

5. RAPPRESENTAZIONI GRAFICHE

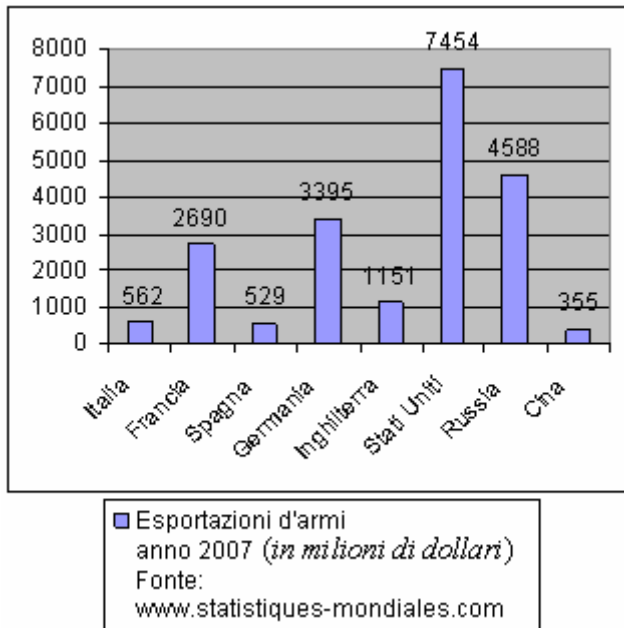
Di fronte a un insieme di dati,
indipendentemente dall'intenzione o meno di "fare della statistica" (= calcolare frequenze, medie ecc.),
è sovente assai utile far ricorso a rappresentazioni grafiche,
 per agevolare la lettura e l'interpretazione dei dati stessi.

Particolarmente utilizzati sono allo scopo

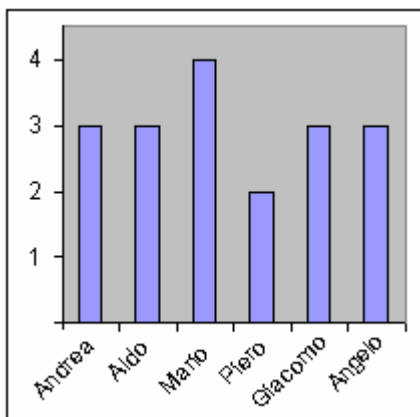
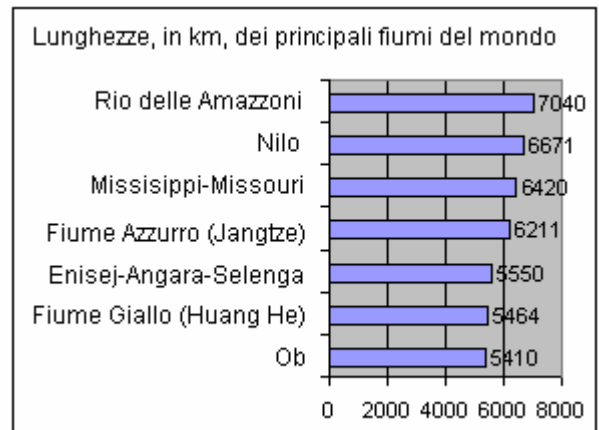
- **I DIAGRAMMI "A BARRE" O "A RETTANGOLI", DETTI ANCHE "ORTOGRAMMI"**
 (rettangoli verticali = "colonne", rettangoli orizzontali = "nastri")
- **E I GRAFICI CARTESIANI.**

Il loro impiego porta subito ad una visualizzazione molto efficace del fenomeno ...

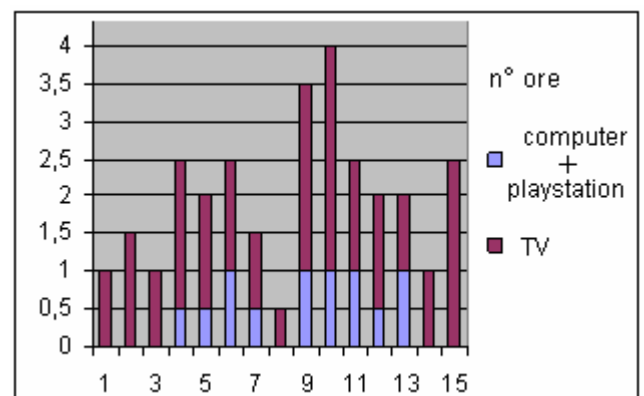
occorre però che I DATI NON SIANO TROPPO NUMEROSI, ALTRIMENTI "CI SI PERDE"!



*Il diagramma qui a sinistra è "a COLONNE",
 quello qui sotto è "a NASTRI".
 Entrambi sono diagrammi "a BARRE".
 In un foglio elettronico però,
 viene detto "diagramma a barre"
 solo quello che noi abbiamo chiamato "a nastri".
 Che pasticcio, a volte, la terminologia!*

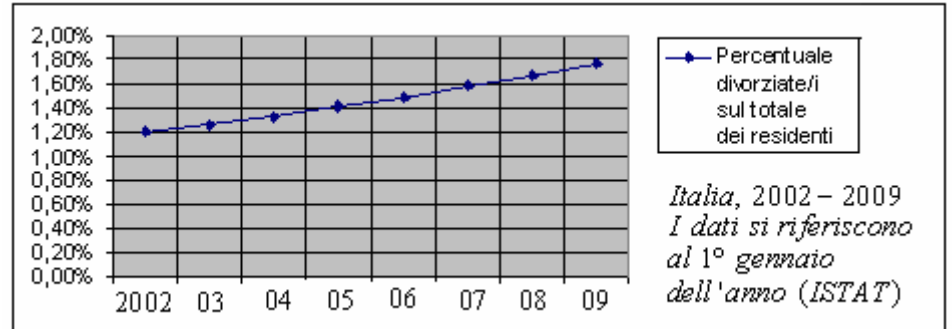


Grado
 di istruzione
 delle persone
 presenti
 al Bar Sport
 in un
 determinato
 giorno e ora
 1 = elementari
 2 = medie
 3 = diploma
 4 = laurea



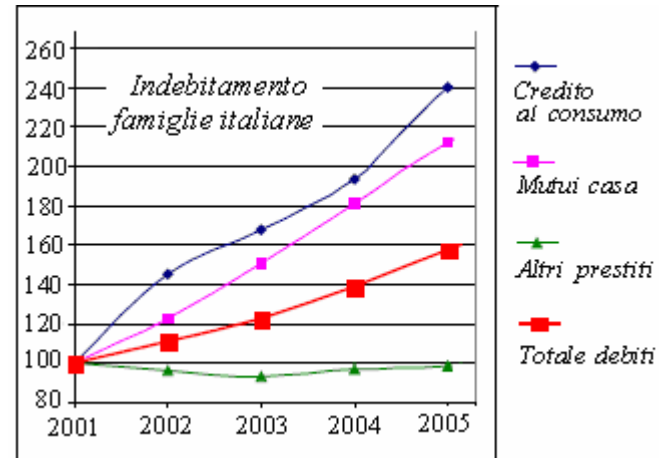
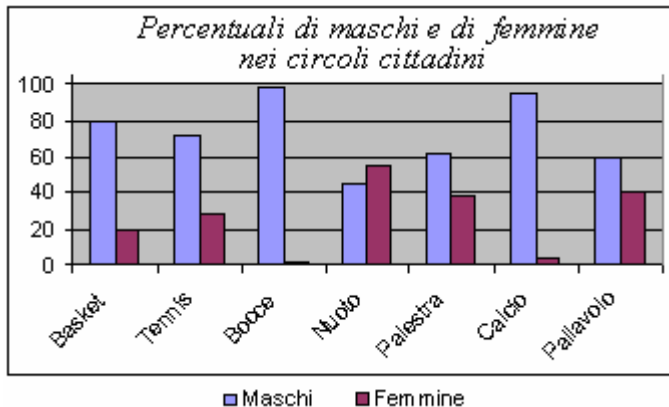
Permanenza media in una giornata davanti a un monitor
 (per i 15 ragazzi dell'Oratorio S. Giuseppe): barre in pila

Per una "**SERIE STORICA**",
 in cui uno stesso dato
 viene rilevato
 in tempi successivi,
 il grafico cartesiano
 può essere preferibile
 rispetto al diagramma
 a rettangoli,
 perché rende meglio l'idea
 dell'evolversi del fenomeno:

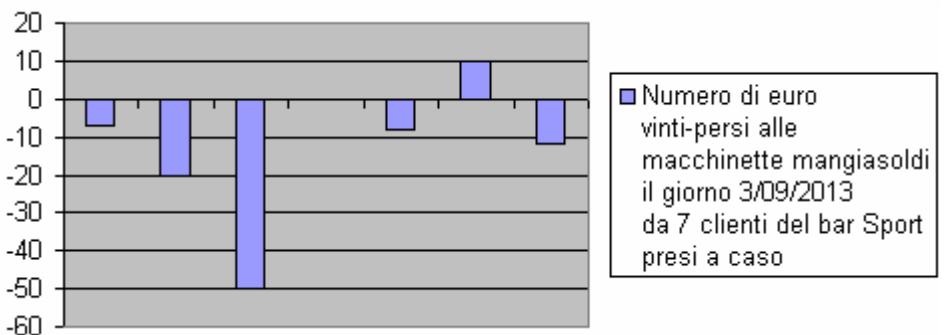


OSSERVAZIONE: le due parole “**DIAGRAMMA**” e “**GRAFICO**” sono in una certa misura intercambiabili, ma in generale sarebbe preferibile limitare l’uso della parola “**grafico**” ai soli casi in cui le quantità che vengono messe in relazione fra loro sono *tutte* espresse da **numeri**, e non da sostantivi/aggettivi/avverbi.

Tanto con un diagramma a rettangoli, quanto con un grafico cartesiano, è possibile anche confrontare *due o più* dati fra loro:

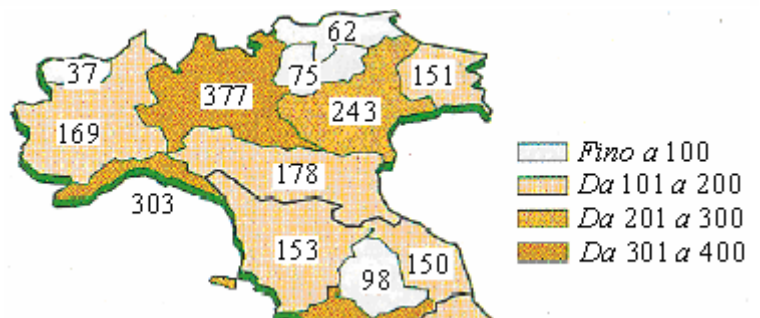


QUANDO I RETTANGOLI SCENDONO AL DI SOTTO DELL'ASSE ORIZZONTALE, esprimono un dato negativo:



In un **CARTOGRAMMA**, le diverse regioni di una cartina geografica vengono colorate con tinte più o meno scure a seconda dell'*intensità* del fenomeno in esame.

Nel cartogramma qui a fianco: intensità della popolazione residente (numero di abitanti per kilometro quadrato, anno 1997).
Fonte: ISTAT



ESERCIZI

1) Rileva le altezze in cm dei tuoi compagni di classe.

Con un “**foglio elettronico**”

(Excel o OpenOffice Calc, ad esempio:

trovi una elementare GUIDA al foglio elettronico nelle pagine successive)
visualizza la situazione.

Tieni presente che

- Excel** chiama “Istogrammi” i “diagrammi a colonne” (rettangoli verticali) e chiama “Barre” i “diagrammi a nastri” (rettangoli orizzontali)
- OpenOffice Calc** chiama rispettivamente “Colonna” un “diagramma a colonne” (rettangoli verticali) e “Barra” un “diagramma a nastri” (rettangoli orizzontali).

Comunque ... niente paura!

In un foglio elettronico, la denominazione è subito accompagnata dalla figura, quindi ... è facilissimo capire!

2) Con un “**foglio elettronico**” (Excel o OpenOffice Calc, ad esempio)
rappresenta la serie storica della popolazione italiana (in milioni di abitanti)
contenuta nella tabella seguente (dati ISTAT):

1901	1911	1921	1931	1936	1951	1961	1971	1981	1991
33,78	36,92	37,86	41,04	42,4	47,52	50,62	54,14	56,56	56,41

E veniamo ora a esaminare i principali tipi di rappresentazione grafica che si utilizzano più specificatamente per illustrare gli esiti di una vera e propria *indagine statistica* (c'è un "collettivo statistico", o "popolazione", e noi andiamo a rilevare qual è la *frequenza* - assoluta, relativa, o percentuale - con cui si riscontrano, nella "popolazione", le varie "modalità" di un determinato "carattere").

DIAGRAMMA A BARRE (= rettangoli verticali o orizzontali)

detto anche "**DIAGRAMMA A RETTANGOLI**" o "**ORTOGRAMMA**".

Se le barre sono **VERTICALI**, si potrà parlare di "**DIAGRAMMA A COLONNE**";

se **ORIZZONTALI**, di "**DIAGRAMMA A NASTRI**"

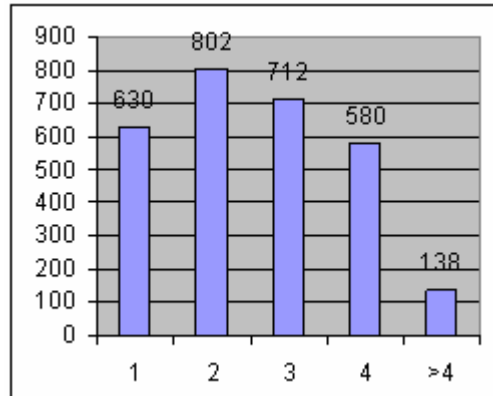
(come però abbiamo già fatto notare, **NEI "FOGLI ELETTRONICI"**

LA TERMINOLOGIA È DIVERSA DA QUESTA. Pazienza, è lo stesso, tanto si capisce ugualmente!)

E' una figura con rettangoli le cui basi sono fra loro uguali, e le cui altezze sono proporzionali alle frequenze (assolute, o relative, o percentuali).

*Indagine statistica
sulle 2862 famiglie di un Comune:
numero di componenti
del nucleo familiare
(conviventi nello stesso appartamento)*

1	630
2	802
3	712
4	580
>4	138
	2862



I rettangoli si possono eventualmente accostare fra loro.

La figura qui sotto mostra una "tabella composta", nella quale le modalità di uno stesso carattere sono riferite a più popolazioni.

Notare anche l'aspetto "3D" (= tridimensionale) del diagramma.

*"Trovei interessanti i talk show politici in TV"?
(Risposte in percentuale su un campione di 964
intervistati, suddivisi per livello di istruzione)*

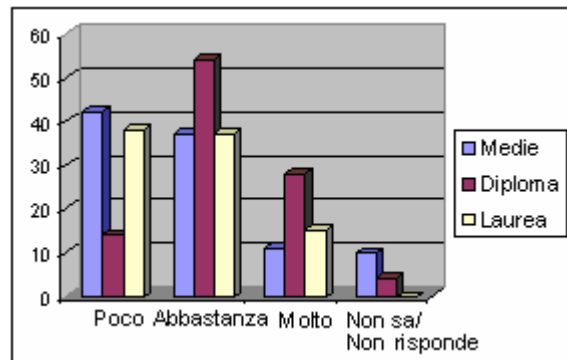
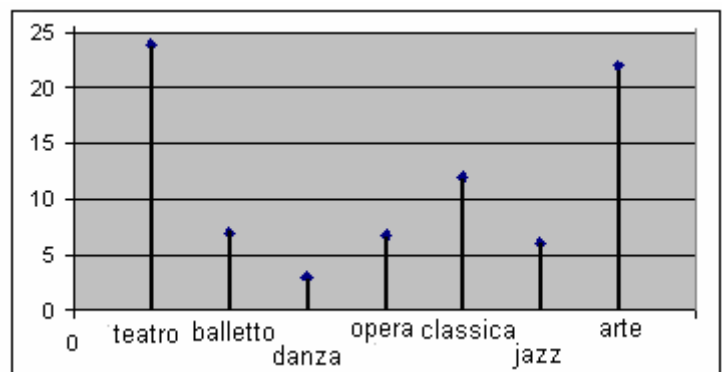


DIAGRAMMA A SEGMENTI (= AD ASTE),

analogo al diagramma a rettangoli, con una delle dimensioni del rettangolo sottilissima.

*Partecipazione ad eventi culturali:
percentuale della popolazione
che ha partecipato
ad almeno un evento culturale
del tipo specificato
(teatro, balletto, danza contemporanea,
opera, musica classica,
jazz, galleria d'arte)
nel periodo considerato.*



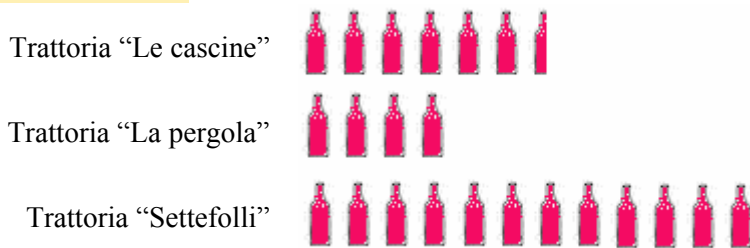
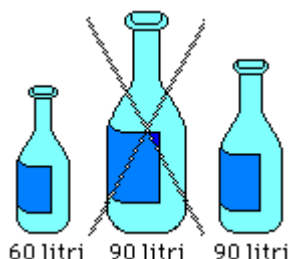
IDEOGRAMMA

Figure opportune, legate al contesto, vengono disegnate in modo che sia proporzionale al dato da rappresentare o il loro numero, oppure la loro estensione.

Nell'ideogramma qui a fianco:
Il consumo annuo di vino in 3 trattorie
(1 bottiglia = 500 litri)



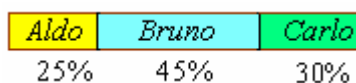
Attenzione, però:

se è l'estensione della figura quella che conta, è facile sbagliare:
ad esempio, nella figura qui a fianco, tratta da <http://macosa.dima.unige.it>,
la bottiglia centrale dovrebbe avere una volta e mezza il volume della prima
($90 = 60 \cdot 1,5$), mentre in realtà, essendo le sue dimensioni 1 volta e mezza,
il volume è $1,5 \cdot 1,5 \cdot 1,5 = 3,375$ volte tanto.

L'ideogramma corretto è la terza bottiglia, quella a destra!

DIAGRAMMA A STRISCE

Ottimo per confrontare le parti con il totale



Il contributo di 3 muratori
alla piastrellatura di un corridoio

DIAGRAMMA A TORTA (o “diagramma a settori circolari”)

Un cerchio è suddiviso in tante fette quante sono le modalità del carattere in esame.

L'angolo al centro di una fetta (ossia: di un settore circolare)
è proporzionale alla frequenza, assoluta o relativa o percentuale, della rispettiva modalità.

Come faccio a determinare di quanti gradi x° dev'essere una data “fetta”?

Semplice:

Se ad esempio le unità statistiche erano 24, e 10 di esse hanno presentato una certa modalità, allora

$$10 : 24 = x^\circ : 360^\circ \text{ da cui } x^\circ = \frac{10 \cdot 360^\circ}{24} = 150^\circ$$

In generale,

$$\frac{\text{frequenza assoluta}}{\text{numerosità}} : \text{numerosità} = x^\circ : 360^\circ \text{ da cui } x^\circ = \frac{\text{frequenza assoluta}}{\text{numerosità}} \cdot 360^\circ$$

(ricordiamo che per “numerosità” di una popolazione si intende il numero totale delle unità statistiche).

Si può anche operare (è lo stesso!), per determinare x° , con la frequenza relativa:

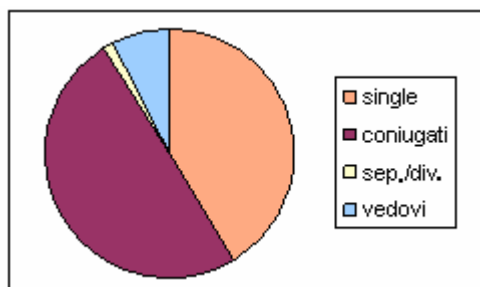
$$\frac{\text{frequenza relativa}}{\underbrace{\frac{\text{somma delle frequenze}}{\text{relative}}}_{=1}} = x^\circ : 360^\circ \text{ da cui } x^\circ = \text{frequenza relativa} \cdot 360^\circ$$

... oppure con la frequenza percentuale:

$$\frac{\text{frequenza percentuale}}{\underbrace{\frac{\text{somma delle frequenze}}{\text{percentuali}}}_{=100}} = x^\circ : 360^\circ \text{ da cui } x^\circ = \frac{\text{frequenza percentuale}}{100} \cdot 360^\circ$$

ESEMPIO: la popolazione (compresi i bambini) di un certo comune per stato civile nel 2013, in migliaia (single, coniugati, separati o divorziati, vedovi). Il fatto che nelle colonne delle frequenze relative e percentuali la somma non sia esattamente 1 o 100 è dovuto agli arrotondamenti.

	assoluta	relativa	percentuale	angolo
single	23516	0,414	41,4	149
coniugati	28185	0,496	49,6	179
sep./div.	854	0,015	1,5	5
vedovi	4224	0,074	7,4	27
TOTALE	56779	1,000	100,0	360



ISTOGRAMMA

(utilizzabile, nella versione “per aree” che qui sotto presentiamo, per i caratteri quantitativi continui; tuttavia, quasi sempre il termine “istogramma” viene inteso piuttosto come sinonimo di “diagramma a colonne”, il buon vecchio diagramma a colonne con basi delle colonne fra loro uguali. La raffigurazione “per aree” di cui stiamo per occuparci ha infatti sovente più svantaggi che vantaggi)

Un'azienda vuole illustrare la “ripartizione dei suoi dipendenti per classi di età”.

Poiché le età di 25 anni, di 45, e di 55, sono normalmente associate a scatti di carriera o comunque appaiono particolarmente adeguate a ripartire i dipendenti in gruppi in qualche modo omogenei (per atteggiamento mentale, per esperienza lavorativa ...), viene compilata la tabella seguente (s'intende, in ogni intervallo, il *primo estremo incluso* e il *secondo escluso*):

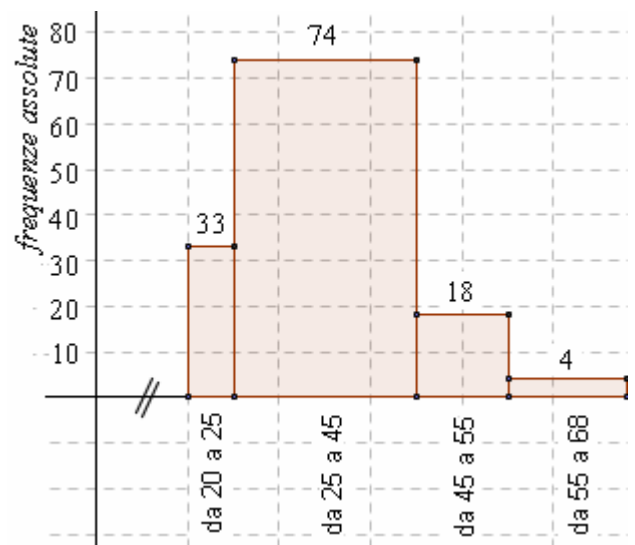
da 20 a 25	da 25 a 45	da 45 a 55	da 55 a 68
33	74	18	4

Le “classi” (= gli intervalli) sono fra loro differenti come ampiezza, per cui sembra opportuno che siano pure differenti fra loro (e proporzionali agli intervalli) le suddivisioni dell'asse delle ascisse.

Tuttavia, se a questo punto noi associassimo a ciascuna classe un rettangolo di altezza proporzionale alla frequenza ...

... questa rappresentazione potrebbe darci un **“colpo d'occhio” distorto** sulla situazione, per almeno due motivi:

- la nostra attenzione è portata spontaneamente a portarsi sull'*area* di ciascun rettangolo, piuttosto che sulla sua *altezza* ... ma da ciò si trarrebbe l'impressione (sbagliata!) che i dipendenti con almeno 45 anni (i 2 intervalli a destra) siano più numerosi di quelli con meno di 25 anni;
- e inoltre, si potrebbe pensare che l'altezza dell'intervallo si riferisca a ogni singolo valore che sta alla base dell'intervallo (quindi, per esempio, che ci siano 33 dipendenti di 20 anni, 33 di 21, 33 di 22, ...)



Viene allora un'altra idea.

Su ciascun intervallo si costruisce un rettangolo la cui **AREA** sia **proporzionale alla frequenza di quella classe**.

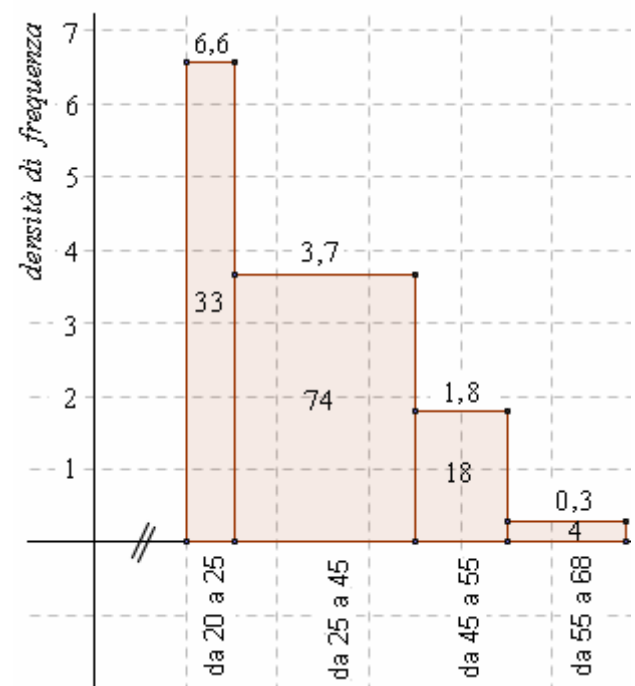
L'altezza del rettangolo si potrà perciò ricavare dividendo la frequenza per l'ampiezza della classe.

Nel nostro esempio, abbiamo ricavato l'altezza del primo rettangolo a sinistra dividendo la frequenza (che era 33) per l'ampiezza dell'intervallo ($25 - 20 = 5$).

Abbiamo ottenuto 6,6, quindi il nostro primo rettangolo a sinistra, avendo base 5 e altezza 6,6, avrà **area** uguale alla frequenza (33) con la quale la modalità “da 20 a 25 anni” si è manifestata nella nostra popolazione statistica (= l'insieme dei dipendenti dell'azienda).

Insomma, **in un istogramma di questo tipo le FREQUENZE non sono date dalle altezze dei rettangoli, bensì dalle loro AREE !**

E le altezze dei rettangoli vengono anche chiamate “densità di frequenza”.



QUALCHE OSSERVAZIONE sui diagrammi precedenti.

- La prima è banale: abbiamo utilizzato il simbolo $//$ per indicare il fatto che il segmento in gioco ha una lunghezza che non va d'accordo con le lunghezze degli altri intervalli sull'asse orizzontale: tale segmento non aveva importanza per il nostro diagramma, ed è stato tagliato per guadagnare spazio.

- Inoltre: il riferimento è, in entrambe le figure, “dimetrico”, cioè con due diverse unità di misura in orizzontale e in verticale.
D'altronde, sono addirittura diverse le due grandezze le cui misure vengono riportate sugli assi: in entrambi i diagrammi, l'asse orizzontale riporta intervalli di età, mentre sull'asse *verticale* abbiamo:

- nel primo dei due diagrammi, una *frequenza assoluta*;
- nel secondo, una “densità di frequenza”, la cui unità di misura ha la dimensione $età^{-1}$

- E' importante, quando si suddividono le modalità in intervalli (= in “classi di frequenza”), specificare con chiarezza se un estremo dell'intervallo è *incluso* oppure è *escluso*.

Noi lo abbiamo fatto dichiarando “s'intende, in ogni intervallo, il *primo estremo incluso* e il *secondo escluso*”.

A volte si indica l'inclusione o esclusione di un estremo in modo schematico: fra i simboli utilizzati, c'è

- $\vdash$ oppure $\bullet \text{---} \circ$ per indicare che il 1° estremo è incluso e il 2° è escluso: $x_1 \leq x < x_2$, $x \in [x_1, x_2)$
- $\dashv$ oppure $\circ \text{---} \bullet$ per indicare il viceversa: $x_1 < x \leq x_2$, $x \in (x_1, x_2]$
- H o anche solo un trattino (—) oppure $\bullet \text{---} \bullet$ per indicare l'inclusione di entrambi gli estremi: $x_1 \leq x \leq x_2$, $x \in [x_1, x_2]$

Poiché diversi Autori potrebbero effettuare scelte diverse, senza magari esplicitare la loro scelta con chiarezza, occorre sempre, quando si consulta una fonte, cercare di capire come si è regolato quel libro o quel sito.

- Nel caso particolare in cui le ampiezze degli intervalli siano tutte uguali, un istogramma non differisce da quello che noi avevamo denominato “diagramma a barre” o “ortogramma”.
- I “fogli elettronici” EXCEL, OPENOFFICE CALC non fanno istogrammi, ma solo **diagrammi a barre ...** **... che tuttavia chiamano “istogrammi”!** E questa abitudine a utilizzare il termine “istogramma” per indicare quelli che, per la precisione, andrebbero chiamati “diagrammi a barre”, o “ortogrammi”, o “diagrammi a rettangoli” è comunque entrata nell'uso anche nel linguaggio comune.

Puoi trovare una brevissima introduzione al “foglio elettronico” nelle pagine seguenti.



Va detto che **GLI ISTOGRAMMI CON INTERVALLI DI UGUALE AMPIEZZA (INDISTINGUIBILI QUINDI DAI “DIAGRAMMI A BARRE”)** SONO AMPIAMENTE PREFERIBILI, perché di più immediata interpretazione. **IL PREZZO DA PAGARE È CHE LA BASE DEL RETTANGOLO PUÒ NON ESSERE PROPORZIONALE ALL'AMPIEZZA DELLA CLASSE, MA ... PAZIENZA!**

Riguardo alle rappresentazioni grafiche, citiamo ancora le “**TABELLE A DOPPIA ENTRATA**”:

ottime per illustrare la distribuzione di due distinti caratteri su di una stessa popolazione.

Ad esempio: indagine, su 303 famiglie, riguardo a reddito annuo lordo, in euro, e numero di autovetture possedute.

Reddito \ n° auto	0	1	2 o più	Totale
<20000	23	27	0	50
20000 $\vdash$ 40000	5	84	18	107
40000 $\vdash$ 60000	0	70	29	99
≥ 60000	0	15	32	47
Totale	28	196	79	303

I TIPI DI RAPPRESENTAZIONE GRAFICA PIU' ... “GETTONATI”

Per i caratteri **QUALITATIVI SCONNESSI** i diagrammi più utilizzati sono:

- il diagramma a **torta** (specialmente se le modalità sono poche);
- il diagramma a **colonne** (che Excel chiama *istogramma*) o il diagramma a **nastri** (che Excel chiama *barre*)

Per i caratteri **QUALITATIVI ORDINATI** si utilizzano prevalentemente

il diagramma a **colonne** (che Excel chiama *istogramma*) o il diagramma a **nastri** (che Excel chiama *barre*)

Per i caratteri **QUANTITATIVI DISCRETI** si utilizza prevalentemente il diagramma ad **aste** (= **segmenti**)

Per i caratteri **QUANTITATIVI CONTINUI** si utilizza prevalentemente il diagramma a **colonne**

Per le **SERIE STORICHE**: grafico **cartesiano**, diagramma a **colonne**

Per le **SERIE GEOGRAFICHE**: **cartogramma**