

Unità Didattica N° 17:

Distribuzioni statistiche e loro rappresentazione grafica

- 1) L'indagine statistica e le sue fasi**
- 2) Universo statistico , campione , unità statistiche**
- 3) Variabile statistica**
- 4) Mutabile statistica**
- 5) Frequenze assolute , relative , percentuali , cumulate**
- 6) Distribuzioni di frequenze**
- 7) Ogiva di frequenza**
- 8) Brevi considerazioni generali sulle rappresentazioni grafiche**
- 9) Rappresentazione grafica di serie statistiche :**
 - a) diagrammi a punti , b) ideogramma , c) diagrammi a nastri d) diagrammi a colonne**
 - e) diagrammi a settori circolari f) diagramma polare**
- 10) Rappresentazione grafica di seriazioni statistiche :**
 - a) diagramma a bastoni c) istogramma**

Definizione , scopi ed applicazioni della statistica

- La **statistica** è la scienza che studia i **fenomeni collettivi** o di massa . Un fenomeno è detto **collettivo** o di massa quando la sua determinazione è possibile solo attraverso una molteplicità di osservazioni . L'occupazione in Italia negli ultimi 10 anni è un esempio di **fenomeno collettivo** in quanto la sua conoscenza presuppone le osservazioni circa il lavoro di una molteplicità di persone . Dato un fenomeno collettivo , la statistica insegna ad individuare le modalità in cui si manifesta , a descriverlo sinteticamente , ed a trarre da esso conclusioni più generali riguardanti fenomeni più ampi .

- I fenomeni oggetto di studio possono essere classificati in due gruppi : **fenomeni tipici** e **fenomeni atipici** . Sono fenomeni tipici quelli che si manifestano , in ogni circostanza che li osserviamo , con le stesse caratteristiche . Sono **fenomeni tipici** una reazione chimica tra un acido ed una base , la caduta di un grave nel vuoto , la temperatura di ebollizione di un liquido . Per la conoscenza delle caratteristiche di questi fenomeni è sufficiente osservarli una sola volta .

Si dicono , invece , **fenomeni atipici** quelli che non si manifestano con le stesse caratteristiche in ogni circostanza in cui li osserviamo . Sono **fenomeni atipici** i redditi delle varie famiglie di una nazione , le stature di un gruppo di ragazzi di 10 anni . Se si vogliono conoscere le caratteristiche dei fenomeni atipici non è sufficiente osservarli nel singolo caso . La conoscenza dei fenomeni atipici richiede che essi vengano osservati in una pluralità di casi , o come suole dirsi in statistica , in un **collettivo** di casi . Si osservi però che le caratteristiche dei fenomeni atipici non risulteranno subito evidenti , anzi l'osservazione di un fenomeno atipico in un **collettivo di casi** darà luogo ad una serie di dati che a prima vista potrebbero apparire disordinati e privi di qualsiasi regolarità .

Sottoponendo tali dati ad opportune elaborazioni sarà possibile fare emergere da essi alcune regolarità che caratterizzano il fenomeno atipico indagato . Dette regolarità prendono il nome di **leggi empiriche** o **leggi statistiche** .

- ridurre ad una sola osservazione . E' un *fenomeno individuale* l'età di una singola persona , l'altezza di una casa , il titolo di un libro .

- Si dice che un **fenomeno** è **collettivo** o di massa quando l'indagine su quel fatto richiede una molteplicità di osservazioni su fenomeni o fatti individuali aventi tutti un carattere comune . Sono **fenomeni collettivi** la distribuzione del reddito dei cittadini di una nazione , la distribuzione per età dei medesimi o quelli della loro statura .

• **La statistica è la scienza che ha come fine lo studio quantitativo o qualitativo dei fenomeni collettivi** . La statistica può essere suddivisa in **statistica descrittiva** e statistica inferenziale . Compito della *statistica descrittiva* è quello di raccogliere , ordinare , riassumere, presentare ed analizzare i dati ottenuti dallo studio di un fenomeno collettivo . Compito della *statistica inferenziale* è quello di estendere i risultati ottenuti con la statistica descrittiva ad un insieme di elementi più vasto di quello analizzato e studiato .

Chiariamo meglio quanto detto . In una indagine in cui vengono raccolti i dati di tutta la popolazione statistica , il fenomeno collettivo risulta sostanzialmente e inequivocabilmente determinato , anche se non ancora chiaro a causa della frammentazione in cui si presenta , a sua volta dovuta alla molteplicità delle informazioni disponibili . Si tratta quindi di analizzare i dati , trovando gli aspetti comuni o le possibili connessioni tra essi , al fine di descrivere sinteticamente ed in maniera efficace l'andamento complessivo del fenomeno . Di tutto ciò si occupa una parte della statistica chiamata **statistica descrittiva** o deduttiva . Il suo compito è dunque quello di analizzare e descrivere una certa quantità di dati senza trarre da essi conclusioni più generali . In un'indagine in cui vengono raccolti dati solo su un campione della popolazione statistica , il fenomeno collettivo non è determinato inequivocabilmente in tutte le sue unità . Però , sulla base dei dati del campione , si possono trarre indicazioni circa l'andamento del fenomeno su tutta la popolazione .

Esempio : durante lo scrutinio dei voti relativi ad una consultazione elettorale , esperti di statistica , dall'esame dei voti scrutinati , fanno previsioni circa i voti di tutta la popolazione . In altre parole , generalizzano i risultati del campione estendendoli a tutta la popolazione . La parte della statistica che studia le condizioni sotto le quali è lecito operare tale generalizzazione o **inferenza** , è detta **statistica inferenziale** .

L 'indagine statistica e le sue fasi

Nello svolgimento di una qualsiasi indagine statistica distinguiamo le seguenti fasi :

- 1) **definizione degli obiettivi**
- 2) **individuazione del fenomeno collettivo**
- 3) **rilevazione dei dati**
- 4) **spoglio dei dati**
- 5) **elaborazione ed analisi**
- 6) **interpretazione dei risultati e divulgazione** .

Prima di iniziare una ricerca bisogna avere ben chiaro lo **scopo** al fine di stabilire con precisione su che cosa si desidera condurre l'indagine e quali sono i caratteri che si vogliono rilevare . Così , volendo studiare la diffusione degli incidenti automobilistici in una certa regione , si preciseranno i tipi di strade su cui effettuare la rilevazione (comunali , provinciali , superstrade) , la zona che si considera , la cilindrata della macchina , ecc..

Ogni *singolo soggetto* in esame (nel nostro caso l ' automobilista) prende il nome di **unità statistica** , il rilievo che si esegue dà luogo all' **osservazione** e l'insieme delle osservazioni costituisce il **dato statistico** . L ' indagine può essere condotta su tutti i soggetti che interessano (**popolazione statistica** o **universo statistico**) oppure , come di norma avviene , su una parte di essi (**campione statistico**) .

Eseguita la scelta del campione , si proseguirà alla **rilevazione** statistica vera e propria , che consiste nell'annotare su apposite schede , opportunamente approntate e che risultano diverse a seconda dell'indagine che si intende compiere , i dati che si vogliono raccogliere . La raccolta dei dati deve essere il più possibile **esatta e completa** . (*)

La mancanza di questi requisiti compromette l'esito dell'indagine . Effettuata la rilevazione si esegue lo **spoglio** delle schede ed i dati raccolti vengono evidenziati per mezzo di :

- a) **tabulazioni** che consistono nel sistemare in tabelle i risultati dello spoglio
 - b) **rappresentazioni grafiche** che ci forniscono immediatamente l'andamento di un dato fenomeno
- Dopo lo *spoglio* viene effettuata l ' **elaborazione dei dati** che consiste nella scelta e nell'applicazione dei procedimenti matematici propri del metodo statistico .

Infine abbiamo l ' **interpretazione dei risultati** , e questa parte del procedimento statistico è interamente affidata alla perizia , al senso critico ed alla capacità interpretativa del ricercatore .

Distribuzioni statistiche semplici

Si chiama **collettività statistica** o popolazione statistica o *universo statistico* un insieme di elementi considerati omogenei rispetto ad uno o più caratteri . Ogni elemento appartenente alla popolazione statistica prende il nome di **unità statistica** .

Nello svolgimento di una qualsiasi indagine statistica distinguiamo le seguenti fasi :

- 1) **definizione degli obiettivi** 2) **individuazione del fenomeno collettivo** 3) **rilevazione dei dati**
- 4) **spoglio dei dati** 5) **elaborazione ed analisi** 6) **interpretazione dei risultati e divulgazione** .

Quando non è possibile o non conviene effettuare indagini sull'intera popolazione statistica , allora si procede per **campione** . I motivi che inducono ad effettuare indagini per campione sono relativi soprattutto ai costi ed ai tempi :

(*) Si chiama **rilevamento statistico** l'operazione di raccolta dei dati statistici riguardanti una determinata popolazione di individui in possesso di un medesimo carattere , relativo al fenomeno che si vuole studiare . Un rilevamento statistico può essere saltuario o continuo , parziale o totale , pubblico o privato . Per i rilevamenti statistici

a) le risorse da impiegare per una raccolta esaustiva di dati possono risultare troppo dispendiose in relazione al risultato da raggiungere

b) i tempi necessari all'elaborazione dei dati crescono al crescere del numero dei dati stessi , fino a raggiungere tempi di attesa eccessivi.

Sono state messe a punto tecniche per individuare **campioni** con caratteristiche rappresentative dell'universo da cui sono estratti . Il problema del **campionamento** e dell ' inferenza statistica saranno argomenti di successivi approfondimenti . Di ogni unità statistica si studiano i **caratteri** , la cui scelta dipende dal particolare problema che si vuole esaminare e dalle ipotesi iniziali .

Ogni **carattere** si manifesta mediante le sue modalità cioè mediante i diversi modi di apparire , ciascuno dei quali è chiamato **modalità** .

I caratteri possono essere **qualitativi** e **quantitativi** , secondo che le modalità con cui si presentano sono espresse mediante attributi , aggettivi , denominazioni , oppure sono espresse mediante valori numerici . Sono **caratteri qualitativi** , il titolo di studio posseduto da ciascuna unità statistica , il tipo di attività esercitato dai lavoratori di un'azienda , la regione di residenza degli italiani .

Sono **caratteri quantitativi** l'altezza delle persone di una provincia , il numero di studenti iscritti ad un determinato tipo di scuola, la quantità di pioggia caduta in un determinato periodo dell'anno , il reddito pro-capite per le diverse categorie di lavoratori . I **caratteri quantitativi** possono essere **discreti** (quando possono assumere valori appartenenti ad insiemi finiti o infiniti ma numerabili) o **continui** (quando hanno modalità appartenenti ad un intervallo di numeri reali) **estensivi** (modalità che indicano la frequenza assoluta) .

I caratteri di tipo quantitativo sono chiamati anche **variabili statistiche** e vengono indicate con lettere maiuscole come **X** o **Y** mentre i caratteri qualitativi prendono il nome di **mutabili statistiche** e , di solito , vengono indicate con la lettera maiuscola **M** .

Il **dato statistico** esprime il numero totale di unità statistiche che presentano la **stessa modalità** .Coincide con la frequenza o la intensità di una modalità di un carattere .

L'insieme dei dati relativi alle diverse modalità di un singolo carattere costituisce una **distribuzione statistica** . In una distribuzione statistica relativa a mutabili o a variabili discrete interessa studiare la **frequenza** .

DEFINIZIONE

sono necessari appositi registri,moduli,questionari,schede...su cui riportare e ordinare tutte le informazioni raccolte sul carattere o i caratteri del fenomeno in esame,per potere,alla fine,costruire le cosiddette **tabelle statistiche** relative

Si chiama **frequenza assoluta** di una modalità il numero delle volte che tale modalità compare nel collettivo statistico osservato . Si chiama **frequenza relativa** di una modalità il rapporto tra la sua frequenza assoluta e il numero di unità statistiche del collettivo osservato . Se il valore viene rapportato a 100 si ha una **frequenza relativa percentuale** . Si chiama **distribuzione di frequenze** l'insieme delle coppie ordinate il cui primo elemento corrisponde alla modalità del carattere ed il secondo elemento alla sua frequenza assoluta o relativa .

n_k = **frequenza assoluta** , N = **numero totale delle unità statiche** dell'intero collettivo

$$r_k = \frac{n_k}{N} = \text{frequenza relativa} \quad p_k = \frac{n_k}{N} \cdot 100 = \text{frequenza percentuale}$$

- a) La somma delle frequenze assolute è uguale al numero N di unità statistiche considerate
- b) la somma delle frequenze relative è uguale ad 1
- c) la somma delle frequenze percentuali è uguale a 100

DEFINIZIONE

Si chiama **frequenza relativa cumulata** ad una modalità x_k il numero di unità che presentano modalità minore o uguale ad x_k . Possiamo avere **frequenze assolute cumulate** , frequenze relative cumulate , frequenze percentuali cumulate .

$$C_k = n_1 + n_2 + \dots + n_k = \text{frequenze assolute cumulate}$$

$$F_k = \frac{n_1}{N} + \frac{n_2}{N} + \dots + \frac{n_k}{N} = r_1 + r_2 + \dots + r_k = \text{frequenze relative cumulate}$$

$$P_k = 100 \cdot F_k = \frac{n_1}{N} \cdot 100 + \frac{n_2}{N} \cdot 100 + \dots + \frac{n_k}{N} \cdot 100 = p_1 + p_2 + \dots + p_k =$$

$$= \text{frequenze percentuali cumulate}$$

Dalla definizione di frequenza cumulata deriva che il valore relativo all'ultima modalità è uguale al totale complessivo N per le frequenze assolute , ad **1** per le frequenze relative , a **100** per quelle percentuali . La seguente tabella riporta le frequenze assolute , le frequenze assolute cumulate , le frequenze relative cumulate , le frequenze percentuali cumulate dei voti di italiano riportati dagli alunni di un liceo .

voto dello	Studenti che hanno riportato tale voto in italiano
------------	--

scrutinio finale modalità x_i	frequenze					
	assolute n_i	assolute cumulate C_i	relative r_i	relative cumulate F_i	percentuali p_i	percentuali cumulate P_i
$x_1 = 3$	$n_1 = 10$	$C_1 = 10$	$r_1 = 0,03$	$F_1 = 0,03$	$p_1 = 3$	$P_1 = 3$
$x_2 = 4$	$n_2 = 25$	$C_2 = 35$	$r_2 = 0,08$	$F_2 = 0,11$	$p_2 = 8$	$P_2 = 11$
$x_3 = 5$	$n_3 = 34$	$C_3 = 69$	$r_3 = 0,12$	$F_3 = 0,23$	$p_3 = 12$	$P_3 = 23$
$x_4 = 6$	$n_4 = 136$	$C_4 = 205$	$r_4 = 0,46$	$F_4 = 0,68$	$p_4 = 46$	$P_4 = 69$
$x_5 = 7$	$n_5 = 68$	$C_5 = 273$	$r_5 = 0,23$	$F_5 = 0,91$	$p_4 = 23$	$P_5 = 92$
$x_6 = 8$	$n_6 = 22$	$C_6 = 295$	$r_6 = 0,07$	$F_6 = 0,98$	$p_6 = 7$	$P_6 = 99$
$x_7 = 9$	$n_7 = 3$	$C_7 = 298 = N$	$r_7 = 0,01$	$F_7 = 1$	$p_7 = 1$	$P_7 = 100$
	$N = 298$					

Il numero degli studenti insufficienti sono 69 e tale numero corrisponde alla frequenza assoluta cumulata relativa alla modalità 5 .

unità statistica = alunno di un liceo ; **carattere unità statistica** = voto di italiano nello scrutinio finale ; **modalità del carattere** = 3,4,5,6,7,8,9 (si tratta di un carattere quantitativo)

Il **dato statistico** rispetto alla modalità 3 è 10 , rispetto alla **modalità** 8 è 22 .

La classificazione dei dati statistici avviene mediante raggruppamenti in classi omogenee ed i dati statistici , alla fine , si presentano sotto forma di tabelle .

Le **tabelle statistiche** sono dette tabelle semplici se sono riferite ad un solo carattere , mentre sono dette **tabelle complesse** o tabelle multiple se sono riferite a due o più caratteri .

Le *tabelle a doppia entrata* possono essere :

- 1) di **correlazione** se i due caratteri sono entrambi quantitativi
- 2) di **contingenza** se i due caratteri sono qualitativi
- 3) **miste** se uno dei due caratteri è qualitativo e l'altro quantitativo

Si ha una **serie statistica** se il carattere è qualitativo cioè se le modalità sono espresse da attributi , si ha una **seriazione statistica** se il carattere è quantitativo , cioè se le modalità sono espresse da numeri .

Data una distribuzione di frequenza di un carattere , si definisce :

variabile statistica X l'insieme delle modalità x_i ($i = 1, 2, \dots, k$) di un **carattere quantitativo** e delle frequenze n_i associate. Essa può essere indicata in uno dei seguenti modi :

$$X = \begin{cases} x_1, x_2, \dots, x_k \\ n_1, n_2, \dots, n_k \end{cases} \text{ per le } \mathbf{seriazioni di frequenza}$$

SERIAZIONE DI FREQUENZE		
Modalità del carattere quantitativo X	Frequenze del carattere quantitativo X	Tabella di frequenze per una variabile statistica
x_1	n_1	
x_2	n_2	
$\vdots$	$\vdots$	
x_k	n_k	

Seriazione	Frequenze
x_1	n_1
x_2	n_2
$\vdots$	$\vdots$
x_i	n_i
$\vdots$	$\vdots$
x_k	n_k
Totale	N

Carattere discreto per modalità

x_1	n_1
x_2	n_2
$\vdots$	$\vdots$
x_i	n_i
$\vdots$	$\vdots$
x_k	n_k
Totale	N

Carattere discreto per classi

$x_1 - x_2$	n_1
$x_3 - x_4$	n_2
$\vdots$	$\vdots$
$x_i - x_{i+1}$	n_i
$\vdots$	$\vdots$
$x_{k-1} - x_k$	n_k
Totale	N

Voto di italiano X	Alunni n
--------------------	----------

$x_1 = 3$	$n_1 = 1$
$x_2 = 4$	$n_2 = 2$
$x_3 = 5$	$n_3 = 4$
$x_4 = 6$	$n_4 = 14$
$x_5 = 7$	$n_5 = 3$
$x_6 = 8$	$n_6 = 1$
Totale	$N = 25$

Distribuzione degli alunni
di una classe secondo
il voto di italiano

Classi di stature (cm)	Reclute n_i
fino a 150	10
150 – 160	40
160 – 170	180
170 – 180	210
189 – 190	50
oltre 190	10
totale	500

Distribuzione
di un gruppo
di reclute
secondo la
statura

Classi di ampiezza (numero posti letto)	Cliniche n_i
fino a 25	23
26 – 50	158
51 – 75	134
76 – 100	101
101-125	43
126 – 150	50
151 – 200	27
201 – 250	10
251 – 350	11
351 e più	3
totale	560

Distribuzione
di un gruppo
di cliniche
private
secondo
l'ampiezza

Data una distribuzione di frequenza di un carattere , si definisce :

mutabile statistica M l'insieme delle modalità osservate m_i ($i = 1, 2, \dots, k$) di un **carattere qualitativo** e delle frequenze n_i ad esse associate. Essa può essere indicata in uno dei seguenti

modi:
$$M = \begin{cases} M_1, M_2, \dots, M_k \\ n_1, n_2, \dots, n_k \end{cases} \text{ per le } \mathbf{serie \text{ di frequenza}}$$

Serie	Frequenze
m_1	n_1
m_2	n_2
$\vdots$	$\vdots$
m_i	n_i
$\vdots$	$\vdots$
m_k	n_k
Totale	N

SERIE DI FREQUENZE		
Modalità del carattere qualitativo M	Frequenze del carattere qualitativo M	Tabella di frequenze per una mutabile statistica
m_1	n_1	
m_2	n_2	
$\vdots$	$\vdots$	
m_k	n_k	

Le **serie statistiche** possono essere di tre specie: di **fatto** o qualitative, di **luogo** o territoriali, di **tempo** o cronologiche o *storiche* a seconda che espongano la composizione qualitativa del fenomeno, o la sua distribuzione nello spazio o la sua successione nel tempo.

Le **serie storiche** o **temporali** scaturiscono dallo spoglio di un carattere temporale. Ecco un esempio di serie storia: Le serie storiche si distinguono in **connesse** e **non connesse**. Sono **serie storiche connesse** quelle in cui la manifestazione quantitativa di un periodo influisce sulla manifestazione quantitativa del periodo successivo, mentre prendono il nome di **serie storiche non connesse** tutte le altre in cui le manifestazioni quantitative dei diversi periodi appaiono fra loro indipendenti.

Anni	Nati vivi
1973	874.546
1974	868.882
1975	827.552
1976	781.638
1977	742.546
1978	709.043
1979	670.078
1980	644.001
1981	621.805
1982	617.507
1983	600.218

Nati vivi in
Italia negli
anni dal
1973 al 1983

Le *serie storiche* possono essere di due tipi : **serie storiche di flusso** (o di *movimento*) e serie storiche **distato** (di fondo) . Le **serie storiche di flusso** sono quelle che riguardano fenomeni che si realizzano nel tempo (produzione , nascite , migrazioni , reddito ,..) e che pertanto possono essere osservati solo con riferimento ad intervalli di tempo . In generale vengono considerati intervalli di tempo adiacenti e tutti della stessa ampiezza . La serie storica dei nati negli anni dal 1973 al 1983 fa riferimento agli 11 intervalli annuali (dal 1° gennaio al 31 dicembre) degli anni 1973,...,1983 e si riferisce ad un fenomeno di flusso . Essa è pertanto una serie storica di flusso . Le **serie storiche di fondo** o di stato sono quelle che riportando le intensità di un fenomeno riferite ad una serie di istanti di tempo .

In genere si considerano istanti di tempo successivi ed equidistanti . nella seguente tabella è riportato il totale della popolazione italiana alla fine degli anni indicati , cioè con riferimento agli istanti di tempo pari alla mezzanotte del 31 dicembre di ciascuno degli anni indicati .

Anni	Popolazione residente
1981	56.536.499
1982	56.742.468
1983	56.929.101
1984	57.080.498

Popolazione residente
in Italia alla fine
degli anni indicati

Le **serie territoriali** o **geografiche** o **spaziali** scaturiscono dallo spoglio di un carattere geografico .
Ecco un esempio di serie territoriale .

Gruppi di cause di morte	Morti
Malattie infettive e parassitarie	3.246
Tumori	124.570
Disturbi psichici	7.956
Malattie del sistema circolatorio	245.068
Malattie dell'apparato respiratorio	33.475
Malattie dell'apparato digerente	29.844
Altre malattie	36.108
Stati morbosi mal definiti	14.487
Cause violente	27.578
Totale	522.332

Distribuzione dei
morti in Italia
nel 1982 secondo
gruppi di cause
di morte

Se il criterio ordinatore si articola in modalità di tipo nominale (e cioè se non esiste un ordine logico o naturale delle modalità che possono essere disposte a piacere) si ha una **mutabile sconnessa** o **serie sconnessa** . Una serie geografica è un caso particolare di serie sconnessa , di una serie le cui modalità non hanno un ordine logico di nessun tipo .

Le distribuzioni di frequenza o di intensità possono essere visualizzate mediante rappresentazioni grafiche . Le più ricorrenti sono :

- 1) i **diagrammi cartesiani** (**diagrammi a punti** , a **segmenti** , a **bastoni**)
- 2) gli **istogrammi**
- 3) gli **ortogrammi** (o diagrammi a nastri)
- 4) i **diagrammi polari**
- 5) i **diagrammi a settori circolari**
- 6) i **cartogrammi** e gli **ideogrammi**

Rappresentazione grafica di serie statistiche

La rappresentazione grafica più semplice di una serie statistica è fornita dal **diagramma a punti** , che consiste nell'individuare , sul piano cartesiano, i punti corrispondenti alle coppie di valori da attribuire uno all'ascissa x e l'altro all'ordinata y .

Anni	Suicidi	
	M	F
1864	516	130
1874	762	253
1884	1115	255
1894	1381	351
1904	1705	451
1914	2294	891
1924	1778	1000
1934	2703	983
1944	1510	615
1954	2032	874

Morti per suicidio secondo il sesso dal 1864 al 1954

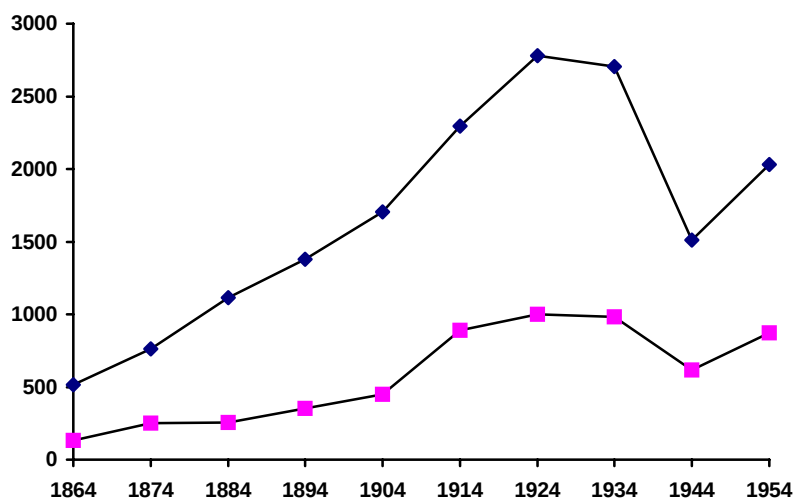


Diagramma a punti

Per facilitare la lettura del grafico si usa congiungere mediante segmenti i punti successivi $P(x,y)$.

Un altro interessante e suggestivo metodo per rappresentare una *serie statistica* è dato dall' **ideogramma**.

In questo tipo di rappresentazione si raffigura la distribuzione di un fenomeno attraverso disegni che richiamano in maniera intuitiva con la loro forma il fenomeno descritto. In genere si usano delle figure schematizzate come: 1) omini 2) spighe 3) monete 4) ciminiere per rappresentare rispettivamente 1) la popolazione 2) la produzione di cereali 3) il reddito 4) la produzione di acciaio

L ' **ideogramma** è una forma di rappresentazione grafica che si presenta particolarmente adatta per scopi pubblicitari . Solitamente gli ideogrammi vengono usati nei libri di geografia per illustrare comparativamente i prodotti (agricoli o di allevamento) delle varie nazioni .

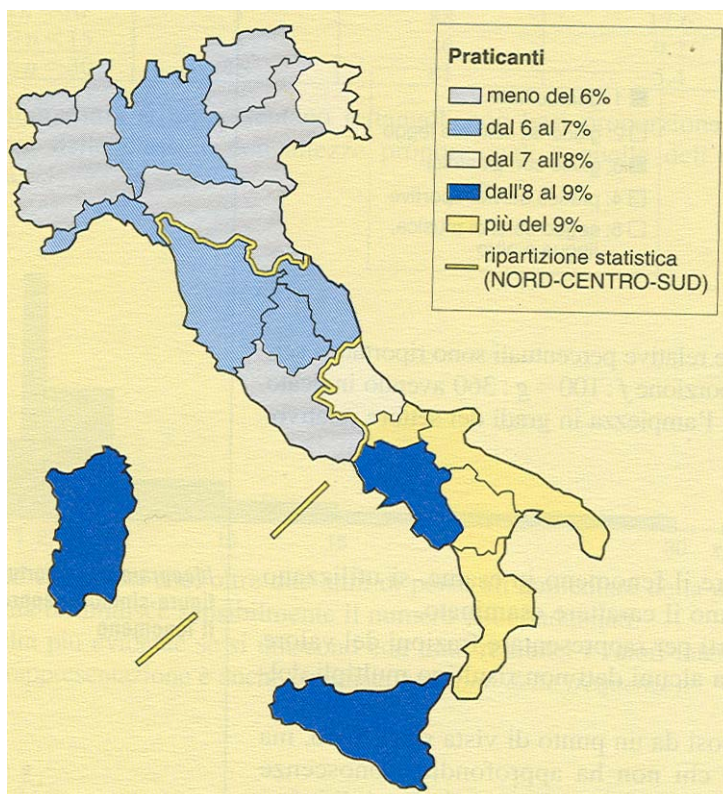
Gli **ideogrammi** utilizzano figure che richiamano il contenuto del fenomeno e servono per dare una visione immediata di un fenomeno .



I **cartogrammi** rappresentano fenomeni legati al territorio geografico sul quale si manifestano . La frequenza del fenomeno è indicata dalla diversità del colore . Il **cartogramma** è una distribuzione statistica territoriale .

Il cartogramma della figura indica la percentuale di un'attività sportiva praticata nelle diverse regioni italiane .

(Da una fonte del **Touring Club Italiano**)



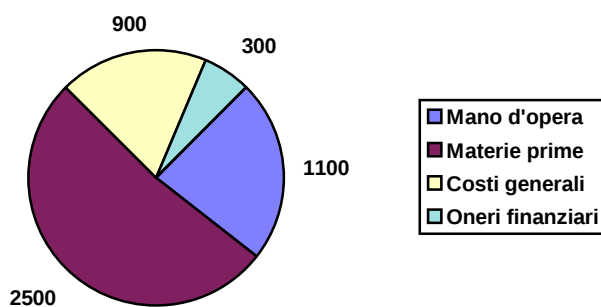
Si può ricorrere anche ad altre rappresentazioni grafiche , quali i **diagrammi a nastri** , i **diagrammi a colonne** , ed i diagrammi a settori circolari la cui realizzazione è piuttosto semplice e rapida . In questi grafici , al fine di valutare il significato statistico della lunghezza dei nastri e delle colonne , è necessario affiancare ad essi una scala di misura di riferimento .

Quando si vuole evidenziare sia il valore assoluto di un certo fenomeno sia la sua composizione si può ricorrere ad un **grafico a settori circolari**.

L'area del cerchio viene resa proporzionale al valore assoluto del fenomeno in esame. L'angolo al centro, pari a 360° , rappresenta il 100% del fenomeno osservato, quindi si individuano nel cerchio diversi settori, ognuno dei quali avrà un angolo al centro di ampiezza proporzionale alla percentuale della caratteristica che si vuole rappresentare.

Costi di produzione	1972	1977
Mano d'opera	900	1100
Materie prime	1800	2500
Costi generali	300	900
Oneri finanziari	200	300

Grafico a settori circolari
Costi di produzione 1977



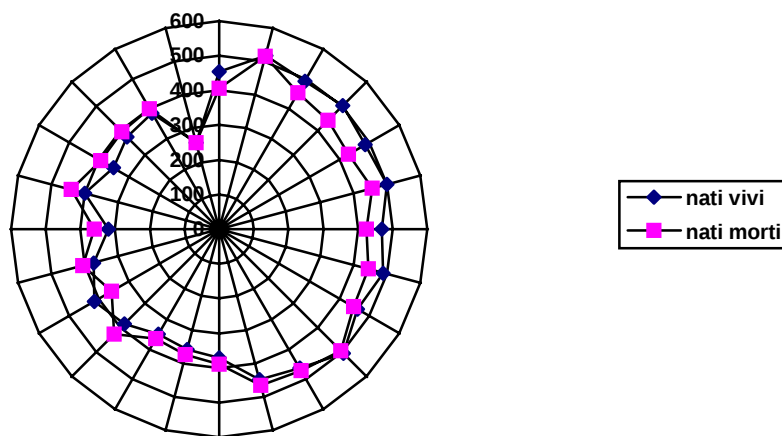
Infine, ci sono le rappresentazioni a **coordinate polari** a cui generalmente si fa ricorso quando si considerano fenomeni ciclici, cioè fenomeni che risentono dell'influsso di fattori che agiscono periodicamente.

Il **diagramma polare** è caratterizzato dall'indicazione delle due coordinate in relazione ad un punto centrale detto **polo** del grafico. Le ordinate sono lineari e vengono misurate dalla distanza dal polo, mentre le ascisse sono angolari, cioè sono costituite dall'angolo formato da **raggio vettore** e dal raggio di riferimento.

La costruzione materiale del grafico si ottiene fissando il polo, indi, scelto il raggio di riferimento, si divide a partire da esso, l'angolo giro in tante parti uguali (che costituiscono le ascisse) quante sono le fasi cicliche che si intendono rappresentare (nel nostro caso 24), mediante raggio partenti dal polo.

Si eseguono poi tanti cerchi con centro comune nel polo che segnano i principali valori della scala delle ordinate . I punti così ottenuti possono essere congiunti mediante una spezzata .

Nati vivi e nati morti secondo l'ora della nascita

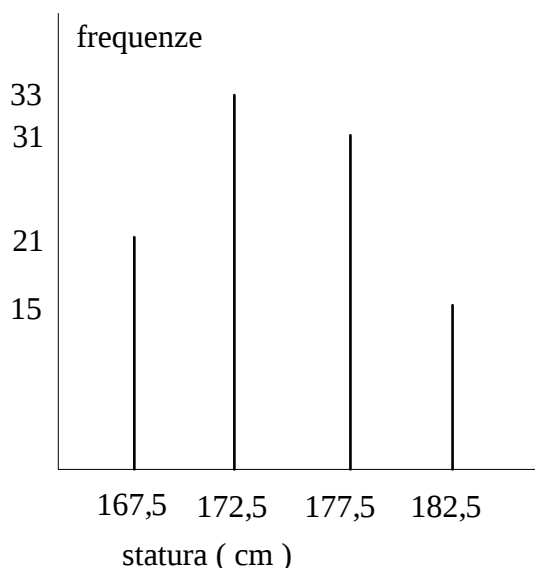


Rappresentazione grafica di seriazioni statistiche

La rappresentazione grafica più semplice di una seriazione statistica è il **diagramma a bastoni** , mediante il quale a ciascun valore della variabile oppure ad ogni classe della stessa si fa corrispondere, parallelamente all'asse delle ordinate , un segmento che s'innalza da un punto la cui ascissa è pari al valore osservato dalla variabile o al valore centrale della classe , mentre la sua lunghezza è proporzionale alla frequenza .

Classi di statura	Valori centrali	% degli iscritti
165–170	167,5	21
170–175	172,5	33
175–180	177,5	31
180-185	182,5	15

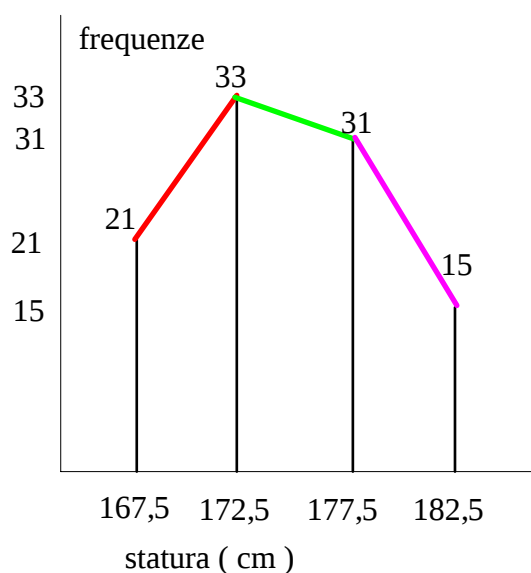
Ripartizione percentuale ,
in alcune classi di statura,
degli iscritti alla leva dei nati
nel 1955 (Toscana)



Ripartizione % , in alcune classi di statura ,
degli iscritti alla leva dei nati nel 1955
(Toscana)

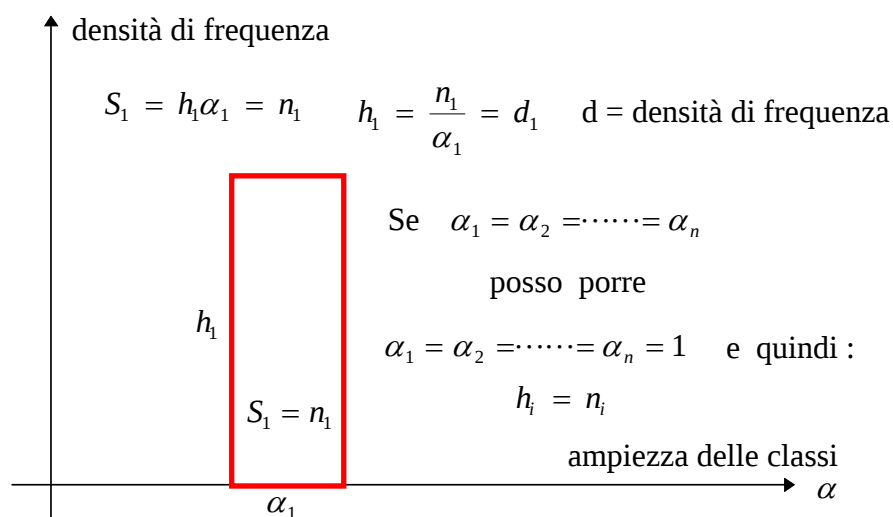
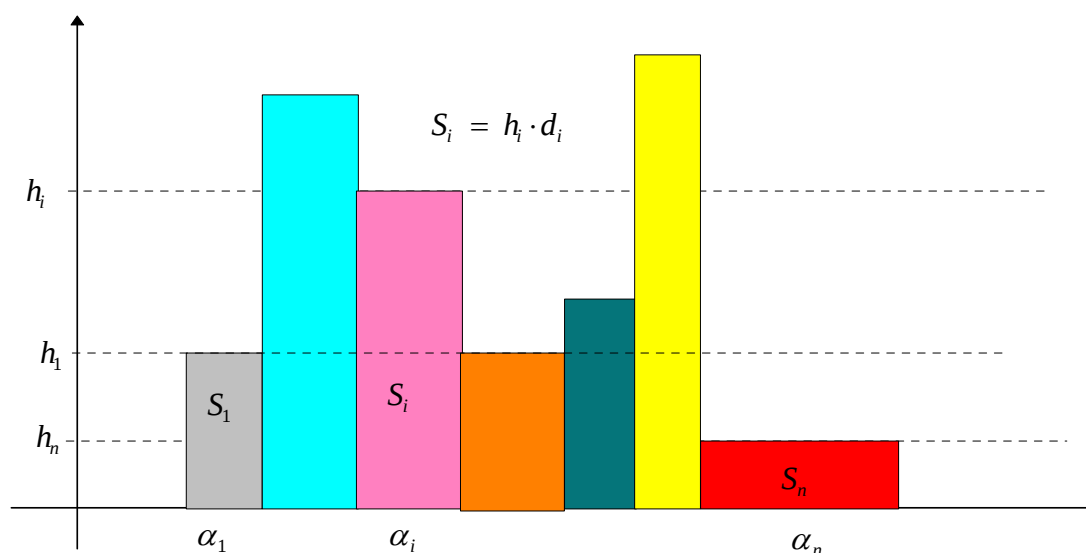
Diagramma a bastoncini

Se uniamo mediante una spezzata i vertici dei segmenti , da un diagramma a bastoncini otteniamo il **poligono di frequenza** .



La rappresentazione grafica di **seriazioni** più frequente in statistica è data dall'**istogramma** . In questo tipo di grafico le superfici rettangolari sono proporzionali , per ciascuna classe , alle corrispondenti frequenze . Pertanto la superficie dell'intero grafico rappresenta il complesso delle frequenze considerate , mentre l'area di un dato rettangolo rappresenta la frequenza dei valori compresi nel corrispondente intervallo .

Gli **istogrammi** sono costituiti da una serie di rettangoli contigui che si sviluppano lungo l'asse orizzontale , ed hanno come basi le ampiezze delle classi e come altezze le rispettive densità di frequenza . L'area dell'intero istogramma ci fornisce il totale complessivo N dell'intera popolazione statistica , mentre l'area di ciascun rettangolo ci fornisce la frequenza assoluta di ciascuna classe .



Se in un istogramma si uniscono i punti medi delle basi superiori dei rettangoli si ottiene una spezzata detta **poligono di frequenza**. Il poligono è esteso anche alle classi estreme aventi frequenza nulla.

E' facile verificare che, quando le basi sono uguali, l'area della parte di piano delimitata dall'asse orizzontale e dal poligono di frequenza è uguale alla somma delle aree dei rettangoli.

Quindi in ascissa riportiamo le ampiezze delle varie classi ed in ordinata i valori delle corrispondenti densità di frequenza. Se tutte le classi hanno la stessa ampiezza possiamo riportare in ordinata le frequenze assolute convenendo di porre numericamente uguale ad uno l'ampiezza di ciascuna classe.

Distribuzioni di frequenza e loro rappresentazione grafica : diagramma di frequenza e funzione di ripartizione

Una distribuzione di frequenza è una distribuzione che ha una delle seguenti strutture :

$$X \begin{cases} x_1 & x_2 & \cdots & x_k \\ n_1 & n_2 & \cdots & n_k \end{cases} \quad \text{distribuzione di frequenza assoluta}$$

$$X \begin{cases} x_1 & x_2 & \cdots & x_k \\ f_1 & f_2 & \cdots & f_k \end{cases} \quad \text{distribuzione di frequenza relativa}$$

$$X \begin{cases} x_1 & x_2 & \cdots & x_k \\ F_1 & F_2 & \cdots & F_k \end{cases} \quad \text{distribuzione di frequenza relativa cumulata}$$

$$X \begin{cases} x_1 & x_2 & \cdots & x_k \\ P_1 & P_2 & \cdots & P_k \end{cases} \quad \text{distribuzione di frequenza percentuale cumulata}$$

Il grafico di una distribuzione di frequenza assoluta o relativa , detto **diagramma di frequenza** , va costruito in base alle seguenti considerazioni :

- in ascisse bisogna riportare il valore della modalità x_i della variabile **X**
- si riporta in ordinate il valore della frequenza assoluta n_i o della frequenza relativa f_i

Costruire il **diagramma di frequenza** della seguente distribuzione di frequenza relativa

(**variabile discreta semplice**)

Famiglie residenti per numero di componenti al 1961			
Ampiezza delle famiglie (numero componenti X)	Numero delle famiglie (frequenza assoluta n_i)	Frequenza relativa f_i	Frequenza Cumulata F_i
$x_1 = 1$	1.464.277	0,10	0,10
$x_2 = 2$	2.693.471	0,196	0,296
$x_3 = 3$	3.086.113	0,27	0,56
$x_4 = 4$	2.797.966	0,204	0,77
$x_5 = 5$	1.726.469	0,13	0,9
$x_6 = 6$	956.537	0,06	0,96
$x_7 = 7$	493.596	0,036	0,996
$x_8 = 8$	254.970	0,002	0,998
$x_9 = 9$	135.531	0,001	0,999
$x_{10} = 10$	69.311	0,001	1
	$N = 13.678.341$		

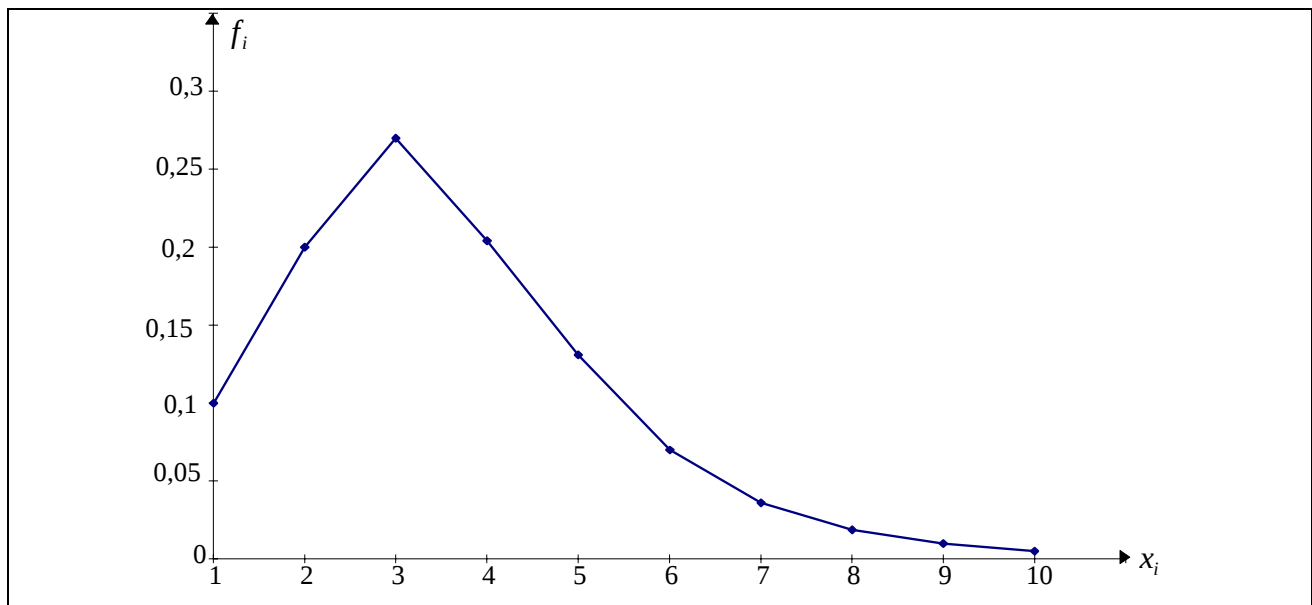


Diagramma di frequenza delle famiglie residenti per componenti al 1961

Funzione di ripartizione

La funzione che ad ogni modalità x_i della variabile statistica X fa corrispondere la frequenza cumulata F_i dicesi **funzione di ripartizione della variabile statistica X** . Essa rappresenta la percentuale F_i della popolazione N il cui carattere X ha modalità x non superiore alla

modalità x_i . In simboli matematici abbiamo :

$$F_i = \sum_{r=1}^i \frac{n_r}{N} = \frac{n_1}{N} + \frac{n_2}{N} + \dots + \frac{n_k}{N}$$

con $i = 1, 2, \dots, k$ se k sono le **modalità** del carattere X $x \leq x_i$

La **funzione di ripartizione di una variabile statistica X** è la funzione F_i che ad ogni modalità x_i

associa il numero $F_i = \sum_{r=1}^i \frac{n_r}{N}$, cioè :

$$F_i : x_i \rightarrow F_i = \sum_{r=1}^i \frac{n_r}{N}$$

Il grafico di una **funzione di ripartizione** si costruisce così :

- 1) in ascisse si riportano le **modalità** x_i del carattere X
- 2) in ordinata si riporta la corrispondente **frequenza relativa cumulata** di ogni modalità
- 3) nel piano cartesiano $x_i F_i$ otteniamo dei punti che si congiungono con dei segmenti a due a due consecutivi

4) La spezzata che si ottiene è il grafico della funzione di ripartizione e dicesi anche **ogiva di frequenza** (**ogiva di frequenza dal basso** ed **ogiva di frequenza dall'alto**)

Osservazione

Per costruire il grafico della funzione di ripartizione di una distribuzione continua (distribuzione di frequenza in classi) dove ogni modalità della variabile **X** non è espressa da un numero intero ma da un intervallo si procede come segue :

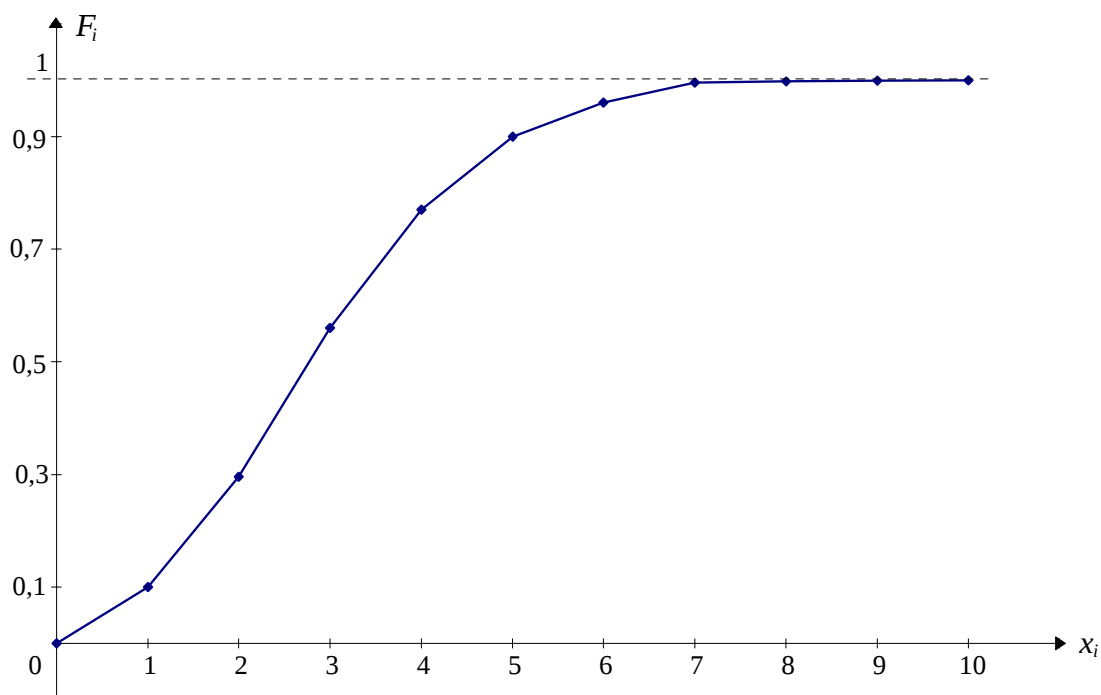
1) si considera come valore x_i il valore centrale della classe cioè la media aritmetica tra i

valori estremi , per cui come valore x_i si prende il valore $\frac{x_{i-1} + x_i}{2}$

2) poi si procede come prima .

In figura è rappresentato il grafico (**ogiva dal basso**) della **funzione di ripartizione della variabile statistica X** << **Famiglie residenti per numero di componenti al 1961** >>

Famiglie residenti per numero di componenti al 1961			
Ampiezza delle famiglie (numero componenti X)	Numero delle famiglie (frequenza assoluta n_i)	Frequenza relativa f_i	Frequenza Relativa Cumulata F_i
$x_1 = 1$	1.464.277	0,10	0,10
$x_2 = 2$	2.693.471	0,196	0,296
$x_3 = 3$	3.086.113	0,27	0,56
$x_4 = 4$	2.797.966	0,204	0,77
$x_5 = 5$	1.726.469	0,13	0,9
$x_6 = 6$	956.537	0,06	0,96
$x_7 = 7$	493.596	0,036	0,996
$x_8 = 8$	254.970	0,002	0,998
$x_9 = 9$	135.531	0,001	0,999
$x_{10} = 10$	69.311	0,001	1
	$N = 13.678.341$		



Funzione di ripartizione delle famiglie residenti per componenti al 1961

Distribuzione integrale di frequenza o distribuzioni cumulate di frequenza : ogiva di frequenza

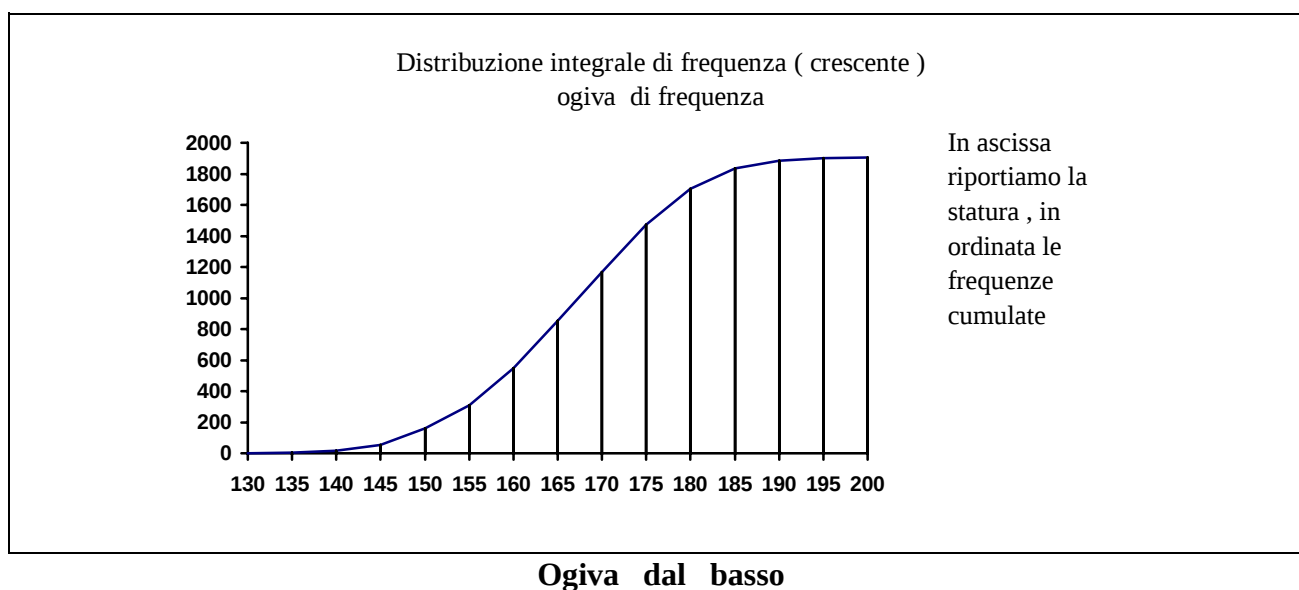
Dato che le distribuzioni di frequenza presentano , in corrispondenza di ciascun valore della variabile indipendente (nel caso di variabili discrete) o di ciascuna classe di valori (per variabili continue) , un determinato numero di osservazioni , potremo dire , considerato l'esempio della tabella , che 315 persone hanno una statura compresa fra 165 e 170 cm . Tuttavia raggruppando le frequenze in modo diverso è corretto parlare di 1.169 persone che hanno una statura non superiore a 170 cm .

Nell'esempio in esame si tratterà semplicemente di sommare progressivamente tutte le frequenze della prima classe fino alla classe da 165 a 170 .

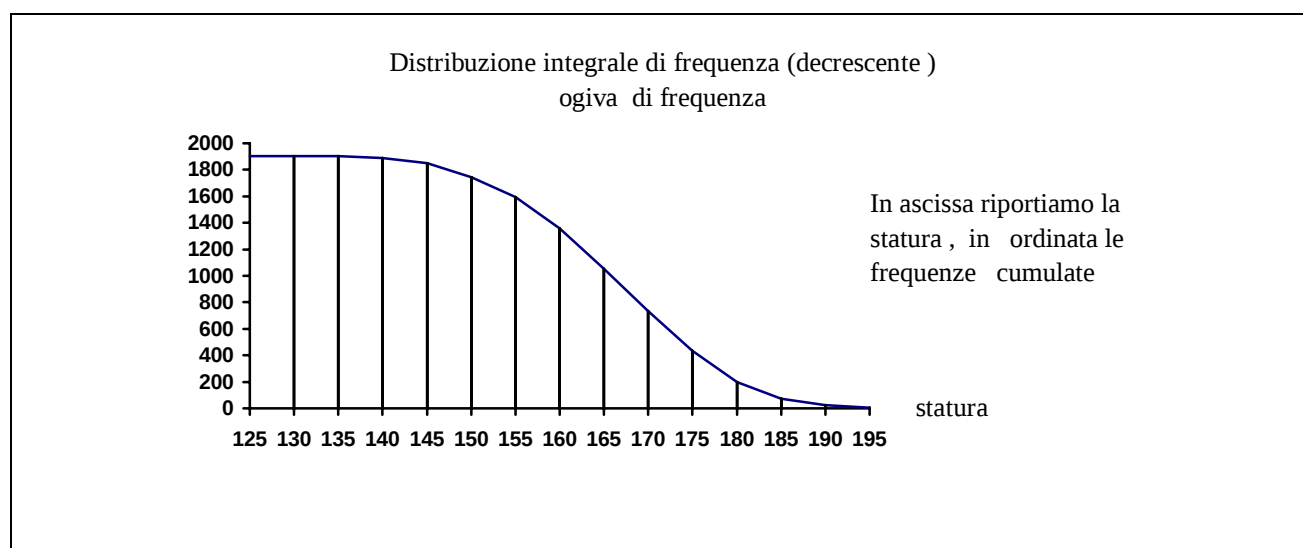
Si ottengono così le **distribuzioni integrali di frequenza** (o distribuzioni cumulate) in cui , invece di considerare le frequenze corrispondenti a ciascun valore o a ciascuna classe di valori , si esaminano i casi che presentano un valore della variabile non superiore o non inferiore ad un dato limite o , per variabili discontinue , ad un dato valore intero .

Se consideriamo i limiti superiori delle successive classi , e se le frequenze relative alla prima , seconda,...,n-esima sono $n_1, n_2, \dots, n_n$, la nuova successione sarà data da n_1 , $n_1 + n_2$, $n_1 + n_2 + n_3$ e così di seguito fino a $\sum_{i=1}^n n_i$ che rappresenta il totale delle frequenze .

La rappresentazione grafica della distribuzione integrale di frequenza è un diagramma , in cui si indicano in ascissa i limiti superiori o inferiori delle classi e in ordinata le corrispondenti frequenze cumulate . Congiungendo le estremità delle ordinate si ottiene una spezzata , che , con un numero di classi elevato , ha l'apparenza di una curva continua detta **ogiva di frequenza** .



Classi di statura $x_1 - x_2$	Frequenze n_i	Limite superiore x_2	Frequenze cumulate (fino ad x_2)	Limite inferiore x_1	Frequenze cumulate (oltre x_1)
125-130	1	130	1	125	1.905
130-135	3	135	4	130	1.905
135-140	11	140	15	135	1.901
140-145	38	145	53	140	1.890
145-150	106	150	159	145	1.852
150-155	152	155	311	150	1.746
155-160	238	160	549	155	1.594
160-165	305	165	854	160	1.356
165-170	315	170	1.169	165	1.051
170-175	302	175	1.471	170	736
175-180	234	180	1.705	175	434
180-185	128	185	1.833	180	200
185-190	50	190	1.883	185	72
190-195	19	195	1.902	190	22
195-200	3	200	1.905	195	3
Totale	1.905				



Ogiva dall'alto