

Mario Mandras

MODELLI STATISTICI

Elementi fondamentali per la ricerca

EDIZIONI MASTRU

Proprietà letteraria riservata

Printed in Italy

© Copyright 2015 by Mario Mandras

S.V. Taniga san Giacomo Medas 4

Sassari

Indice degli argomenti

1.	Raccolta e organizzazione dei dati	pagina 7
1.1	Le basi della quantificazione	7
1.2	I dati statistici	7
1.3	Tipi di dati	8
1.4	Variabili e loro misurazioni	8
1.5	Diversi livelli di misurazione delle variabili	9
	1.5.1 Scale nominali	9
	1.5.2 Scale ordinali	9
	1.5.3 Scale d'intervalli	11
	1.5.4 Scale di rapporti	11
1.6	Esercitazioni	12
1.7	Terminologia statistica	12
2.	La raccolta di dati da un questionario	13
2.1	Esercitazioni	15
2.2	Terminologia statistica	15
3.	L'organizzazione dei dati in tabelle	17
3.1	I dati del Questionario "Stili cognitivi"	17
3.2	L'organizzazione dei dati	19
3.3	Esercitazioni	21
3.4	Terminologia statistica	21
4.	Calcolo della Media Aritmetica	22
4.1	Esercitazione – Calcolo della Media Aritmetica	26
4.2	L'indice statistico della Media Aritmetica	27
4.3	Un modello utile nel lavoro educativo	28
4.3.1	La tendenza dominante nel comportamento	29
4.4	Esercitazioni	30
4.5	Terminologia statistica	30
5.	Le misure della tendenza centrale	31
5.1	Il valore medio	31
5.2	Le medie algebriche	31
5.2.1	La media aritmetica	32
5.2.1.1	Media totale di più gruppi parziali	35
5.2.2	La media geometrica	35
5.2.3	La media armonica	37
5.2.4	La media quadratica	38
5.2.5	La media cubica	40
5.2.6	Le medie potenziate	41
5.2.7	Le medie antiarmoniche	44
5.2.8	Le medie radicali	46
5.2.9	Le medie di posizione	48
5.3.1	La moda	48
5.3.2	La Mediana	50
5.4	Esercitazioni	52
5.5	Terminologia sta	
6.	Scarti dalla media aritmetica, deviazione standard, indici di variabilità	55
6.1	La differenza o distanza da un valore	55
6.2	La differenza o gli scarti dalla media aritmetica	56
6.3	Gli scarti ponderati	57
6.4	Scomposizione degli scarti	57
6.5	Le devianze	58
6.6	Scarto quadratico medio	59
6.7	Calcolo della Deviazione Standard	60
6.8	Calcolo della D.S. direttamente dai punteggi grezzi	65
6.9	Esercitazione – Calcolo della Deviazione Standard	71
6.10	Variabilità e Deviazione Standard	72
6.11	Altri indici di variabilità	73
	Il campo di variabilità	73
	Le gamme intercentiliche	74
	La deviazione media	75
	Scarto probabile	75
	Coefficiente di variazione	76
	Correzione della varianza per dati raggruppati in classi	76
	Varianza totale di più gruppi parziali	76
6.12	Esercitazioni	77
6.13	Terminologia statistica	77
7.	Differenze medie e indici di concentrazione	80
7.1	Le differenze medie	80
7.1.1	Le differenze medie assolute	81
7.1.2	Le differenze medie quadratiche	82
7.2	La concentrazione	83

7.2.1	<i>La curva di Lorenz</i>	84
	<i>Diagramma di Lorenz sui dati grezzi</i>	84
	<i>Diagramma di Lorenz su dati standardizzati</i>	85
7.2.2	<i>Il rapporto di concentrazione</i>	87
7.2.3	<i>Il rapporto di concentrazione come rapporto di aree</i>	91
7.2.4	<i>L'indice δ (delta) del Gini</i>	95
7.2.5	<i>L'indice di concentrazione del Bonferroni</i>	96
7.3	Esercitazioni	98
7.4	Terminologia statistica	98
8.	I punteggi ponderati	101
8.1	I punti standard o punti zeta	101
8.2	I ranghi percentili	106
8.3	Esercitazioni :Calcolo dei punti zeta e dei ranghi percentili	110
8.4	Il modello statistico dei punteggi ponderati	112
8.5	Terminologia statistica	112
9.	I grafici	113
9.1	Il piano cartesiano	113
9.2	Concetto di funzione	116
9.3	Diagrammi	116
9.4	Diagrammi statistici	117
9.4.1	<i>Il poligono di frequenza</i>	117
9.4.2	L'istogramma	
9.4.3	<i>L'Ogiva</i>	
9.5	Esercitazioni	
9.6	Terminologia statistica	
10.	La curva normale o funzione gaussiana	127
10.1	La distribuzione normale completa	129
10.2	La distribuzione normale standardizzata	120
10.3	Esercitazioni	132
10.4	Terminologia statistica	132
11.	Applicazioni della curva normale	134
11.1	I valori della curva normale	137
10.3.1	<i>La tavola delle ordinate</i>	137
10.3.2	<i>La tavola delle aree</i>	138
10.4	<i>Esercitazioni</i>	141
12.	Indici di forma delle distribuzioni	142
12.1	I momenti	143
12.2	Indici di asimmetria	149
12.2.1	<i>Lo skewness di Pearson</i>	151
12.2.2	<i>L'indice γ_1 (gamma uno) di Fisher</i>	152
12.3	L'indice di curtosi	153
12.4	L'utilità del modello	155
12.5	Esercitazioni	156
12.6	Terminologia statistica	157
13.	Le distribuzioni campionarie	158
13.1	Popolazione statistica	158
13.2	Campione statistico	158
13.3	Distribuzioni campionarie	160
13.3.1	<i>Le medie campionarie</i>	161
13.3.2	<i>La varianza campionaria</i>	166
13.3.3	<i>Errori tipici di vari indici</i>	
13.4	Il campionamento	169
13.5	Esercitazioni	
13.6	Terminologia statistica	
14.	Probabilità e curva normale	176
14.1	Probabilità a priori	176
14.2	Probabilità e frequenza relativa	177
14.3	Proprietà matematiche della probabilità a priori	178
14.4	Limiti di fiducia e livelli di significatività	179
14.5	Stima di parametri	182
14.5	Esercitazioni	185
14.6	Terminologia statistica	186
15.	I test statistici	187
15.1	L'articolazione della ricerca scientifica	187
15.2	Il test "t" di Student	189
15.2.1	<i>Significatività di una media</i>	192
15.2.2	<i>Significatività della differenza tra medie indipendenti</i>	192
15.2.3	<i>Significatività della differenza tra proporzioni indipendenti</i>	193
15.2.4	<i>Significatività della differenza tra medie non indipendenti</i>	195
15.2.5	<i>Significatività di un coefficiente di correlazione</i>	196
15.3	Il test F di Snedecor	197
15.4	Il test del χ^2 (chi quadrato)	202

- 15.4.1 *La distanza dal valore medio in una curva normale* 202
- 15.4.2 *Formule del χ^2* 206
- 15.4.3 *Applicazioni del χ^2 con scale ad intervallo* 207
- 15.4.4 *Verifica della normalità della nostra distribuzione* 208
- 15.4.5 *Confronto tra frequenze del campione e frequenze della popolazione* 209
- 15.4.6 *Verifica dell'indipendenza di due variabili in una tabella di contingenza* 211
- 15.4.7 *Caratteristiche del test χ^2* 214
- 15.5 **Esercitazioni** 215
- 15.6 Terminologia statistica 216

16. Regressione e correlazione 219

- 16.1 Tabelle e diagrammi per variabili in connessione o correlazione 219
- 16.2 Connessione fra caratteri qualitativi 230
- 16.2.1 *Indipendenza e dipendenza di connessione* 230
- 16.2.2 *Contingenza e tabella di contingenza* 231
- 16.2.3 *L'indice chi quadrato (χ^2) di Pearson* 232
- 16.3 Interpolazione statistica e principio dei minimi quadrati 232
- 16.3.1 *L'interpolazione matematica* 233
- 16.3.2 *Interpolazione statistica* 234
- 16.3.2.1 *La condizione dei minimi quadrati* 234
- 16.3.2.2 *L'interpolante lineare e curvilinea* 236
- 16.4 La regressione 239
- 16.4.1 La regressione lineare a minimi quadrati 242
- 16.5 La correlazione 246
- 16.5.1 Il coefficiente di correlazione lineare 247
- 16.5.2 Calcolo del coefficiente di correlazione lineare 247
- 16.5.3 Proprietà del coefficiente di correlazione 251
- 16.6 **Esercitazioni** 252
- 16.7 Terminologia statistica 253

17. La distribuzione binomiale, la gaussiana e la distribuzione di Poisson 255

- 17.1 Cenni di calcolo combinatorio 255
 - 1. *n! fattoriale* 255
 - 2. *Principio fondamentale* 255
 - 3. *Tipi di raggruppamenti* 256
 - 3.1 *Tipologie di raggruppamenti* 256
 - 3.2 *Permutazioni senza ripetizione* 257
 - 3.3 *Permutazione con ripetizione* 257
 - 3.4 *Disposizioni senza ripetizione* 258
 - 3.5 *Disposizioni con ripetizione* 259
 - 3.6 *Combinazioni senza ripetizione* 259
 - 3.7 *Combinazioni con ripetizione* 259
 - 3.8 *Combinazioni con ripetizione* 260
- 17.3 La distribuzione di Gauss 263
- 17.3.1 *Funzione gaussiana e probabilità* 265
- 17.4 *Calcolo di approssimazione della binomiale alla gaussiana* 267
- 17.5 La distribuzione di Poisson 268
- 17.6 **Esercitazioni** 272
- 17.7 Terminologia statistica 273

1. Raccolta e organizzazione dei dati

1.1 Le basi della quantificazione

L'ambiente che ci circonda è caratterizzato dalla varietà degli elementi e dal divenire delle situazioni.

Possiamo identificare i diversi elementi e le mutevoli caratteristiche di oggetti e di eventi con concetti *qualitativi* che fanno riferimento a modalità dell'essere e a classi come : animali-vegetali, mammiferi-uccelli, varie gamme di colore.

I concetti qualitativi che fanno riferimento a modalità opposte costituiscono una prima rozza ed insufficiente misurazione, come caldo-freddo, ampio-stretto, leggero-pesante, lungo-corto, ...

Il confronto di diverse modalità omogenee dell'essere e la valutazione delle differenze pone le basi per un *approccio quantitativo* alla realtà che ci circonda in quanto si stabiliscono il concetto di misura, di unità e di parte dell'intero.

I numeri rappresentano manifestazioni della realtà; le operazioni sui numeri servono a valutare ed interpretare i fatti e gli eventi.

La raffigurazione di superfici e di oggetti consente di sottoporre ad analisi i disegni ottenuti e di stabilire un rapporto tra disegno geometrico e realtà cui si riferisce.

Osservando oggetti ed eventi simili si rileva che esiste tra loro una grande variabilità; così le altezze delle persone i colori delle rose, ecc. sono mutevoli da unità ad unità.

Tutti gli oggetti ed eventi che possono essere raggruppati in classi presentano tra loro infinite differenze.

Volendo descrivere un insieme di dati legati in qualche modo tra loro si ricorre al concetto di *valore medio* della distribuzione e al concetto di *dispersione* o di *variabilità*.

Osservando infine il rapporto tra due eventi o variabili si può rilevare una *dipendenza* oppure una *indipendenza* tra le variabili considerate.

I concetti di media, di variabilità e di dipendenza sono i fondamenti dell'approccio statistico agli eventi del mondo che ci circonda.

1.2 I dati statistici

I dati sono gli elementi primi sui quali opera l'indagine scientifica e statistica. L'intera indagine statistica consiste nell'ordinare, classificare, analizzare ed interpretare i dati considerati.

I dati sono informazioni ottenute dall'osservazione spontanea, da strumenti d'indagine appositamente costruiti o test; i dati possono anche essere ottenuti in modo casuale o per via ipotetica.

I dati possono riferirsi a rilevamenti saltuari, se in una specifica occasione, oppure a rilevamenti sistematici, se raccolti con periodicità o in base ad un piano di ricerca.

I dati, che si intende sottoporre ad indagine statistica, possono essere incompleti, esigui, non attendibili, oppure ampi, rappresentativi.

I dati ricavati dall'osservazione vengono in genere espressi mediante valori numerici.

1.3 Tipi di dati

I dati possono riferirsi ad eventi reali, ad esperienze, al vissuto soggettivo dell'esperienza, o anche ad eventi immaginati o fantasticati.

I dati possono riguardare *l'aspetto fisico* dell'esperienza: si pensi alle dimensioni di lunghezza, massa, forza e tempo. I dati possono anche riferirsi all'aspetto *biologico-vegetativo*, all'aspetto *biologico-istintivo*, all'aspetto *biologico-emotivo* e infine all'aspetto *sociologico-relazionale*; in questi casi l'esperienza è caratterizzata dalla distanza relazionale con l'origine e dall'avvicinamento a nuove fonti vitali.

Emerge pertanto una differenza tra i dati che riguardano l'aspetto fisico della realtà e i dati che riguardano la dimensione biologica e la dimensione relazionale.

Nell'analisi dell'aspetto fisico (si pensi alla dimensione della lunghezza, ad una forza, al tempo) i dati assumono caratteri quantitativi e sono espressi da numeri. Nell'analisi della dimensione biologica e soprattutto relazionale intervengono caratteri qualitativi che rendono approssimativo e parziale l'uso dei numeri per cui risulta problematico e parziale il trattamento matematico.

I caratteri quantitativi possono essere raggruppati in due grandi classi:

1. *I caratteri discontinui o discreti*, quelli che variano per quantità finite e sono per lo più rappresentati da numeri interi e positivi. Sono anche chiamati caratteri enumerabili perché derivano da conteggi: si pensi al numero di soggetti in un gruppo, o al numero di elementi costituenti un intero, ecc.
2. *I caratteri continui*, quelli che possono variare liberamente. Si pensi alle possibili misure di lunghezza, di peso, ecc. I caratteri continui sono il risultato di una qualsiasi *misurazione*. Misurare significa rapportare la grandezza in esame con una specifica ed appropriata unità di misura; questo rapporto viene espresso con un numero reale. Così se la lunghezza di un palo, rapportato al metro, è di 3,27 si dirà che il palo misura 3,27 metri. I caratteri continui possono essere raggruppati in classi, cioè discretizzati.

1.4 Variabili e loro misurazioni

Il concetto di variabile deriva dal linguaggio matematico e nella sua prima accezione indica il tipo di misurazione metrica. Nelle scienze del comportamento il concetto di variabile si è ampliato tanto da comprendere "il criterio in base al quale viene suddiviso una determinata popolazione di oggetti" (Boudon, 1970), o di osservazioni. Esempi di variabili sono il sesso, l'età, l'altezza, il grado d'intelligenza ecc. Pertanto il concetto di variabile presuppone due o più livelli o valori. Tutte le caratteristiche che si distribuiscono in grado diverso tra le persone sono chiamate variabili: intelligenza, attitudini, interessi ecc. Tutto ciò che nell'ambiente circostante determina una risposta o reazione da parte di un organismo può chiamarsi variabile. Variabile è lo stimolo, semplice o complesso, che determina un comportamento da parte di un organismo. Variabile è anche ogni tipo di comportamento che caratterizza un determinato individuo.

E' fondamentale la distinzione tra variabile indipendente e variabile dipendente. La variabile indipendente è la causa presunta di un determinato comportamento, mentre la variabile dipendente ne sarebbe l'effetto.

Le variabili controllate dallo sperimentatore, in quanto le può variare sistematicamente o mantenere costanti, sono le variabili indipendenti. Si dia il caso che lo sperimentatore desideri osservare le risposte di un gruppo di studenti ad una domanda sulla Divina Commedia; il quesito sulla Divina Commedia rappresenta la variabile indipendente, mentre le risposte fornite dagli alunni sono la

variabile dipendente. Lo sperimentatore se vuole confrontare il rendimento dei singoli studenti deve porre a tutti la medesima domanda sulla Divina Commedia nelle stesse condizioni.

1.5 Diversi livelli di misurazione delle variabili

E' facile misurare la variabile altezza o peso delle persone. Ma è possibile misurare i colori dell'iride, l'intelligenza, l'aggressività, l'adattamento, la sicurezza e le capacità in genere di una persona?

Stevens (1951) afferma che misurare equivale assegnare dei numerali ad eventi od oggetti rispetto ad una regola.

Una scala di misura si costruisce attribuendo alcune caratteristiche dei numeri a determinate proprietà di oggetti. Ricordiamo inoltre come la funzione della misurazione sia quella di connettere un concetto con la realtà allo scopo di permettere al ricercatore di verificare sperimentalmente le sue teorie e proposizioni. La misurazione pertanto rende possibile l'interazione tra elaborazione teorica e verifica sperimentale.

Stevens (1951) distingue quattro livelli di misurazione delle variabili: le scale nominali, le scale ordinali, le scale di intervalli e le scale di rapporti. La scala di livello superiore permette tutte le operazioni della scala di livello inferiore, oltre alle proprie.

1.5.1 Scale nominali

Costruire una scala significa assegnare numeri, o altri simboli, ad una proprietà di oggetti per poter attribuire in questo modo alcune caratteristiche dei numeri alla proprietà presa in esame.

Le caratteristiche dei numeri sono numerose; così rileviamo che in una serie di numeri ogni numero è diverso, può essere maggiore o minore di un altro, può essere sommato, moltiplicato ecc. con un altro.

Le scale nominali usufruiscono solo della proprietà per cui un numero è diverso da tutti gli altri. Altri simboli che servono ad indicare le proprietà delle scale nominali sono le lettere dell'alfabeto, in quanto ogni lettera è diversa dalle altre.

Le scale nominali servono a suddividere in categorie, mutualmente escludentesi, le variabili, tra le cui modalità vi sono relazioni di eguale-diverso. La variabile sesso può essere suddivisa nelle categorie "maschio" e "femmina", che si possono denominare rispettivamente categoria 1 e categoria 2, oppure categoria A e categoria B.

La suddivisione fatta da Esquirol tra demenza e idiozia e la classificazione tra ritardati mentali e idioti di Seguin sono i primi esempi di scale nominali in psicologia. Secondo Esquirol la classe demenza differisce dalla classe idiozia in quanto nel primo caso si ha una regressione dello sviluppo psichico, mentre nella seconda si ha un arresto e un non sviluppo. Risulta evidente che un demente non può essere classificato idiota, né due soggetti che presentino un non sviluppo possono essere posti in classi diverse.

Tutte le classificazioni nosologiche sono esempi di scale nominali. Nello stabilire delle scale nominali è importante determinare con precisione i criteri per cui un evento o un oggetto viene incluso in una classe piuttosto che in un'altra, cioè l'unicità della categoria.

1.5.2 Scale ordinali

E' possibile ordinare secondo una scala ordinale una variabile qualora tra le sue modalità o caratteristiche esistano relazioni di minore-eguale-maggiore.

Quando Esquirol suddivise l'idiozia in innumerevoli classi in base alle sfumature di degradazione umana, classificò un ordinamento continuo, che andava dall'ultimo limite della degradazione umana fino ai limiti della normalità, fornendo un esempio di scala ordinale.

La maggior parte delle variabili utilizzate nelle scienze del comportamento sono, al massimo, a livello di scala ordinale: così le attitudini, gli interessi, gli atteggiamenti, il Q.I., i tratti della personalità ecc.

Binet precisò in questo senso il concetto di misura proprio della psicologia; concetto che non ha un senso matematico, quante volte cioè una quantità ne contiene un'altra, ma indica una posizione in graduatoria. Binet (1909) afferma che l'idea di misura si riconduce a quella di classificazione gerarchica. Così egli classificò in modo gerarchico tutte le categorie dell'intelligenza: brillante, superiore, media, lenta, debole mentale, imbecillità, idiozia.

Da un punto di vista matematico, mentre le scale nominali riguardano la caratteristica di unicità che possiedono i numeri, le scale ordinali si avvolgono delle relazioni ordinali proprie dei numeri. Così il numero 1 è inferiore a 2, che a sua volta è inferiore a 3 e così via. La relazione ordinale si riferisce pertanto unicamente alla posizione in graduatoria.

Perché sia possibile costruire una scala ordinale per misurare una variabile è necessario che questa abbia caratteristiche unitarie e che i fenomeni siano ordinati dal ricercatore in relazione alla variabile stessa. Esistono molte procedure per la costruzione di scale ordinali in psicometria; accenneremo alle graduatorie costruite per somma e alle scale di Likert.

Per ottenere una *graduatoria mediante somma* è sufficiente attribuire punti diversi a risposte, scelte o comportamenti diversi ed infine calcolare il punteggio complessivo.

Supponiamo che il tratto psicologico definito "fiducia in sé" sia valutabile dal fatto che non si abbia difficoltà a parlare con persone importanti, che non si abbia timore di essere notato da tanta gente, che non si abbia paura a sostenere esami, che non si abbia la sensazione di essere un fallito, che non si sia pessimisti ecc. E' possibile attribuire un punteggio alle varie situazioni nel modo seguente.

		zero punti	1 punto
1.	Mi trovo in difficoltà a parlare con persone importanti	si	No
2.	Ho timore di essere notato da tanta gente	si	No
3.	Quando devo dare esami ho tanta paura	si	No
4.	Mi sembra di essere un fallito	si	No
5.	Sono pessimista	si	No

Dal momento che alla risposta "no" corrisponde un punto e alla risposta "si" zero punti, il punteggio complessivo di un soggetto consiste nel sommare i punti ottenuti. Il soggetto che rispondesse sempre "no" otterrebbe 5 punti, quello che rispondesse sempre "si" zero punti.

Mediante questa scala si possono ordinare i soggetti in relazione alla variabile "fiducia in sé". Evidentemente il soggetto che ottiene 4 punti non ha 4 volte più fiducia in sé rispetto al soggetto che ha ottenuto 1 punto. Una simile illazione non avrebbe senso. Questa è una scala ordinale ed esprime semplicemente un ordinamento gerarchico e nulla più.

Per sostenere la validità della scala costruita è necessario che i 5 items utilizzati siano stati sottoposti in precedenza a tecniche di analisi, che indichino che i 5 items misurano un medesimo tratto della personalità. Solo in questo modo infatti è possibile considerare, ad esempio, identico il punteggio di 2 conseguito da tre soggetti, dei quali il primo abbia risposto affermativamente all'item 1 e 5, il secondo al 3 e 4, e il terzo al 2 e 5 item.

Le scale tipo Likert rappresentano una tecnica più complessa per costruire un ordinamento mediante somma. Questa tecnica prevede 5 possibilità di risposta:

- sono perfettamente d'accordo;
- sono d'accordo;
- non so decidermi;
- non sono d'accordo;
- non sono affatto d'accordo.

Ad ogni affermazione viene attribuito un punteggio: 4 a sono perfettamente d'accordo, 3 a sono d'accordo, 2 a non so decidermi, 1 a non sono d'accordo e zero a non sono affatto d'accordo.

Altre possibilità di risposta potrebbero essere: moltissimo, molto, abbastanza, poco e niente.

I 5 items relativi al tratto "fiducia in sé" potrebbero contemplare 5 possibilità di risposta (moltissimo, molto, abbastanza, poco, niente) o per praticità 3 possibilità:

	0	1	2
1. Mi trovo in difficoltà a parlare con persone importanti	si	a volte	no
2. Ho timore di essere notato da tanta gente	si	a volte	no
3. Quando devo dare esami ho tanta paura	si	a volte	no
4. Mi sembra di essere un fallito	si	a volte	no
5. Sono pessimista	si	a volte	no

Questo tipo di scala permette una maggiore discriminazione tra i soggetti e si rivela perciò più efficace rispetto alla precedente.

1.5.3 Scale d'intervalli

Con questo tipo di scala non solo è possibile ottenere una classificazione gerarchica delle modalità della variabile in questione, ma è anche possibile stabilire l'esatta distanza tra le diverse modalità. Questa proprietà implica che le scale d'intervalli possiedano un'unità costante, cioè un'unità di misura precisa e replicabile. Un esempio di scala d'intervalli è dato dalla scala centigrada, che ha come unità costante il grado centigrado.

Mentre le scale nominali possiedono la relazione di equivalenza e le scale ordinali le proprietà dell'ordinamento semplice, le scale d'intervalli possiedono le proprietà di addizione e di sottrazione. Posso così affermare che la differenza tra 20°C e 18°C è identica a quella fra 3°C e 1°C .

Mediante il metodo di Thurstone (1929) degli intervalli apparentemente uguali è possibile ottenere scale che abbiano le proprietà delle scale d'intervalli.

Il metodo consiste essenzialmente nel far classificare a dei giudici scelti delle affermazioni, concernenti la variabile che si vuoi misurare, in determinati mucchi scaglionati lungo un continuum che va dalle affermazioni più favorevoli a quelle più contrarie.

1.5.4 Scale di rapporti

Queste scale oltre a possedere tutte le proprietà delle scale nominali, ordinali e di intervalli, fanno riferimento ad uno zero assoluto. In questo tipo di scale sono possibili le operazioni di moltiplicazione e divisione. Le misure di peso, di tempo, di lunghezza sono esempi di scale di rapporti. Un soggetto di 80 chili pesa il doppio di uno di 40 chili.

La legittimità dell'uso di un determinato livello di misurazione dipende dalla natura della variabile che si vuole esaminare e dagli strumenti che si hanno a disposizione. In psicometria le variabili comunemente prese in esame appartengono al livello di misurazione delle scale ordinali. Tuttavia una volta che le modalità di queste variabili vengono quantificate e tradotte in numeri, si eseguono le operazioni proprie delle scale d'intervalli ottenendo degli indici (media, deviazione standard, indici di correlazione ecc.) che appartengono chiaramente a scale d'intervalli.

Il problema non è tanto sulla legittimità dell'applicazione di procedimenti statistici di un determinato livello a numeri che sono in relazione con fenomeni appartenenti ad un livello inferiore, quanto sulla correttezza dell'attribuzione di valori numerici ad oggetti ed eventi e soprattutto sul tipo di applicazione degli indici numerici ottenuti agli oggetti od eventi presi in esame.

1.6 Esercitazioni

2. Costruire esempi di scale nominali, ordinali e d'intervallo

1.7 Terminologia statistica

Approfondire con esemplificazione i seguenti termini :

1. Dati
2. Osservazioni
3. Esperimenti
4. Variabili
5. Misurazione
6. Quantificazione
7. Variabile quantitativa
8. Variabile qualitativa
9. Livelli di misurazione
10. Scala di misurazione
11. Scala nominale
12. Scala ordinale
13. Scala di intervalli
14. Scala di rapporti
15. Intervalli apparentemente uguali
16. Valutazione per intuito
17. Scheda di rilevamento
18. Questionario
19. Caratteri discreti
20. Variabile discreta
21. Caratteri continui
22. Graduatoria mediante somma
23. Scale tipo Likert
24. Zero assoluto

2. La raccolta di dati da un questionario

Nelle scienze dell'educazione i dati possono essere ricavati mediante l'osservazione e la registrazione del comportamento spontaneo o indotto, attraverso l'utilizzo di test o di risposte fornite dai soggetti a determinati compiti.

Nella nostra analisi ci serviremo dei dati ottenuti da un campione di studenti universitari frequentanti il corso di modelli statistici per l'analisi e la valutazione dei processi educativi.

Lo strumento adoperato per la raccolta dei dati è *un questionario sperimentale di personalità* che mira a rilevare *gli stili cognitivi* di soggetti che svolgono attività didattica o educativa.

Il test utilizzato è il **Questionario Stili Cognitivi** che è composto da 30 items o domande ed è suddiviso in tre parti: la prima parte, che comprende i primi dieci items, mira a valutare alcuni *aspetti del sé* riguardanti il senso di sicurezza psicologica del soggetto; la seconda parte, altri dieci items, intende valutare il grado di autonomia del soggetto *nel campo dell'apprendimento*; la terza parte, ultimi dieci items, mira a valutare il grado di indipendenza e autonomia *nell'attività di insegnamento* con eventuali alunni; il punteggio totale, formato dai 30 items, fornisce una stima del grado *di autonomia e indipendenza del proprio stile cognitivo*.

Per semplicità indicheremo la prima parte con: **I Aspetti del Sé**, la seconda parte con: **II Apprendimento**, la terza parte con: **III Insegnamento**, infine il punteggio complessivo con: **IV Totale**.

Per *stile cognitivo* si intende la modalità personale di indipendenza o dipendenza *dal campo*, cioè dall'ambiente fisico o sociale circostante. Lo stile cognitivo **di indipendenza** dal campo coincide con una marcata autonomia del soggetto, che nelle proprie scelte ed atteggiamenti fa riferimento a se stesso e alle proprie variabili interne, mentre il soggetto **dipendente** dal campo tende a privilegiare e seguire le indicazioni provenienti dall'ambiente fisico e sociale in cui vive, manifestando dipendenza e bisogno di approvazione sociale.

Trattandosi di un questionario sperimentale è necessario procedere preventivamente all'analisi degli items, alla determinazione delle norme statistiche, alla verifica degli indici di attendibilità e validità per poterlo utilizzare come strumento di valutazione psicologica ed educativa dei soggetti.

Il nostro obiettivo è semplicemente quello di utilizzarlo come strumento per la raccolta di dati e l'esemplificazione di modelli e procedure per l'analisi e la valutazione statistica nel campo educativo.

Il questionario è stato costruito per poter disporre di uno strumento utile nell'orientamento educativo e didattico dei docenti; la formulazione degli items ha tenuto conto di un adeguato quadro di riferimento teorico e di relativi indicatori; sono state standardizzate le condizioni di somministrazione e di determinazione del punteggio. Il questionario è stato somministrato collettivamente e si è mantenuto l'anonimato. Terminata la somministrazione si è proceduto all'autocorrezione indicando quale delle quattro alternative ai singoli items valesse 3 punti, oppure 2 punti, oppure 1 punto, oppure zero; la somma dei primi dieci punti grezzi ha fornito il punteggio grezzo di **I Aspetti del Sé**, la somma dei seguenti dieci punteggi ha fornito il punteggio grezzo di **II Apprendimento**, la somma degli ultimi dieci punteggi ha fornito il punteggio grezzo di **III Insegnamento**, infine la

somma di tutti gli items ha fornito il punteggio grezzo **IV Totale**.

Una volta raccolti, i dati della ricerca vengono organizzati mediante la costruzione di tabelle in cui inserirli.

Nel nostro caso è stata approntata la seguente tabella:

ITEM RELATIVI AL QUESTIONARIO
“STILI COGNITIVI”

	Aspetti del Sé	Apprendimento	Insegnamento	Totale
SOGGETTI	I	II	III	TOT

Nella tabella n°1 vengono riportati i dati grezzi delle risposte del campione di 112 soggetti al Questionario *Stili Cognitivi*

Tabella n°1 Punteggi grezzi Questionario “stili cognitivi”				
Soggetti	I	II	III	Totale
1	17	18	26	61
2	14	14	15	43
3	14	20	17	51
4	16	16	19	51
5	21	21	21	63
6	14	18	17	53
7	17	16	10	50
8	19	16	21	45
9	14	22	20	57
10	13	15	18	48
11	12	12	18	42
12	19	21	18	58
13	15	10	19	44
14	11	18	15	44
15	15	20	20	55
16	11	15	20	46
17	18	17	20	55
18	17	7	15	39
19	19	13	18	50
20	14	15	20	49
21	16	17	20	53
22	15	15	21	51
23	14	15	18	47
24	17	16	18	51
25	15	14	18	47
26	21	15	18	54
27	15	15	21	51
28	17	14	14	45
29	7	17	15	39
30	14	11	16	41
31	10	13	15	38
32	14	11	16	41
33	11	13	16	40

CAMPIONE N=112

Colonna 2: “Aspetti del sé”
Colonna 3: “Apprendimento”
Colonna 4: “Insegnamento”
Colonna 5: “Totale”

34	14	10	19	43
35	18	17	20	55
36	20	16	20	56
37	14	22	23	59
38	13	19	20	52
39	14	20	19	53
40	16	24	17	57
41	15	16	24	55
42	22	23	22	67
43	16	17	21	54
44	19	27	23	49
45	13	16	18	47
46	12	17	19	48
47	14	17	25	56
48	14	20	18	52
49	16	22	23	61
50	14	20	22	56
51	18	20	22	60
52	14	14	14	42
53	15	14	18	47
54	17	19	17	53
55	12	20	26	58
56	11	17	22	50
57	10	15	21	46
58	16	14	18	48
59	16	18	21	55
60	15	13	21	49
61	9	11	15	35
62	14	13	16	43
63	10	12	24	46
64	15	20	20	55
65	9	12	12	35
66	15	19	21	55
67	19	19	20	58
68	11	19	21	51
69	19	21	13	53
70	16	15	16	47
71	14	14	16	44
72	14	16	22	52
73	11	12	16	39
74	16	17	18	51
75	15	21	16	52
76	13	18	19	50
77	21	19	18	58
78	14	14	15	53
79	13	13	18	44
80	14	11	16	41
81	14	11	16	41
82	19	17	17	43
83	18	15	19	42
84	21	15	16	52
85	15	19	18	52
86	16	14	18	48
87	16	10	19	45
88	20	18	12	50
89	20	15	17	52
90	19	20	18	57
91	16	17	20	53

92	17	18	23	58
93	17	18	22	57
94	17	21	18	56
95	17	13	20	50
96	17	19	25	61
97	10	11	18	39
98	19	19	19	57
99	15	18	19	52
100	11	13	14	38
101	15	22	17	54
102	15	21	17	53
103	17	21	16	54
104	17	15	16	48
105	17	16	19	52
106	12	17	20	49
107	14	19	16	49
108	22	19	16	57
109	19	16	19	54
110	17	7	19	43
111	17	17	14	48
112	15	17	22	55

2.1 Esercitazioni

1. Con il foglio di calcolo elettronico (Excel) costruire una tabella simile a tab.1
2. Costruire *questionari di vario genere* per la raccolta di dati sperimentali.

2.2 Terminologia statistica

Approfondire con esemplificazione i seguenti termini :

1. Tabella
2. Questionario di personalità
3. Stile cognitivo
4. Personalità
5. Aspetti del sé
6. Stile di apprendimento
7. Stile di insegnamento
8. Dipendenza
9. Autonomia
10. Insicurezza
11. Test
12. Subtest
13. Soggetto
14. Punteggio grezzo
15. Punteggio totale
16. Item

3. L'organizzazione dei dati in tabelle

3.1 I dati del Questionario “*Stili cognitivi*”

I Punteggi conseguiti dai soggetti al Questionario costituiscono i dati grezzi. Per poter interpretare i punteggi dei singoli soggetti è necessario esaminare i risultati di ogni soggetto e fare il confronto con i risultati degli altri soggetti.

Pertanto è necessario eseguire un'analisi dettagliata e minuziosa di tutti i punteggi ottenuti.

Il primo modo di analisi dei dati consiste nel disporre i dati medesimi in tabelle che rendano facile e spedita l'individuazione dei punteggi dei diversi soggetti e consentano il confronto tra i vari punteggi.

La tabella N° 1 , o tabella di spoglio dei risultati, è molto importante e consente d'individuare celermente il punteggio di ogni soggetto e consente un primo confronto tra i risultati di tutti i soggetti.

Per un'analisi più approfondita è necessario approntare nuove tabelle che consentano sintesi maggiori dei vari dati.

Osservando i punteggi grezzi relativi alla variabile **I Aspetti del Sé** si nota che i punteggi vanno da 7, punteggio minimo ottenuto, a 22, punteggio massimo ottenuto. Pertanto è possibile calcolare quanti soggetti hanno ottenuto i vari punteggi grezzi da 7 fino a 22. Indicando con il termine **frequenza assolute** il numero di soggetti che hanno ottenuto un determinato punteggio grezzo è utile costruire la tabella delle frequenze assolute relativamente ai quattro aspetti del Questionario: I Aspetti del Sé, II Apprendimento, III Insegnamento e IV Totale.

Per costruire una **tabella di frequenze** è necessario ordinare i dati dal più piccolo al più grande e quindi calcolare le frequenze dei vari punteggi; i punteggi possono anche essere ordinati dal più grande al più piccolo; per costruire una tabella di frequenze sono sufficienti due colonne: nella prima vanno ordinati i dati, nella seconda vanno indicate le frequenze assolute, cioè il numero di volte che si verifica la presenza di quel punteggio.

Viene riportata la tabella di frequenze per: I Aspetti del Sé (tab.N.2); per: II Apprendimento (tab.N.3); per : III Insegnamento (tab.N.4); per: IV Totale (tab.N. 5) .

Tab. N.2 – Frequenze assolute

Campione N=112 I : “Aspetti del sé”	
PUNTEGGI	FREQUENZE
7	1
9	2
10	4
11	7
12	4
13	5
14	22
15	16
16	12
17	16
18	4
19	10
20	3
21	4
22	2

Tab. N.3 – Frequenze assolute

Campione N=112 II: “Apprendimento”	
PUNTEGGI	FREQUENZE
7	2
10	3
11	6
12	4
13	8
14	9
15	13
16	10
17	14
18	9
19	11
20	9
21	7
22	4
23	1
24	1
27	1

Tab. N.4 – Frequenze assolute

Campione N=112 III : “Insegnamento”	
PUNTEGGI	FREQUENZE
10	1
12	2
13	1
14	4
15	7
16	15
17	8
18	19
19	13
20	14
21	11
22	7
23	4
24	2
25	2
26	2

Tab. N.5 – Frequenze assolute

Campione N=112 IV : Totale			
PUNTEGGI	FREQUENZE	PUNTEGGI	FREQUENZE
33	1	51	7
35	1	52	10
38	2	53	8
39	4	54	6
40	1	55	7
41	4	56	4
42	2	57	6
43	5	58	5
44	4	59	1
45	3	60	1
46	3	62	3
47	5	63	1
48	6	67	1
49	4	69	1
50	6		

3.2 L'organizzazione dei dati

In generale i dati possono riferirsi ad una *frequenza*, qualora indichino il numero di diverse modalità del fenomeno. Ad esempio, se in un'urna ci sono 10 palline nere, 15 bianche e 20 rosse, i numeri 10, 15 e 20 indicano il numero dei casi e pertanto le rispettive frequenze assolute.

Si parla invece di *intensità* quando si rileva la misurazione di un fenomeno quantitativo; dicendo che un muro è alto metri 3,28, il dato esprime il rapporto tra l'altezza del muro e l'unità di misura rappresentata dal metro.

Si ha una *distribuzione di frequenza o di intensità* quando i dati vengono organizzati in *tabelle*, dove vengono indicate le diverse modalità ed il totale; la tabella 5.1 e la tabella 5.2 sono degli esempi.

Tab.5.1- Distribuzione di frequenza . Palline contenute nell'urna.

<i>Colore delle palline</i>	<i>Frequenze assolute (f)</i>
Nere	10
Bianche	15
Rosse	20
totale	45

Tab.5.2- Distribuzione di intensità. Peso di un contenitore

<i>Oggetti contenuti</i>	<i>Peso in chili</i>
<i>Forbici</i>	0,3
<i>Martello</i>	1
<i>Mazza</i>	4
totale	5,3

Mentre con *frequenza assoluta* o semplicemente *frequenza* ci si riferisce al numero delle osservazioni effettuate di un determinato fenomeno, con *frequenza relativa* si indica il rapporto fra la frequenza assoluta e il numero complessivo dei casi:

$$(1) \quad f_{rel} = \frac{f}{N}$$

dove:

f_{rel} = frequenza relativa

f = frequenza assoluta

N = numero dei casi

Il totale della somma delle frequenze relative è pari ad 1. Il totale si indica con il simbolo Σ (corrispondente alla lettera dell'alfabeto greco S maiuscola) e si legge *sommatoria*.

$$(2) \quad \Sigma f_{rel} = 1$$

Volendo calcolare le frequenze relative della tab.1 si avrà:

Tab.5.3- Distribuzione di frequenze assolute e relative

<i>Colore palline</i>	<i>f (assolute)</i>	<i>f_{rel}</i>
Nere	10	10:45 = 0,22
Bianche	15	15:45 = 0,33
Rosse	20	20:45 = 0,44
<i>Totale Σ</i>	<u>45</u>	<u>1</u>

Le frequenze cumulate si ottengono ordinando in ordine crescente la distribuzione di frequenza e sommando tutte le frequenze fino ad includere quella considerata (la frequenza *iesima* indicata con x_i):

$$(3) \quad f_{cum_i} = \sum_{j=1}^i f_j$$

dove:

f_{cum_i} = frequenza cumulata *iesima*, cioè la frequenza cumulata considerata

$\sum_{j=1}^i$ = sommatoria di j che va da 1 ad i , cioè la somma della frequenza considerata con le precedenti

Riferendoci ai dati della tab.5.1 si otterrà la tab.5.4.

Tab.5.4- Frequenze assolute e frequenze cumulate

<i>Colore palline</i>	<i>f</i>	<i>f_{cum}</i>
Nere	10	10
Bianche	15	25 (=10+15)
Rosse	20	45 (=25+20)
<i>Totale</i>	45	

L'uso di *tabelle* facilita molto l'organizzazione e la rappresentazione dei dati. Si va dalle tabelle semplici, come la distribuzione di frequenze (tab.5.1), a quelle più complesse come le tabelle a doppia entrata, le tabelle che suddividono in classi le osservazioni, ecc.

Con serie e seriazioni statistiche si intendono le grandezze ordinate riportate su una tabella; le serie temporali riguardano i caratteri rilevati in relazione a tempi diversi; le seriazioni sono successioni di coppie di numeri dei quali quello di destra esprime sempre la frequenza (tab.5.5).

Tab.5.5- Seriazioni. Classificazione di 20 alunni ad una prova di matematica

<i>Classi punteggi</i>	<i>frequenze</i>
1-3	4
4-6	10
7-9	6
<i>Totale</i>	20

3.3 Esercitazioni

1. **Con il foglio elettronico (Excel)** costruire tabelle simili alle tabelle 2- 3- 4- 5.
2. Riprodurre **al foglio elettronico (Excel)** le tabelle da 5.1 a 5.5.
3. Immettere **nel foglio elettronico** i dati ricavati dalla somministrazione di questionari e da tabelle di rilevamento appositamente costruite.

3.4 Terminologia statistica

Approfondire con esemplificazione i seguenti termini :

1. Frequenza; frequenza assoluta; frequenza relativa; frequenza cumulata
2. Distribuzione di frequenza; distribuzione d'intensità.

Indicare il significato dei seguenti simboli

1. $f_{rel} = \frac{f}{N}$
2. $\sum f_{rel} = 1$
3. $f_{cum_i} = \sum_{j=1}^i f_j$

4. Calcolo della Media Aritmetica

La tabella di frequenze consente ulteriori ed importanti analisi dei punteggi ottenuti dai soggetti. Bisogna a questo punto presentare alcuni simboli che vengono usati nel calcolo statistico.

I punteggi grezzi ottenuti dai soggetti vengono indicati con la lettera maiuscola **X**.

Così il punteggio 7 di Aspetti del Sé si indica con X_1 , il punteggio 9 con X_2 , il punteggio 10 con X_3 , e così via; con X_i si indica un qualsiasi punteggio da 7 a 22; le X_i sono i vari punteggi.

Le frequenze assolute si indicano con la lettera minuscolo **f**; anche in questo caso f_i indica una qualsiasi frequenza connessa ai diversi punteggi.

Con **Media Aritmetica** si intende quel punteggio ottenuto dalla *somma di tutti i punteggi grezzi diviso il numero dei punteggi*. La media aritmetica si indica con **M.A.** oppure con una $\bar{X}$ maiuscolo soprasssegnata da un trattino:

$$(4) \quad \text{Media Aritmetica} = \text{M. A.} = \bar{X}$$

Con **Sommatoria** si intende la somma dei numeri indicati; la sommatoria si indica con la lettera dell'alfabeto greco Σ , che corrisponde alla esse maiuscola.

La somma può riguardare alcuni dati oppure tutti i dati; per sapere quali numeri vanno sommati si indicano sotto e sopra la lettera Σ alcuni indici. Così volendo sommare tutti i dati si scriverà:

$$\sum_{i=1}^N X_i$$

che si legge: sommatoria delle X_i con i che va da 1 ad N ; cioè sommatoria di tutte le X , ossia sommatoria di tutti i punteggi. Volendo semplificare si può scrivere:

$$\sum X_i \quad \text{o anche} \quad \sum X$$

La sommatoria va calcolata al termine delle colonne delle tabelle.

Volendo sintetizzare con una formula il calcolo della Media Aritmetica si avrà:

$$(5) \quad \text{Media Aritmetica} = \frac{\text{somma di tutti i punteggi}}{\text{numero dei punteggi}} = \frac{\sum_{i=1}^N f_i X_i}{N} = \frac{\sum fX}{N}$$

dove:

Σ = simbolo della sommatoria

f_i = simbolo delle frequenze

X_i = i punteggi grezzi

N = numero dei dati o del campione

Si ricordi che il numero dei punteggi **N** è uguale alla sommatoria di tutte le frequenze assolute:

$$(6) \quad \sum f_i = N$$

L'uso di tabelle semplifica il calcolo della Media Aritmetica. Per il calcolo della Media Aritmetica è sufficiente approntare una tabella come indicato in tab. N.6

Tab. N.6

Tabella per il calcolo della Media Aritmetica (M.A. = $\bar{X}$)

1	2	3
PUNTEGGI GREZZI ORDINATI	FREQUENZE	PUNTEGGI per FREQUENZE
X_i	f_i	$f_i X_i$
	$\sum f_i = N$	$\sum f_i X_i$

$$\text{Col. } 3 = 1 \times 2$$

$$\bar{X} = 3/2$$

La tab.1 contiene 4 tipi di punteggi o variabili, pertanto dovremo calcolare 4 diverse Medie Aritmetiche; a tal fine prepqriamo 4 tabelle: la tab.7 per il calcolo della M.A. della variabile I Aspetti del Sé; la tab.8 per il calcolo della M.A. per la variabile II Apprendimento; la tab. 9 per il calcolo della M.A. per la vabiabile Insegnamento; la tab. 10 per il calcolo della vaiabile IV Totale.

Tab.7 – Calcolo della Media Aritmetica (M.A. = $\bar{X}$)

Questionario “Stili cognitivi” I “Aspetti del sé” Campione N=112		
X_i	Frequenze f_i	Punteggi grezzi per frequenze $X_i * f_i$
7	1	7
9	2	18
10	4	40
11	7	77
12	4	48
13	5	65
14	22	308
15	16	240
16	12	192
17	16	272
18	4	72
19	10	190
20	3	60
21	4	84
22	2	44
	$\sum f_i = N = 112$	$\sum f_i X_i = 1717$

$$\text{Media Aritmetica} = M.A. = \bar{X} = \frac{\sum f_i X_i}{N} = \frac{1717}{112} = 15.33$$

Tab.8 – Calcolo della Media Aritmetica (M.A.)

Questionario “Stili cognitivi” II “Apprendimento” Campione N=112		
Punteggi grezzi X_i	Frequenze f_i	Punteggi grezzi per frequenze $X_i \times f_i$
7	2	14
10	3	30
11	6	66
12	4	48
13	8	104
14	9	126
15	13	195
16	10	160
17	14	238
18	9	162
19	11	209
20	9	180
21	7	147
22	4	88
23	1	23
24	1	24
27	1	27
	$\sum f = N =$ 112	$\sum f_i X_i =$ 1841

$$\text{Media Aritmetica} = \text{M.A.} = \bar{X} = \frac{\sum f_i X_i}{N} = \frac{1841}{112} = 16.44$$

Tab. 9 – Calcolo della Media Aritmetica (M.A.)

Questionario “Stili cognitivi” III “Insegnamento” Campione N=112		
Punteggi grezzi X_i	Frequenze f_i	Punteggi grezzi per frequenze $X_i \times f_i$
10	1	10
11	0	0
12	2	24
13	1	13
14	4	56
15	7	105
16	15	240
17	8	136
18	19	342
19	13	247
20	14	280
21	11	231
22	7	154
23	4	92
24	2	48
25	2	50
26	2	52
	$\sum f_i = N =$ 112	$\sum f_i X_i =$ 2080

$$\text{Media Aritmetica} = \text{M.A.} = \bar{X} = \frac{\sum f_i X_i}{N} = \frac{2080}{112} = \mathbf{18.57}$$

Tab.10 – Calcolo della Media Aritmetica (M.A.)

Questionario “Stili cognitivi” IV Totale Campione N=112		
Punteggi grezzi X_i	Frequenze f_i	Punteggi grezzi per frequenze X_i * f_i
33	1	33
35	1	35
38	2	76
39	4	156
40	1	40
41	4	164
42	2	84
43	5	215
44	4	176
45	3	135
46	3	138
47	5	235
48	6	288
49	4	196
50	6	300
51	7	357
52	10	520
53	8	424
54	6	324
55	7	385
56	4	224
57	6	399
58	5	290
59	1	59
60	1	60
61	3	183
62	0	0
63	1	63
67	1	67
69	1	69
	$\sum f_i = N =$ 112	$\sum f_i X_i =$ 5695

$$\text{Media Aritmetica} = \text{M.A.} = \bar{X} = \frac{\sum f_i X_i}{N} = \mathbf{50.84}$$

4.1 Esercitazione – Calcolo della Media Aritmetica

Ad un campione di 40 soggetti è stato somministrato il **Questionario “Stili Cognitivi”**. I punteggi grezzi ottenuti da questo campione sono riportati nella **tab. 11**.

Si chiede di organizzare i dati in tabelle di frequenze e di calcolare la Media Aritmetica delle quattro variabili del Questionario per il campione di 40 soggetti disponendo i dati in tabelle.

Tabella 11

Punteggi grezzi Questionario “*Stili Cognitivi*” -
Campione N=40

Soggetti	I	II	III	Totale
1	11	15	17	43
2	13	19	18	50
3	11	20	16	47
4	18	21	19	58
5	13	11	16	40
6	19	22	19	60
7	17	19	17	53
8	16	12	22	50
9	10	16	25	51
10	11	16	13	40
11	17	24	20	61
12	11	14	14	39
13	14	13	23	50
14	10	21	20	51
15	16	16	12	44
16	18	19	15	52
17	20	18	18	56
18	12	15	16	43
19	21	25	19	65
20	20	17	12	49
21	14	15	17	46
22	14	22	19	55
23	17	24	22	63
24	11	19	25	55
25	14	15	24	53
26	12	9	18	39
27	16	14	17	47
28	14	18	26	58
29	12	17	23	52
30	17	7	14	38
31	15	15	18	48
32	17	15	15	47
33	16	24	22	62
34	16	19	21	56
35	16	19	19	54
36	18	17	16	51
37	15	17	15	47
38	15	22	19	56
39	16	17	15	48
40	20	16	16	52

CAMPIONE N=40

Colonna 2: “Aspetti del sé”
Colonna 3: “Apprendimento”
Colonna 4: “Insegnamento”
Colonna 5: “Totale”

Verificare se i risultati ottenuti corrispondono a quelli riportati nella tab.12

Tab.12 – Risultati relativi alle medie aritmetiche del campione di N=40

<i>Questionario “Stili cognitivi”</i>	$\sum f_i$	$\sum f_i X_i$	$M.A. = \frac{\sum f_i X_i}{N}$	
I Aspetti del Sé	40	603	15.07	
II Apprendimento	40	694	17.87	
III Insegnamento	40	732	13.3	
IV Totale	40	2029	50.72	

4.2 L'indice statistico della Media Aritmetica

Siamo partiti dai *punteggi grezzi* ottenuti dai soggetti al *Questionario “Stili Cognitivi”* e siamo giunti attraverso semplici calcoli algebrici , cioè mediante l'applicazione delle quattro operazioni, al calcolo della Media Aritmetica delle 4 dimensioni o variabili del *Questionario*.

L'insieme dei punteggi grezzi del campione costituiscono **la distribuzione dei punteggi**. La tab. 1 riporta le quattro distribuzioni dei punteggi del campione di 112 soggetti (N=112).

Osservando ogni distribuzione dei punteggi si nota che *i dati vanno da un minimo ad un massimo*; nel caso di *I Aspetti del Sé* i punteggi vanno da un minimo di 7 ad un massimo di 22; nel caso di *II Apprendimento* da un minimo di 7 ad un massimo di 27; nel caso di *III Insegnamento* da un minimo di 10 ad un massimo di 26; nel caso di *IV Totale* da un minimo di 33 ad un massimo di 69.

Questa variazione di una distribuzione di punteggi viene chiamata **gamma dei punteggi** e si indica con la lettera greca Γ , che sarà uguale alla *differenza tra il punteggio massimo ed il punteggio minimo di una distribuzione* ed indica **il campo di variazione dei punteggi**.

Esprimendoci con una formula sarà:

$$(7) \quad \text{gamma} = \Gamma = (\text{puntegg. massimo} - \text{puntegg. minimo}) + 1$$

La prima osservazione è che i punteggi dei soggetti *variano entro una certa gamma*.

Ora esiste *un punteggio che si colloca al centro della gamma* o variazione della distribuzione.

La Media Aritmetica è il punteggio che si colloca al centro della gamma.

Volendo dare una definizione più precisa osserviamo che la *somma di tutti i punteggi della gamma* fornisce un determinato risultato; così la somma dei 112 punteggi di *I Aspetti del Sé* è uguale a

1717; sostituendo ad ognuno dei 112 punteggi di *I Aspetti del Sé* la media aritmetica (15,33035714) *la loro somma* fornisce lo stesso risultato di 1717.

Pertanto possiamo definire **la media aritmetica quel valore che sostituito ai dati consente di ottenere lo stesso risultato** (definizione del Chisini).

Usando le formule avremo:

$$(7.1) \quad X_1 + X_2 + X_3 + X_4 = \text{M.A.} + \text{M.A.} + \text{M.A.} + \text{M.A.}$$

$$(7.2) \quad \sum X_i = N \times \text{M.A.}$$

$$(7.3) \quad \text{Media Aritmetica} = \text{M.A.} = \frac{\sum X_i}{N}$$

La *media aritmetica* è compresa fra il dato più piccolo e il dato più grande e pertanto è *un valore medio*.

Come vedremo in seguito esistono altri valori medi.

Questa media è detta *aritmetica* perché calcolata su un insieme di *dati crescenti* di numero dispari *in progressione aritmetica* corrisponde al *valore centrale* di detta progressione.

Così, avendo la progressione

$$6+8+10+12+14$$

$$\text{M.A.} = \bar{X} = \frac{\sum X_i}{N} = \frac{6+8+10+12+14}{5} = 10$$

Cioè la **media aritmetica** coincide con il **valore centrale 10**.

La *media aritmetica* è il *valore medio* più usato perché le grandezze sommative si incontrano facilmente.

Mentre la *media aritmetica di un campione* si indica con $\bar{X}$ la *media aritmetica di una popolazione* si indica con la lettera greca μ (mu).

La *media aritmetica* costituisce un *indice statistico*, in quanto rappresenta un valore che riassume l'intera distribuzione dei dati

Parleremo in seguito di campione e di popolazione e delle proprietà della media aritmetica

4.3 Un modello utile nel lavoro educativo

Il lavoro educativo ha il suo centro nevralgico *nella relazione d'aiuto* che si viene a stabilire tra l'educatore e *l'altro*, sia nella fase evolutiva e formativa che nelle situazioni di conflitto e di disagio personale, familiare e sociale.

La qualità della relazione d'aiuto contraddistingue le diverse modalità nelle quali si articola il lavoro educativo attualizzato nelle numerose *figure professionali* che hanno come punto di riferimento o come dimensione costitutiva e caratterizzante *l'educazione*.

Non è nostro obbiettivo trattare le *dinamiche delle relazioni educative*.

Faremo brevi cenni della relazione propria *dell'educatore*.

L'educatore nel suo lavoro stabilisce con il soggetto e con la comunità *una relazione d'aiuto relativamente stabile nel tempo* : una relazione *maieutica e genitoriale*, in quanto suscitatrice di nuovi atteggiamenti e comportamenti intrapersonali e intersoggettivi più realistici, creativi e amicali.

L'educatore non agisce in maniera automatica e predeterminata dai suoi bisogni razionali e irrazionali interni o dalle forze ambientali e sociali. Il suo profondo sentire, le sue reazioni, il suo comportamento è empatico, positivo, realistico, creativo, aperto al futuro, *in sintonia con il vissuto del soggetto e della comunità di cui fa parte*.

L'educatore orienta anche *la propria spontaneità* in maniera realistica e maieutica.

L'educatore deve interagire con la personalità del soggetto e con le dinamiche della comunità sviluppando la capacità di pronta, realistica valutazione dei comportamenti dei soggetti.

L'educatore *raccoglie e registra* i numerosi comportamenti del soggetto, li *confronta* con i precedenti, con quelli di altri soggetti e con i comportamenti attesi, infine *reagisce* con il proprio comportamento.

I *modelli statistici* sono di grande utilità nel lavoro educativo ed affinano la capacità d'intuito e di reazione oltre a consentire interventi d'analisi e di valutazione più realistici, motivanti ed efficaci.

Ci soffermiamo brevemente sulla necessità di un'attenta raccolta dei dati.

4.3.1 *La tendenza dominante nel comportamento*

Nella relazione d'aiuto l'educatore s'impegna *con autenticità* in un'interazione *sentita e creativa*, ma nello stesso tempo è orientato a promuovere *un realistico sviluppo ed un'ottimale crescita* del soggetto.

Atteggiamenti e comportamenti del soggetto sono costantemente rilevati, valutati, usati come feed back, integrati nella reazione che caratterizza la relazione d'aiuto dell'educatore.

La qualità e l'efficacia del lavoro educativo è legata a questa *capacità di rilevazione, valutazione ed integrazione* da parte dell'educatore.

Il primo problema è pertanto la rilevazione di atteggiamenti e comportamenti del soggetto.

I metodi che l'educatore può usare sono molteplici: *osservazione spontanea* nei vari ambienti di vita del soggetto, *indagine in situazioni mirate o sperimentali* (discussioni, gioco dei ruoli, compiti specifici...), *risposte ai test, resoconti e relazioni, valutazioni di altre persone*,...

La raccolta dei dati deve essere attenta, ampia, scevra da pregiudizi, rispondente alla realtà del soggetto, rispettosa della privacy, aperta al futuro.

Il modello statistico ci indica un estremo rigore nella raccolta dei dati , che devono essere veri e rappresentativi, cioè appartenenti al soggetto.

In riferimento al soggetto alcuni comportamenti o fatti hanno la caratteristica dell'unicità, come la nascita e la morte, mentre altri sono soggetti a ripetizione. In riferimento ad un campione il fatto più significativo è la diversa frequenza del fenomeno.

L'educatore deve prestare la massima attenzione alla ripetizione di determinati comportamenti e cercare di capirne l'obiettivo e la motivazione.

La salute mentale e l'equilibrio dipendono dalla frequenza o assenza di determinati comportamenti.

Il modello della tendenza centrale orienta l'educatore alla considerazione della personalità del soggetto tenuto conto dei diversi e contrastanti atteggiamenti rilevati.

L'educazione mira a consolidare nel soggetto comportamenti che evitino atteggiamenti estremistici:
in medio stat virtus.

4.4 Esercitazioni

1. *Al foglio elettronico(Excel)* ricostruire le tabelle 7 – 8 – 9 – 10. Dopo aver copiato le prime due colonne ottenere la terza colonna con il comando “aggiungi funzione”: le caselle della colonna C si ottengono moltiplicando le rispettive caselle di A e di B (tenendo attiva ad es. la casella C2 si scriverà $= (A2 * B2)$)

4.5 Terminologia statistica

Approfondire con esemplificazione i seguenti termini :

1. Questionario di personalità
2. Variabili
3. Punteggi grezzi
4. Distribuzione di dati
5. Tabella
6. Gamma dei punteggi
7. Punteggi ordinati
8. Frequenze assolute
9. Sommatoria
10. Media Aritmetica
11. Indice statistico
12. Progressione aritmetica
13. Definizione del Chisini

Indicare il significato dei seguenti simboli:

1. X_i
2. f_i
3. $f_i X_i$
4. Σ
5. $\overline{X}$
6. $\sum f_i$
7. $\sum f_i X_i$
8. N
9. $\frac{\sum f_i X_i}{N}$
10. M.A.
11. μ
12. $\sum fX$

5. Le misure della tendenza centrale

L'uso di tabelle e di grafici consente di rappresentare i dati e di intravedere l'andamento e le caratteristiche di una distribuzione.

Con *l'elaborazione matematica* i dati possono essere analizzati e valutati in relazione al problema che si intende affrontare.

I valori caratteristici di un insieme di dati sono espressi mediante funzioni matematiche, che mettono in luce determinati aspetti dei dati stessi.

Gli aspetti e le caratteristiche fondamentali di ogni distribuzione di dati sono: *il valore medio e la variabilità*.

5.1 Il valore medio

Come si è rilevato nelle pagine precedenti, una distribuzione di dati comprende la presenza di più dati, che possono essere uguali tra loro, tutti diversi, oppure alcuni ripetuti.

Il numero di volte che un dato si ripete si chiama *frequenza assoluta* (f_i oppure n_i); il dato che non si ripete avrà come frequenza: 1.

Ogni valore di una distribuzione di dati può essere considerato valore medio o centrale della distribuzione.

In una distribuzione di dati si possono calcolare:

1. *le medie algebriche*
2. *le medie di posizione*.

Per il calcolo delle *medie*, come si è visto, è opportuno disporre i dati in tabelle sotto forma di distribuzione di frequenze (vedi *tab. 2* e seguenti). In una distribuzione *la somma* di tutte le frequenze assolute è *uguale a N*.

5.2 Le medie algebriche

Si chiamano *medie algebriche* quelle calcolate mediante un'espressione algebrica.

La media algebrica esprime *l'equidistribuzione dei dati*, cioè il valore che ogni dato assumerebbe se la distribuzione fosse equilibrata.

Per calcolare *la media* è pertanto importante verificare *il tipo di legame* che unisce i dati. I dati di una distribuzione possono avere carattere additivo, moltiplicativo...

Quelle più comunemente usate sono la media aritmetica (M.A.), la media geometrica (M_g), la media armonica (M_a), la media quadratica (M_q), la media cubica (M_c), le medie potenziate, le medie antiarmoniche e le medie radicali

La media si ottiene dai dati mediante calcoli algebrici. Secondo il Chisini, la media di un insieme di dati è tale se collocata al posto dei singoli dati ottiene lo stesso valore della funzione dei dati.

Ad esempio, nella funzione di somma :

$$5+2+4+1 = 3+3+3+3$$

I dati (5,2,4,1) hanno come media 3 che collocata al posto dei dati ottiene la stessa funzione di somma (12).

5.2.1 La media aritmetica

Avendo una successione di numeri, corrispondenti ad una variabile la cui caratteristica è sommativa, è possibile calcolare la loro somma.

Così se

$$X_1 + X_2 + X_3 + X_4 = 2 + 6 + 7 + 10$$

$$\Sigma X_i = 25$$

Ora secondo la definizione del Chisini la media è quel valore che sostituito ai dati consente di ottenere lo stesso risultato; per cui

$$X_1 + X_2 + X_3 + X_4 = M + M + M + M$$

Indicando con N il numero dei dati si avrà:

$$\Sigma X_i = N * M$$

Da cui:

$$(7.4) \quad M = \frac{\Sigma X_i}{N}$$

Dove:

$M =$ media aritmetica

$\Sigma X_i =$ sommatoria di tutti i dati o di tutte le X

$N =$ numero dei dati

Pertanto la media aritmetica è uguale alla somma dei dati diviso per il numero dei dati. Essa corrisponde all'equadistribuzione di grandezze additive.

Per il calcolo della media aritmetica ci si può servire di una tabella di distribuzione di frequenza

aggiungendo una terza colonna come indicato nella(tab.12.1)

Tab.12.1- Calcolo della media aritmetica

1	2	3
Valori della variabile	frequenze assolute	valori della variabile moltiplicati per la rispettiva frequenza
X_i	f_i	$f_i X_i$
X_1		
X_2		
X_3		
$\vdots$		
X_n		
	$\Sigma f_i = N$	$\Sigma f_i X_i$

La formula per il calcolo della *media aritmetica* è la seguente:

$$(7.5) \quad M.A. = \bar{X} = \frac{\Sigma f_i X_i}{N}$$

Dove:

$M.A.$ = Media Aritmetica

$\bar{X}$ = simbolo che indica comunemente la Media Aritmetica

Σ = simbolo che indica la sommatoria

f_i = le frequenze assolute

X_i = i singoli dati

N = numero dei dati

Ecco un esempio di calcolo della media aritmetica (tab.12.2).

Tab.12.2-Calcolo della media aritmetica

1 Punteggi	2 frequenze assolute	3 (colonna 1 per colonna 3)
X_i	f_i	$f_i X_i$
2	1	2
3	1	3
4	2	8
5	6	30
7	3	21
8	2	16
	$\Sigma f_i = N =$ = 15	$\Sigma f_i X_i =$ = 80

$$M.A. = \bar{X} = \frac{\Sigma fX}{N} = \frac{\text{somma di tutti i punteggi}}{\text{numero dei punteggi}} = \frac{\text{colonna 3}}{\text{colonna 2}}$$

$$M.A. = \bar{X} = \frac{80}{15} = 5,33$$

La media aritmetica è compresa fra il dato più piccolo e il dato più grande e pertanto è un valore medio.

Questa media è detta aritmetica perché calcolata su un insieme di dati crescenti di numero dispari in progressione aritmetica corrisponde al valore centrale di detta progressione.

Si abbia la seguente progressione:

$$6+8+\mathbf{10}+12+14$$

$$M.A. = \bar{X} = \frac{\Sigma fX}{N} = \frac{6+8+10+12+14}{5} = 10$$

La media aritmetica è il valore medio più usato perché le grandezze sommative si incontrano facilmente.

Mentre la media aritmetica di un campione si indica con $\bar{X}$, la media aritmetica di una popolazione si indica con la lettera greca μ (si legge *mu*).

5.2.1.1 *Media totale di più gruppi parziali*

Conoscendo la numerosità e la media di due gruppi congiunti è possibile calcolare la media complessiva dell'insieme di tutti i dati applicando la seguente formula:

$$(7.5.1) \quad M_{1+2} = \frac{(N_1 \times M_1 + N_2 \times M_2)}{N_1 + N_2}$$

Dove:

M_{1+2} = media aritmetica combinata o complessiva

N_1 = numerosità del primo campione

M_1 = media aritmetica del primo campione

N_2 = numerosità del secondo campione

M_2 = media aritmetica del primo campione

5.2.2 *La media geometrica*

Se il tipo di legame che unisce i dati in esame non è additivo ma moltiplicativo l'equidistribuzione dei dati si ottiene mediante il calcolo della *media geometrica*.

Così ad es. se

$$X_1 \times X_2 \times X_3 = M \times M \times M$$

$$X_1 \times X_2 \times X_3 = M^3$$

Avremo:

$$M = \sqrt[3]{X_1 * X_2 * X_3}$$

Pertanto la *media geometrica* (M_g) di n dati positivi è uguale alla radice ennesima positiva del prodotto di tutti i dati:

$$(7.6) \quad M_g = \sqrt[n]{X_1 * X_2 \dots * X_n}$$

Per semplificare il calcolo della media geometrica è possibile trasformare i dati in logaritmi e dividere la loro somma per N ed infine trasformare il logaritmo così ottenuto in un numero.

Avremo pertanto dalla (7.6)

$$\log M_g = \frac{1}{n} (\log X_1 + \log X_2 \dots + \log X_n)$$

$$(7.8) \quad \log M_g = \frac{\sum \log X_i}{N}$$

Nella tab.12.3 viene riportato un esempio di calcolo della media geometrica.

Tab.12.3- Calcolo della media geometrica

1	2	3	4
X_i	f_i	$\log X_i$	$f_i \log X_i$
2	1	0,30	0,30
3	2	0,48	0,96
7	4	0,85	3,40
9	3	0,96	2,85
12	2	1,08	2,16
$\sum f_i = N = 12$			$\sum f_i \log X_i = 9,67$

$$M_g = \sqrt[12]{2 * 3^2 * 7^4 * 9^3 * 12^2}$$

$$\log M_g = \frac{1}{12} (\sum \log 2 + 2 \log 3 + 4 \log 7 + 3 \log 9 + 2 \log 12)$$

$$\log M_g = \frac{\sum f_i \log X_i}{N} = \frac{9,67}{12} = 0,81$$

$$M_g = \text{anti log } 0,81 = 6,39$$

Anche la media geometrica, calcolata in una progressione geometrica di dati dispari, coincide con il dato centrale della progressione stessa.

Così i dati in progressione geometrica di ragione 2:

$$2, 4, 8$$

avranno come media geometrica:

$$M_g = \sqrt[3]{2 * 4 * 8} = \sqrt[3]{64} = 4$$

5.2.3 *La media armonica*

Quando alcune grandezze sono inversamente proporzionali ad altre, l'equidistribuzione dei dati si ottiene attraverso il calcolo della *media armonica* (M_a).

Avendo a che fare con la somma dei reciproci dei dati si avrà:

$$\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + \dots + \frac{1}{X_n} = \frac{1}{M} + \frac{1}{M} + \frac{1}{M} + \dots + \frac{1}{M}$$

cioè

$$\Sigma \frac{1}{X_i} = \frac{N}{M}$$

Pertanto la *media armonica* (M_a) sarà :

$$(7.9) \quad M_a = \frac{N}{\Sigma \frac{1}{X_i}}$$

Se si calcolano le frequenze si avrà:

$$(7.10) \quad M_a = \frac{N}{\Sigma \frac{f_i}{X_i}}$$

Nel calcolo della media per una successione di dati in progressione armonica, dove cioè i reciproci dei dati sono in progressione aritmetica, la media armonica coincide con il valore centrale della distribuzione se i dati sono in numero dispari.

Si abbia la seguente successione armonica:

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5} \quad (N=5)$$

$$M_a = \frac{N}{\Sigma \frac{1}{X_i}} = \frac{5}{1+2+3+4+5} = \frac{1}{3}$$

Il nome armonica deriva dai pitagorici i quali chiamarono l'accordo *do-mi-sol* armonico o perfetto. Essi scoprirono che se una corda sonora vibrando interamente produce il *do*, vibrando per i 4/5 il *mi* e vibrando per i 2/3 produrrà il *sol*. Ora 4/5 è la media armonica fra 1 e 2/3.

5.2.4 *La media quadratica*

Quando la relazione tra i dati è espressa dalla somma dei loro quadrati, l'equa ripartizione dei valori si ottiene con la media quadratica.

Infatti se

$$X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2 = M^2 + M^2 + M^2 + \dots + M^2$$

$$\Sigma X_i^2 = N * M^2$$

la media quadratica sarà:

$$(7.11) \quad M_q = \sqrt{\frac{\Sigma X_i^2}{N}}$$

Pertanto la media quadratica (M_q) è la radice quadrata della media dei quadrati dei dati.

Dal momento che il quadrato è sempre un valore positivo i dati possono anche essere negativi.

Se i dati sono ordinati in tabelle di frequenza la media quadratica sarà :

$$(7.12) \quad M_q = \sqrt{\frac{\Sigma f_i X_i^2}{N}}$$

dove:

M_q = media quadratica

f_i = frequenze assolute

X_i^2 = dati al quadrato

N = numero dei dati

Per il calcolo della media quadratica ci si può servire della tab.12.4.

Tab.12.4 – Calcolo della media quadratica

1	2	3	4
Dati	frequenze	dati al quadrato	frequenze per dati al quadrato
X_i	f_i	X_i^2	$f_i \times X_i^2$
	$\Sigma f_i = N =$		$\Sigma f_i X_i^2 =$

$$M_q = \sqrt{\frac{4col.}{2col.}}$$

$$(7.13) \quad M_q = \sqrt{\frac{\Sigma f_i X_i^2}{N}}$$

E' necessario distinguere la *media quadratica* dalla media dei quadrati. La media quadratica è la *radice quadrata* della media dei quadrati (13). La media dei quadrati è uguale al quadrato della media quadratica

$$(7.14) \quad M_q^2 = \frac{\Sigma f_i X_i^2}{N}$$

quadrato della media quadratica media dei quadrati

Nella tab.12.5 viene riportato un esempio di calcolo della M_q .

Tab.12.5 – Calcolo della media quadratica

X_i	f_i	X_i^2	$f_i X_i^2$
2	3	4	12
5	4	25	70
7	6	49	254
8	1	64	64
	$\Sigma f_i = 14$		$\Sigma f_i X_i^2 = 400$

$$M_q = \sqrt{\frac{\sum f_i X_i^2}{N}} = \sqrt{\frac{400}{14}} = 5,35$$

La media quadratica insieme alla media aritmetica sono indici molto usati nella statistica.

5.2.5 *La media cubica*

Se la relazione tra i dati è espressa dalla somma dei loro cubi si calcolerà la media cubica. Così se:

$$X_1^3 + X_2^3 + X_3^3 + \dots + X_n^3 = M^3 + M^3 + M^3 + \dots + M^3$$

$$\sum X_i^3 = NX^3$$

la *media cubica* sarà

$$(7.15) \quad M_c = \sqrt[3]{\frac{\sum X_i^3}{N}}$$

Con le distribuzioni di frequenza si avrà:

$$(7.16) \quad M_c = \sqrt[3]{\frac{\sum f_i X_i^3}{N}}$$

La *media cubica* M_c è uguale alla radice cubica dei cubi dei dati diviso per il numero dei dati. Per il calcolo della media cubica ci si può servire della tab.12.6

Tab.12.6 – Calcolo della media cubica

1 Dati	2 frequenze	3 dati al cubo	4
X_i	f_i	X_i^3	$f_i X_i^3$
	$\sum f_i = N =$		$\sum f_i X_i^3 =$

$$M_c = \sqrt[3]{\frac{4col.}{2col.}} = \sqrt[3]{\frac{\sum f_i X_i^3}{N}}$$

Esempio di calcolo della media cubica(tab.12.7):

Tab.12.7 – Calcolo della media cubica

1 X_i	2 f_i	3 X_i^3	4 $f_i X_i^3$
3	2	27	54
5	4	125	500
6	5	216	1080
7	3	343	1029
9	2	729	1458

$$\sum f_i = 16$$

$$\sum f_i X_i^3 = 4121$$

$$M_c = \sqrt[3]{\frac{\sum f_i X_i^3}{N}} = \sqrt[3]{\frac{4121}{16}} = 6,36$$

5.2.6 *Le medie potenziate*

I dati possono essere elevati a qualsiasi potenza. In questo caso l'equa ripartizione è ottenuta con la media potenziata.

Indicando con M_k la *media potenziata* avremo:

$$(7.17) \quad M_k = \sqrt[k]{\frac{\sum X_i^k}{N}}$$

e per le distribuzioni di frequenze:

$$(7.18) \quad M_k = \sqrt[k]{\frac{\sum f_i X_i^k}{N}}$$

La formula (7.18) può anche essere scritta come segue:

$$(7.19) \quad M_k = \sqrt[k]{\frac{\sum f_i X_i^k}{N}} = \left(\frac{\sum f_i X_i^k}{N} \right)^{\frac{1}{k}}$$

dove:

M_k = media potenziata d'ordine k

k = esponente

f_i = frequenze

X_i = i dati

N = numero dei dati

Tutte le medie possono essere considerate come medie potenziate. Così avremo :

$$(7.20) \quad \text{per } k = 0 \quad M_0 = \sqrt[n]{X_1 * X_2 * \dots * X_n} = M_g = \text{media geometrica}$$

$$(7.21) \quad \text{per } k = 1 \quad M_1 = \frac{\sum X_i}{N} = M.A. = \bar{X} = \text{media aritmetica}$$

$$(7.22) \quad \text{per } k = -1 \quad M_{-1} = \left(\frac{\sum X_i^{-1}}{N} \right)^{-1} = \frac{1}{\frac{\sum \frac{1}{X_i}}{N}} = \frac{N}{\sum \frac{1}{X_i}} = \text{media armonica}$$

$$(7.23) \quad \text{per } k = 2 \quad M_2 = \sqrt{\frac{\sum X_i^2}{N}} = M_q = \text{media quadratica}$$

$$(7.24) \quad \text{per } k = 3 \quad M_c = \sqrt[3]{\frac{\sum X_i^3}{N}} = M_c = \text{media cubica}$$

Se sugli stessi dati si calcolano le diverse medie si troverà che la grandezza del valore medio va crescendo dalla media armonica alla media cubica secondo il seguente prospetto:

$$(7.25) \quad \dots \leq M_{arm} \leq M_{geom} \leq M_{arit} \leq M_{quadr} \leq M_{cub} \leq \dots$$

In ordine di grandezza avremo pertanto:

1. la media cubica
2. la media quadratica
3. la media aritmetica
4. la media geometrica
5. la media armonica.

Solo nel caso che si abbia a che fare con dati tutti uguali le varie medie saranno uguali tra loro.

Si nota inoltre che le medie potenziate con esponente sempre più piccolo tendono al dato minore, mentre le medie potenziate con esponente sempre più grande tendono al dato maggiore. Pertanto le infinite medie sono tutte comprese tra il valore minore ed il valore maggiore di una serie di dati.

Si voglia ad esempio calcolare le medie potenziate di ordine -1 (media armonica), di ordine 0 (media geometrica), di ordine 1 (media aritmetica), di ordine 2 (media quadratica), di ordine 3 (media cubica) dei seguenti dati :

3, 5, 6, 7, 9

compilando la tab.12.8 si avrà:

*Tab.12.8 – Calcolo delle medie potenziate d'ordine $-1, 0, 1, 2, 3$
N=5 (3;5;6;7;9)*

X_i	$\frac{1}{X_i}$	X_i^2	X_i^3
3	0,33	9	27
5	0,20	25	125
6	0,17	36	216
7	0,14	49	343
9	0,11	81	729
$\Sigma X_i = 30$	$\Sigma \frac{1}{X_i} = 0,95$	$\Sigma X_i^2 = 200$	$\Sigma X_i^3 = 1440$

$$M_{-1} = \frac{N}{\Sigma \frac{1}{X_i}} = \frac{5}{0,95} = 5,26$$

$$M_0 = \sqrt[5]{X_1 * X_2 * X_3 * X_4 * X_5} = \sqrt[5]{3 * 5 * 6 * 7 * 9} = \sqrt[5]{5670} = 5,63$$

$$M_1 = M.A. = \bar{X} = \frac{\sum X_i}{N} = \frac{30}{5} = 6$$

$$M_2 = \sqrt{\frac{\sum X_i^2}{N}} = \sqrt{\frac{200}{5}} = \sqrt{40} = 6,32$$

$$M_3 = \sqrt[3]{\frac{\sum X_i^3}{N}} = \sqrt[3]{\frac{1440}{5}} = \sqrt[3]{288} = 6,6$$

Si constata che:

$$(7.26) \quad M_{-1} \leq M_0 \leq M_1 \leq M_2 \leq M_3$$

5.2.7 *Le medie antiarmoniche*

Mediante particolari procedure di calcolo è possibile ottenere nuove medie algebriche. Se si stabilisce un rapporto tra le frequenze e i dati, la formula della media aritmetica, così ponderata, fornirà nuove medie algebriche. Così se:

$$f_i = \frac{1}{X_i}$$

si avrà:

$$(7.27) \quad M = \frac{\sum X_i \frac{1}{X_i}}{\sum \frac{1}{X_i}} = \frac{N}{\sum \frac{1}{X_i}}$$

La (7.27) corrisponde alla media armonica che può essere considerata una media aritmetica ponderata, con pesi uguali all'inverso delle intensità.

Nel caso in cui:

$$(7.28) \quad f_i = X_i^k$$

il calcolo della media aritmetica sarà

$$(7.29) \quad M.A. = \frac{\sum X_i f_i}{\sum f_i} = \frac{\sum X_i X_i^k}{\sum X_i^k} = \frac{\sum X_i^{k+1}}{\sum X_i^k}$$

L'ultimo rapporto (7.29) viene chiamata *media antiarmonica d'ordine k*.

La *media antiarmonica* d'ordine k è espressa dal rapporto tra la sommatoria delle $k+1$ potenze dei dati e la sommatoria della k potenze.

Indicando con $\overline{A}_n$ le medie antiarmoniche si avrà:

$$(7.30) \quad \overline{A}_n = \frac{\sum X_i^{k+1}}{\sum X_i^k}$$

Nella tab.12.9 vengono calcolate le medie antiarmoniche per alcuni ordini,

Tab.12.9 – calcolo delle medie antiarmoniche d'ordine 1, 2, 3, 4, 5

	X_i	X_i^2	X_i^3	X_i^4	X_i^5	X_i^6
	4	16	64	256	1024	4096
	9	81	729	6561	59048	531441
Σ	13	97	793	6817	60072	535537

$$\overline{A}_n 1 = \frac{\sum X_i^{k+1}}{\sum X_i^k} = \frac{4^2 + 9^2}{4 + 9} = \frac{97}{13} = 7,46$$

$$\overline{A}_n 2 = \frac{4^3 + 9^3}{4^2 + 9^2} = \frac{64 + 729}{97} = \frac{793}{97} = 8,18$$

$$\overline{A}_n 3 = \frac{6817}{793} = 8,596$$

$$\overline{A}_n 4 = \frac{60072}{6817} = 8,812$$

$$\overline{A}_n 5 = \frac{535537}{60072} = 8,915$$

5.2.8 Le medie radicali

Ponendo nella media aritmetica ponderata:

$$(7.31) \quad f_i = \sqrt[k]{X_i} = X_i^{\frac{1}{k}}$$

si avrà la *media radicale d'ordine k* :

$$(7.32) \quad M.A. = \frac{\sum X_i \sqrt[k]{X_i}}{\sum \sqrt[k]{X_i}} = M_{rad.k} = M.R._k$$

2

La media radicale è indicata con $M_{rad.k}$ e anche con $M.R._k$. La media radicale d'ordine k è espressa dal rapporto tra la sommatoria della radice dei dati moltiplicata per i dati stessi e la sommatoria della radice dei dati.

Calcolando alcune medie radicali d'ordine positivo collocate a destra e superiori alla M.A. di 4 e 9 (tab.12.10) si avrà:

Tab.12.10 – Calcolo di medie radicali d'ordine positivo

	X_i	f_i	$\sqrt[0]{X_i}$	$X_i \sqrt[0]{X_i}$	$\sqrt[1]{X_i}$	$X_i \sqrt[1]{X_i}$	$\sqrt[2]{X_i}$	$X_i \sqrt[2]{X_i}$
	4	1	1	4	4	16	2	8
	9	1	1	9	9	81	3	27
Σ	13	2	2	13	13	97	5	35

$$M.R._0 = \frac{\sum f_i X_i}{\sum f_i} = \frac{\sum X_i \sqrt[0]{X_i}}{\sum \sqrt[0]{X_i}} = \frac{13}{2} = 6,5 = M.A. = \overline{X}$$

$$M.R._1 = \frac{97}{13} = 7,46 = \overline{A}_{n_1} = \text{media antiarmonica d'ordine 1}$$

$$M.R._2 = \frac{35}{5} = 7$$

$$M.R._3 = \frac{25,07}{3,667} = 6,837$$

$$M.R._4 = \frac{21,245}{3,146} = 6,753$$

$$M.R._8 = \frac{16,601}{2,505} = 6,627$$

$$M.R._{16} = \frac{14,687}{2,238} = 6,5626$$

Come si può osservare la media radicale d'ordine zero corrisponde alla *media aritmetica*.

La media radicale d'ordine 1 corrisponde alla *media antiarmonica d'ordine 1*, ed è quella che fra le medie radicali *ha il valore più alto*.

Le medie radicali positive decrescono da $M.R._1$ e tendono con $k = +\infty$ alla M.A., la quale coincide con $M.R._0$.

La media antiarmonica di primo ordine, coincidente con la media radicale d'ordine 1, è quella che rende *minima la somma dei quadrati degli scarti relativi* alla X_i , divisi per la somma stessa.

Avremo così:

$$(7.33) \quad \Sigma \frac{(X_i - M.R._1)^2}{M.R._1} = \text{minimo}$$

Si osservi come le medie radicali d'ordine positivo, vadano decrescendo ad iniziare da quella d'ordine 1 e tendano asintoticamente al valore della M.A., la quale coincide con la media radicale d'ordine zero.

Ponendo

$$(7.34) \quad f_i = \sqrt[k]{\frac{1}{X_i}} = \frac{1}{\sqrt[k]{X_i}} = X_i^{-\frac{1}{k}}$$

e adoperando la formula della media aritmetica

$$(7.35) \quad M.A. = \frac{\sum f_i X_i}{\sum f_i}$$

si ottengono le medie radicali collocate a sinistra della media aritmetica.
La media radicale sinistra con esponente -1 corrisponde alla media armonica:

$$(7.36) \quad M.R._{-1} = \frac{\sum f_i X_i}{\sum f_i} = \frac{\sum \frac{1}{X_i} X_i}{\sum \frac{1}{X_i}} = \frac{N}{\sum \frac{1}{X_i}} = M_{ar}$$

Le medie radicali sinistre, con esponente negativo, al crescere dell'indice di radice k tendono alla Media Aritmetica, che coincide con la media radicale zero

Le medie radicali sinistre ad iniziare dalla media armonica vanno crescendo e tendono asintoticamente al valore della Media Aritmetica.

5.3 Le medie di posizione

Le medie di posizione prendono in considerazione i valori medi indicanti la posizione di un particolare dato. Accenneremo alla *moda*, alla *mediana* e ai *centili*.

5.3.1 La moda

La moda, indicata con M_o , viene definita come *quel valore che in una distribuzione di frequenza corrisponde al massimo della distribuzione*. La moda è il valore prevalente o dominante, quello cioè che si verifica con maggior frequenza.

Nella tab.12.11, la *moda* coincide con la frequenza massima ($f = 20$), per cui:

$$M_o = 20$$

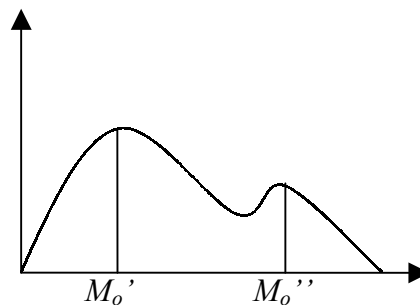
Tab.12.11 - Distribuzione di frequenza . Palline contenute nell'urna.

Colore delle palline	Frequenze assolute (f)
Nere	10
Bianche	15
Rosse	20 ← Moda
totale	45

Quando le frequenze sono tutte uguali la distribuzione è priva di moda.

Le distribuzioni che hanno una sola moda si chiamano *unimodali*. Quando le distribuzioni hanno due o più mode secondarie si chiamano *bimodali* (fig.1a), *trimodali*,...

Fig.1a. Distribuzione bimodale



In genere due distribuzioni unimodali, aventi mode diverse, unendosi originano una distribuzione bimodale.

Le scale nominali hanno come indice della tendenza centrale unicamente la moda.

Quando le frequenze sono raggruppate in classi la *classe modale* è quella che contiene la frequenza massima. In questo caso la determinazione della moda si ottiene, attraverso l'ipotesi della *ripartizione uniforme*, applicando la seguente formula:

$$(7.37) \quad M_o = L_{inf} + \frac{I.C.(f_{max} - f_{inf})}{2f_{max} - (f_{sup} + f_{inf})}$$

Dove:

M_o = moda

L_{inf} = limite inferiore della classe modale

I.C. = intervallo di classe

f_{max} = la frequenza massima, cioè la frequenza della classe modale

f_{inf} = la frequenza della classe modale immediatamente inferiore alla classe modale

f_{sup} = la frequenza della classe immediatamente superiore alla classe modale

Nelle scale ad intervalli, la moda può essere calcolata, secondo Pearson, con la seguente relazione:

$$(7.38) \quad M_o = 3 M_d - 2 M$$

Dove:

M_o = Moda

M_d = Mediana

M = Media aritmetica

5.3.2 La Mediana

La mediana coincide con il valore centrale di una distribuzione e si definisce come *quel valore che in una successione ordinata dei dati, occupa il posto centrale, cioè che ha al di sopra e al di sotto, un ugual numero di dati.*

Considerando le frequenze cumulate, la mediana è il valore che sulla curva cumulata corrisponde alla metà del numero totale dei casi. Trattandosi della curva cumulata delle frequenze relative, è quel valore che corrisponde a 0.5.

Per il **calcolo della Mediana** vanno considerati diversi casi.

a. *Dati non raggruppati in classi e non ripetuti:*

- 1) Quando il numero dei dati (N) è dispari, il posto occupato dalla Mediana si calcola mediante la seguente formula:

$$(7.39) \quad \text{posto occupato dalla Mediana in una serie ordinata} = \frac{N+1}{2}$$

dove:

N = numero dei dati

Nella seguente serie di 7 dati ordinati

$$X_i = 2 ; 4; 11; 14; 18; 21; 32$$

la Mediana corrisponde al valore che si trova al 4° posto, cioè *al valore 14* ; infatti il valore 14 è *preceduto e seguito dallo stesso numeri di dati.*

2) Quando il numero dei dati (N) è pari, la Mediana viene a trovarsi tra due valori consecutivi; solitamente la Mediana in questo caso si ottiene calcolando la Media o semisomma dei due termini mediani.

Nella seguente successione crescente di 6 termini

$$X_i = 3; 6; 8; 12; 15; 20$$

La Mediana è compresa tra il valor 8 ed il valore 12, i quali occupano il 3° ed il 4° posto ($6+1/2 = 3.5$). Calcolando la Media di questi valori si ottiene

$$M_d = \frac{8+12}{2} = 10$$

b) *Dati ripetuti*, con conseguente ripetizione di frequenze. In questo caso si ricorre alle frequenze cumulate, come illustrato nel seguente esempio.

Si voglia ad esempio calcolare la mediana della distribuzione riportata nella tab. (12.11.1) colonne 1 e 2.

Tab. 12.11.1 – Calcolo della mediana per una distribuzione di frequenze

1	2	3	4
PUNTEGGI GREZZI	FREQUENZE	FREQUENZE CUMULATE	FREQUENZE INIZIALI E FINALI DI OGNI CLASSE
X_i	f_i	$N_i = \sum_1^i N_j$	
7	1	1	1
9	2	3	2 e 3
10	4	7	4 e 7
11	7	14	8 e 14
12	4	18	15 e 18
13	5	23	19 e 23
14	22	45	24 e 45
15	16	61	46 e 61
16	12	73	62 e 73
17	16	89	74 e 89
18	4	93	90 e 93
19	10	103	94 e 103
20	3	106	104 e 106
21	4	110	107 e 110
22	2	112	111 e 112

$\Sigma = 112$

Nella colonna 3 si calcolano le frequenze cumulate, e nella colonna 4 le frequenze iniziali e le frequenze finali di ciascuna classe (in questo caso di ogni punteggio)

Avendo la distribuzione un numero pari di termini (112) la mediana si ottiene facendo la semisomma dei due termini che occupano il posto mediano; i due termini che occupano il posto mediano sono 56 e 57; la loro media è uguale a 56,5; ora 56,5 viene a trovarsi nella classe ottava, che comprende i termini da 46 a 61 e corrisponde al valore $X_i = 15$. Pertanto la mediana della distribuzione esaminata è 15:

$$Md = 15$$

c) *Dati raggruppati in classi*.

Per i dati raggruppati in classi si può ricorrere alle seguenti formula:

$$(7.39) \quad M_d = X'_{inf} + \frac{\frac{N}{3} - f_b}{f'} \times I.C.$$

Oppure

$$(7.40) \quad M_d = X'_{sup} - \frac{f_a - \frac{N}{2}}{f'} \times I.C.$$

Dove:

M_d = mediana

X'_{inf} = limite inferiore dell'intervallo di classe contenente il caso mediano

X'_{sup} = limite superiore dell'intervallo di classe contenente il caso mediano

f_b = frequenza cumulativa della classe immediatamente inferiore alla classe contenente la M_d ; cioè numero totale dei casi al di sotto del limite inferiore dell'intervallo di classe contenente la M_d

I.C. = intervallo di classe

f_a = frequenza cumulativa corrispondente alla classe contenente la M_d ; cioè il numero totale dei casi al di sotto del limite superiore della classe contenente la mediana.

5.4 Esercitazioni

1. Calcolare le varie medie (geometrica, armonica, quadratica, cubica, potenziate e antiarmoniche) delle distribuzioni di frequenza della tab.11
1. Calcolare le mediane delle distribuzioni di frequenze di tab.11
2. Calcolare le medie aritmetiche per Aspetti del sé dei soggetti 1-33, 34-91 di tab.11 infine calcolare la media aritmetica totale di questi due gruppi parziali.
3. *Al foglio elettronico (Excel) costruire le tabelle da 12.1 a 12.11*

5.5 Terminologia statistica

Approfondire con esemplificazioni i seguenti termini

1. tendenza centrale
2. valore medio
3. medie algebriche
4. media aritmetica
5. media geometrica
6. media armonica
7. media quadratica

8. media cubica
9. medie potenziate
10. medie antiarmoniche
11. medie di posizione
12. la moda
13. la mediana
14. ranghi percentili

Indicare il significato dei seguenti simboli:

1. $5+2+4+1 = 3+3+3+3$
2. $X_1 + X_2 + X_3 + X_4 = M + M + M + M$
3. $M = \sqrt[3]{X_1 * X_2 * X_3}$
4. $M = \frac{\sum X_i}{N}$
5. $M_{I+2} = \frac{(N_1 \times M_1 + N_2 \times M_2)}{N_1 + N_2}$
6. $M_a = \frac{N}{\sum \frac{1}{X_i}}$
7. $M_q = \sqrt{\frac{\sum X_i^2}{N}}$
8. $M_c = \sqrt[3]{\frac{\sum f_i X_i^3}{N}}$
9. $M_k = \sqrt[k]{\frac{\sum X_i^k}{N}}$
10. $\dots \leq M_{arm} \leq M_{geom} \leq M_{arit} \leq M_{quadr} \leq M_{cub} \leq \dots$
11. $M = \frac{\sum X_i \frac{1}{X_i}}{\sum \frac{1}{X_i}} = \frac{N}{\sum \frac{1}{X_i}}$
12. $M_o = L_{inf} + \frac{I.C.(f_{max} - f_{inf})}{2f_{max} - (f_{sup} + f_{inf})}$
13. $M_o = 3 M_d - 2 M$

$$14. \mathbf{M}_d = \mathbf{X}'_{inf} + \frac{\frac{N}{3} - f_b}{f'} \times \mathbf{I.C}$$