura della possibile variazione della media o deviazione standard misurata su un campione statisticamente rappresentativo anziché sull'intera popolazione.

Esercitazione 1 RILEVAZIONI NAZIONALI DEGLI APPRENDIMENTI 2013-14 - RAPPORTO RISULTATI

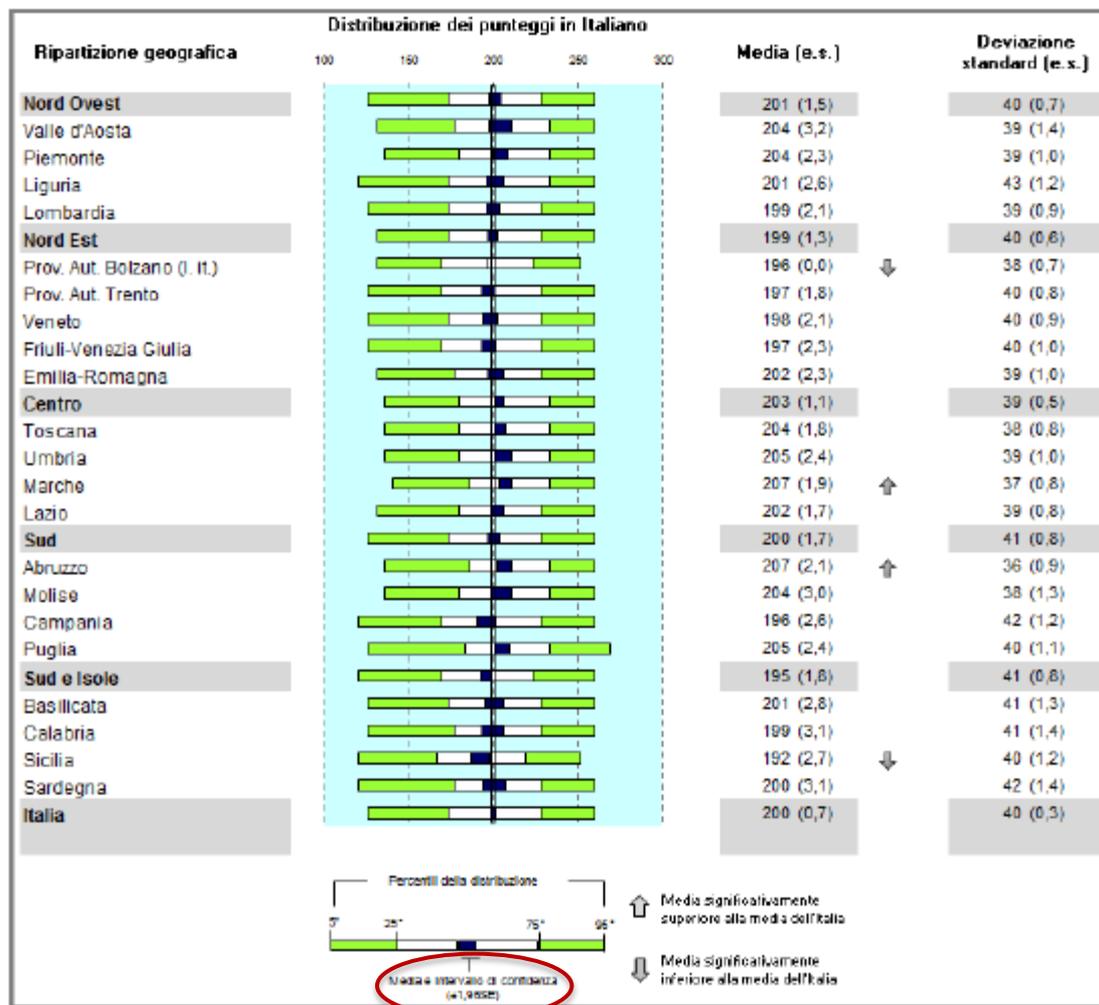


Figura 4.1: Distribuzione dei punteggi della prova di Italiano – classe II primaria

L'**intervallo di confidenza** della media campionaria è l'intervallo di valori intorno alla media campionaria, entro il quale si ritiene sia compresa con un livello di probabilità prefissato (generalmente il 95%) il parametro ignoto (la media) della popolazione.

Nel grafico l'*intervallo di confidenza* della media osservata nel campione è rappresentato dalla zona centrale in blu di ognuna delle barre orizzontali.

Esso misura l'intervallo di punteggi entro il quale oscilla il punteggio "vero" della popolazione.

I limiti superiore e inferiore dell'intervallo sono dati dalla media stimata sul campione più o meno l'*errore standard* di misura, moltiplicato per la costante 1,96.

Esercitazione 1 RILEVAZIONI NAZIONALI DEGLI APPRENDIMENTI 2013-14 - RAPPORTO RISULTATI

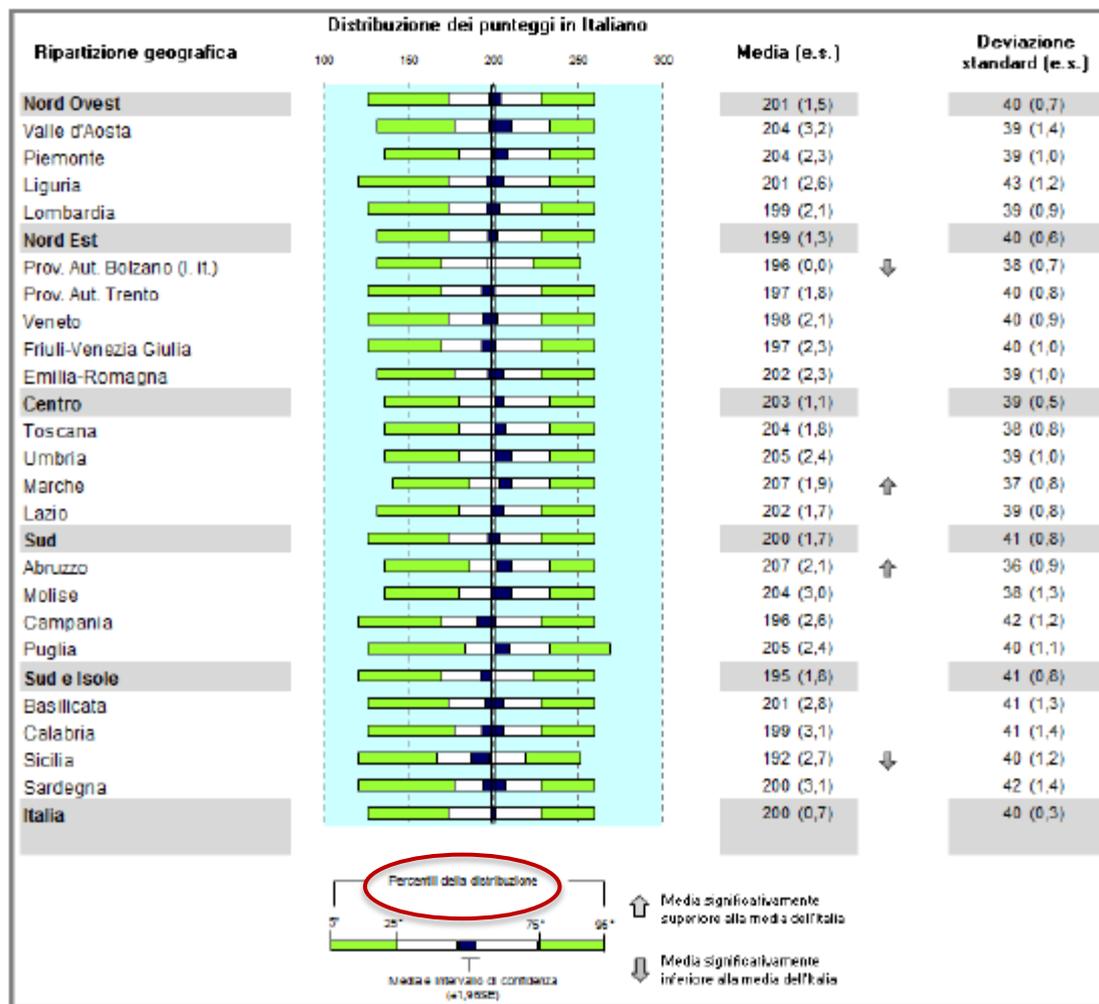


Figura 4.1: Distribuzione dei punteggi della prova di Italiano – classe II primaria

Il **percentile** è un valore della distribuzione che permette di individuare alcune soglie della distribuzione stessa. Supponendo di ordinare i risultati di tutti gli allievi che hanno sostenuto una prova, dal più basso al più alto, il 5° percentile identifica il punteggio che ha conseguito l'allievo che occupa la posizione corrispondente al primo 5% ordinato delle posizioni. (La mediana coincide con il 50° percentile)

I rettangoli verdi all'estremità sinistra e destra di ogni barra rappresentano, rispettivamente, i punteggi compresi fra il 5° e il 25° **percentile**, e tra il 75° e il 95° percentile, mentre i rettangoli bianchi, al cui centro si trova la zona blu che corrisponde all'intervallo di confidenza della media, rappresentano i punteggi situati fra il 25° e il 75° percentile

La lunghezza totale delle barre offre un'immediata rappresentazione dell'ampiezza della dispersione dei punteggi nelle varie aree e regioni

Esercitazione 2 QUESTIONARI STUDENTI, INSEGNANTI E GENITORI – RAPPORTO

Tabella 2. *Correlazione* tra i punteggi medi delle tre scale del questionario insegnanti

	1	2	3	4	5	6
1. Clima scolastico	-					
2. Direzione della scuola	.58	-				
3. Collaborazione con le famiglie	.55	.60	-			
4. Formazione degli insegnanti	.48	.62	.57	-		
5. Collaborazione tra insegnanti	.60	.43	.48	.40	-	
6. Politiche scolastiche	.64	.57	.62	.56	.61	-

Nota. Le correlazioni sono significative con $p < .01$

Correlazione (1/3)

Data una matrice di dati: Tab.1 - Distribuzioni semplici di frequenze

Carattere X	Variabile 1 (frequenze assolute)	Variabile 2 (frequenze assolute)	Variabile 3 (frequenze assolute)
Modalità x_1	n_1	m_1	s_1
Modalità x_2	n_2	m_2	s_2
.	.	.	.
Modalità x_i	n_i	m_i	s_i
.	.	.	.
Modalità x_k	n_k	m_k	s_k
Totale	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k m_i = m$	$\sum_{i=1}^k s_i = s$

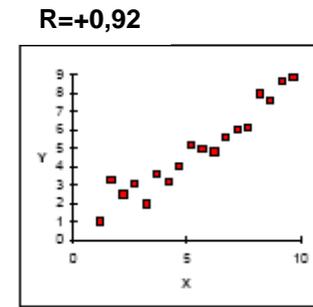
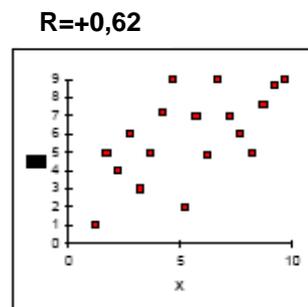
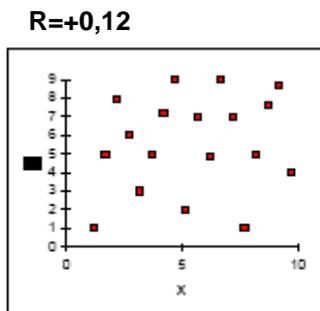
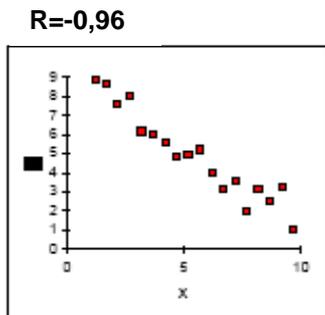
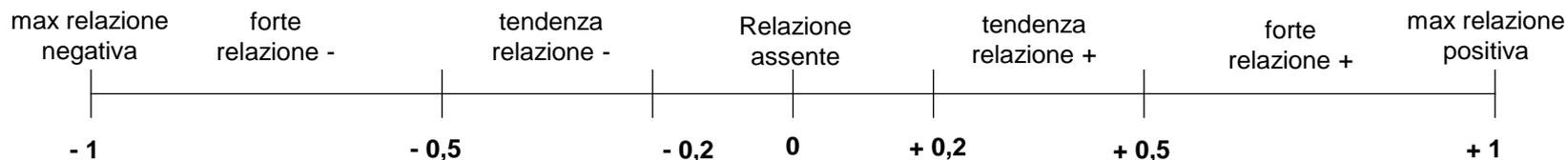
è interessante:

- conoscere se tra due variabili c'è o meno una relazione
- (se c'è relazione) verificarne il verso/direzione: verificare cioè se tra due variabili c'è una relazione positiva o negativa
- (se c'è relazione, positiva o negativa) misurare la forza del legame

Correlazione (2/3)

Il **coefficiente di correlazione di Pearson** permette di studiare il *grado di intensità* del legame lineare tra coppie di variabili.

Il coefficiente di correlazione varia da -1 a 1 e convenzionalmente:



Correlazione (3/3)

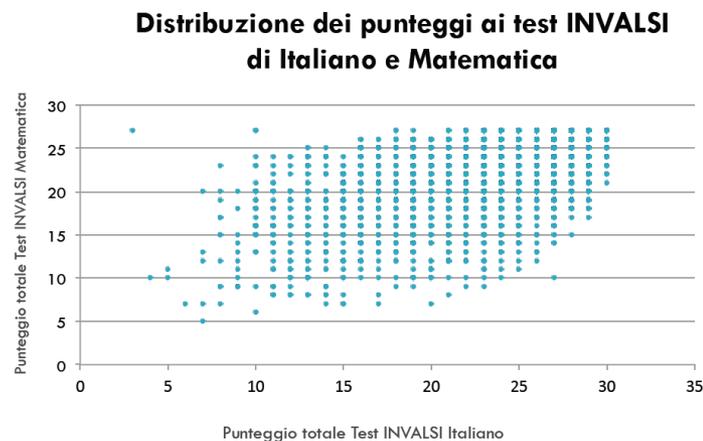
ESEMPIO: Dati Invalsi 2009 sui test di matematica e di italiano su studenti della classe III di scuola secondaria di primo grado (3774 studenti provenienti da 171 scuole).

Ci si chiede se esiste una relazione tra i voti conseguiti in matematica e quelli conseguiti in italiano:

- se non c'è relazione, i risultati conseguiti dagli studenti in uno dei due test non sono legati ai risultati conseguiti nell'altro test
- se la relazione è positiva, gli studenti che ottengono i risultati migliori nel test di italiano, tendono anche ad avere i punteggi più alti nel test di matematica
- se la relazione è negativa, migliori sono i risultati in italiano, peggiori i risultati in matematica

Nell'esempio che utilizza i dati INVALSI:
 $r = 0,54$ quindi tra risultati al test di italiano e al test di matematica c'è una forte relazione positiva.

Gli studenti che ottengono i risultati migliori nel test di italiano, tendono anche ad avere i punteggi più alti nel test di matematica



Esercitazione 2 QUESTIONARI STUDENTI, INSEGNANTI E GENITORI – RAPPORTO

Tabella 2. *Correlazione tra i punteggi medi delle tre scale del questionario insegnanti*

	1	2	3	4	5	6
1. Clima scolastico	-					
2. Direzione della scuola	.58	-				
3. Collaborazione con le famiglie	.55	.60	-			
4. Formazione degli insegnanti	.48	.62	.57	-		
5. Collaborazione tra insegnanti	.60	.43	.48	.40	-	
6. Politiche scolastiche	.64	.57	.62	.56	.61	-

Nota. Le correlazioni sono significative con $p < .01$

**Relazioni positive e
significative**

Esercitazione 3 (OCSE PISA 2012 RAPPORTO NAZIONALE)

Tav. 6.6. – Scomposizione della varianza dei punteggi PISA 2012 e RILEVAZIONI NAZIONALI 2011-12, nel complesso e per i soli 15enni in Grade 10 / “regolari” (valori percentuali)

	Letture / Italiano				Matematica			
	Varianza TRA plessi scolastici	Varianza INTERNA ai plessi scolastici	Varianza TOTALE	Coefficiente di variazione	Varianza TRA plessi scolastici	Varianza INTERNA ai plessi scolastici	Varianza TOTALE	Coefficiente di variazione
PISA 2012								
Tutti gli studenti	57,5	42,5	100	18,4	54,6	45,4	100	17,9
Solo 15-enni in Grade 10 (1)	56,3	43,7	100	16,6	54,3	45,7	100	16,6
Rilevazioni nazionali								
Tutti gli studenti	41,9	58,1	100	28,2	45,8	54,2	100	36,5
Solo studenti “regolari”	39,4	60,6	100	27,1	44,9	55,1	100	35,3
Rilev. nazionali (2)								
Tutti gli studenti	37,6	62,4	100	28,2	41,0	59,0	100	36,5
Solo studenti “regolari”	34,0	66,0	100	27,1	39,8	60,2	100	35,3

Fonte: elaborazioni su dati OCSE-PISA e INVALSI-RN. I dati INVALSI-RN sono corretti per il cheating. Stima HLM- Linear Mixed Model, con (per PISA) pesi-studente normalizzati (la somma dei pesi restituisce il numero complessivo di studenti) e pesi-scuola pari alla somma dei pesi-studente di ogni singola scuola (plesso). (1) Per consentire un confronto più omogeneo con i dati RN, sono esclusi gli studenti dei Corsi di Formazione Professionale (CFP). - (2) Prendendo come unità di analisi la scuola e non il plesso.

Scomposizione della varianza

L'**analisi della Varianza** è una tecnica dell'inferenza statistica che permette di scomporre la variabilità totale.

L'ipotesi alla base dell'analisi della varianza è che dati n gruppi, sia possibile scomporre la varianza in due componenti:

- **varianza tra i gruppi** (Between) misura le differenze esistenti tra un gruppo e l'altro
- **varianza interna ai gruppi** (Within) misura la variabilità esistente attorno alla media aritmetica di ogni gruppo

La variabilità dei dati è dovuta sia al fatto che i soggetti appartengono a gruppi diversi, **VARIANZA TRA GRUPPI**, sia ad una variabilità individuale tra i soggetti anche di uno stesso gruppo, **VARIANZA ENTRO I GRUPPI**

Esercitazione 3 (OCSE PISA 2012 RAPPORTO NAZIONALE)

Tav. 6.6. – Scomposizione della varianza dei punteggi PISA 2012 e RILEVAZIONI NAZIONALI 2011-12, nel complesso e per i soli 15enni in Grade 10 / “regolari” (valori percentuali)

	Letture / Italiano				Matematica			
	Varianza TRA plessi scolastici	Varianza INTERNA al plessi scolastici	Varianza TOTALE	Coefficiente di variazione	Varianza TRA plessi scolastici	Varianza INTERNA al plessi scolastici	Varianza TOTALE	Coefficiente di variazione
PISA 2012								
Tutti gli studenti	57,5	42,5	100	18,4	54,6	45,4	100	17,9
Solo 15-enni in Grade 10 (1)	56,3	43,7	100	16,6	54,3	45,7	100	16,6
Rilevazioni nazionali								
Tutti gli studenti	41,9	58,1	100	28,2	45,8	54,2	100	36,5
Solo studenti “regolari”	39,4	60,6	100	27,1	44,9	55,1	100	35,3
Rilev. nazionali (2)								
Tutti gli studenti	37,6	62,4	100	28,2	41,0	59,0	100	36,5
Solo studenti “regolari”	34,0	66,0	100	27,1	39,8	60,2	100	35,3

Fonte: elaborazioni su dati OCSE-PISA e INVALSI-RN. I dati INVALSI-RN sono corretti per il cheating. Stima HLM- Linear Mixed Model, con (per PISA) pesi-studente normalizzati (la somma dei pesi restituisce il numero complessivo di studenti) e pesi-scuola pari alla somma dei pesi-studente di ogni singola scuola (plesso). (1) Per consentire un confronto più omogeneo con i dati RN, sono esclusi gli studenti dei Corsi di Formazione Professionale (CFP). - (2) Prendendo come unità di analisi la scuola e non il plesso.

I gruppi in cui è suddivisa la popolazione sono rappresentati dai plessi scolastici:

La varianza totale dei punteggi PISA 2012 e Rilevazioni Nazionali 2011-12 è scomposta nelle due componenti:

Varianza TRA plessi scolastici (Within) misura le differenze nei punteggi esistenti tra plessi scolastici

varianza INTERNA al plesso scolastico (Between) misura la variabilità esistente attorno alla media aritmetica di ogni plesso scolastico

L'appartenenza a plessi scolastici diversi ha influenza sulla variabilità?