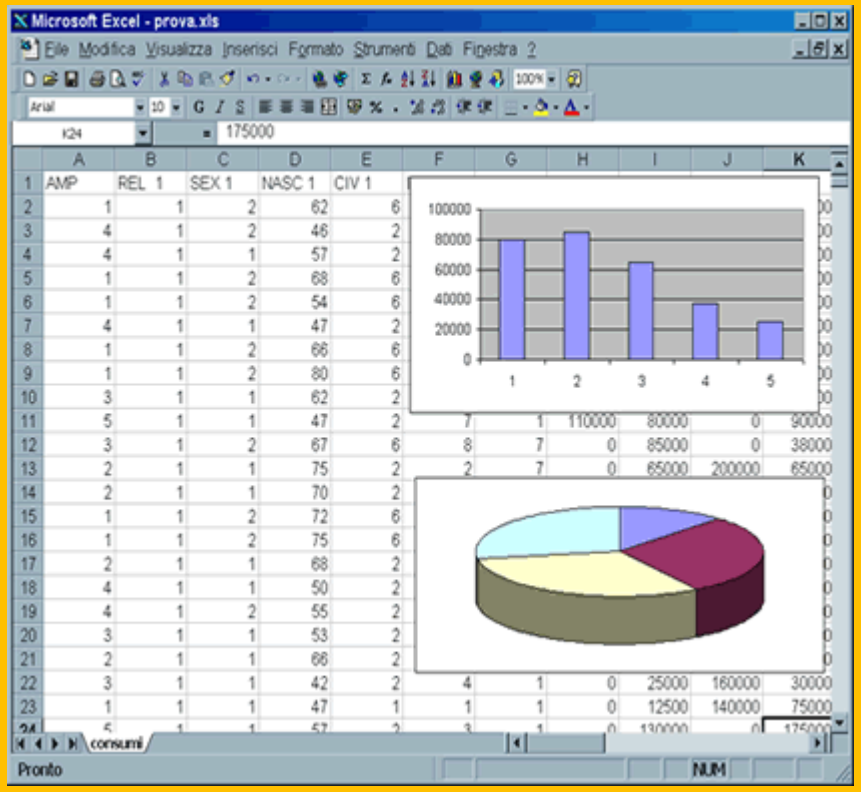


UNIVERSITA' DEGLI
STUDI DI PERUGIA

STATISTICA MEDICA

Prof.ssa Donatella Siepi
donatella.siepi@unipg.it
tel: 075 5853525


Microsoft Excel - prova.xls
File Modifica Visualizza Inserisci Formato Strumenti Dati Finestra ?
175000
AMP REL 1 SEX 1 NASC 1 CIV 1
consumi
Pronto
NUM

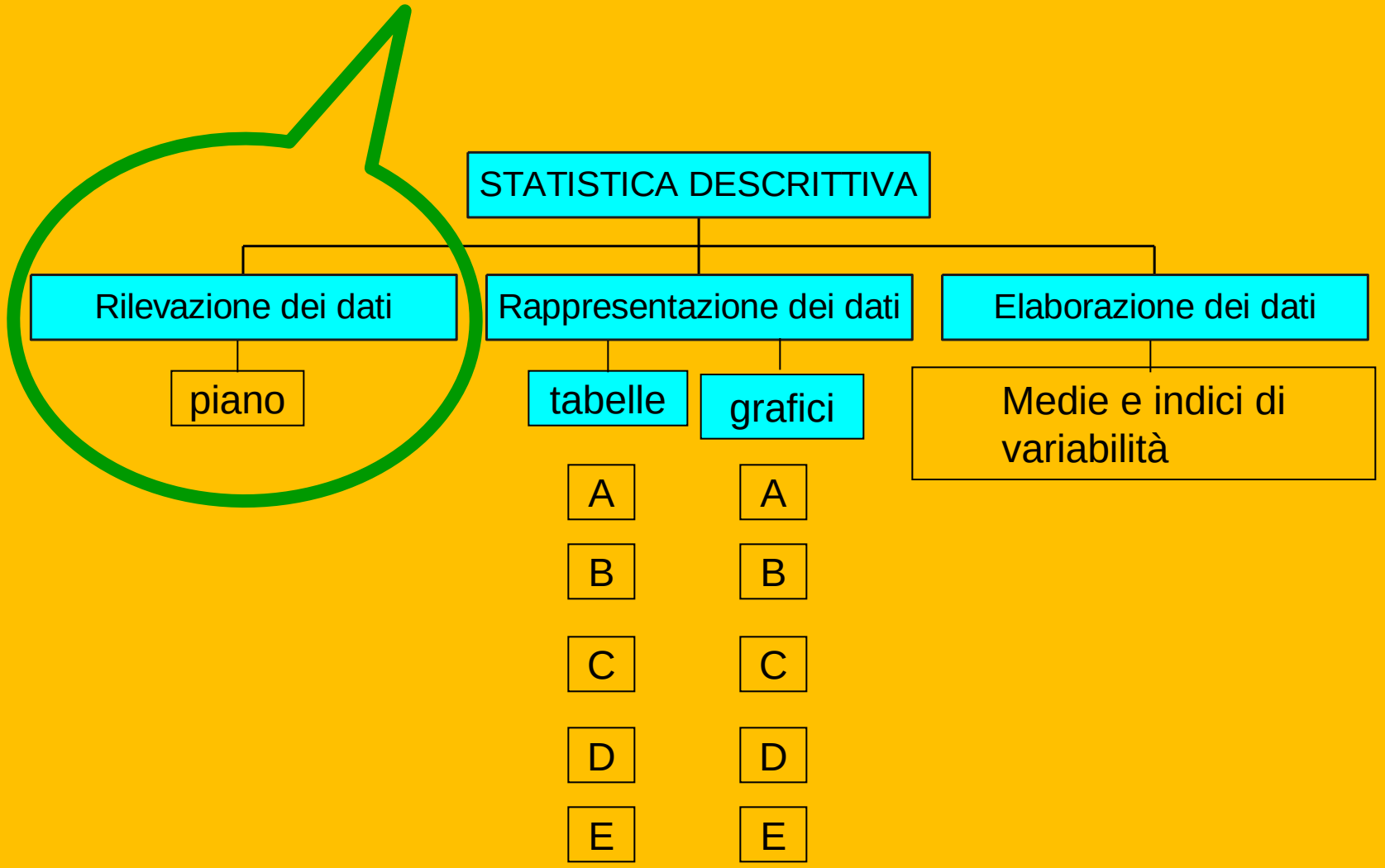
2° LEZIONE


125 m
100
75
50





Statistica descrittiva



PER OTTENERE QUESTO

- Per condurre in modo corretto una ricerca scientifica, cioè per raccogliere un campione con un numero sufficiente di dati, tenendo in considerazione sia le condizioni esistenti nella popolazione, sia la successiva applicazione dei test, occorre seguire alcuni passaggi metodologici, riassumibili in 4 fasi:
- **- il disegno sperimentale,**
- - il campionamento,
- - la descrizione statistica,
- - la scelta dei test per l’inferenza.

Disegno sperimentale

- 1 - Il **disegno sperimentale è necessario per scegliere e programmare le osservazioni in natura e** le ripetizioni in laboratorio, in funzione della ricerca e delle ipotesi esplicative. Già nella prima fase della ricerca, chiamata con termine tecnico appunto "disegno sperimentale" (***programmazione* dell'esperimento), occorre avere chiara la formulazione dell'ipotesi che si intende verificare.**
- Raccogliere i dati prima di aver chiaramente espresso le finalità della ricerca conduce spesso ad analisi non adeguate e quindi a risultati poco attendibili.
- Con la formulazione dell'ipotesi, si deve rispondere alle domande:
- - "*Le eventuali differenze riscontrate tra due o più gruppi di dati, oppure di una serie di osservazioni con quanto è atteso, possono essere imputabili a* ***fattori causali specifici o solamente*** *a* ***fattori casuali ignoti?***
- *- Le differenze riscontrate sono generate dalla naturale variabilità delle misure e del materiale utilizzato oppure più probabilmente esiste una causa specifica che le ha determinate?"*

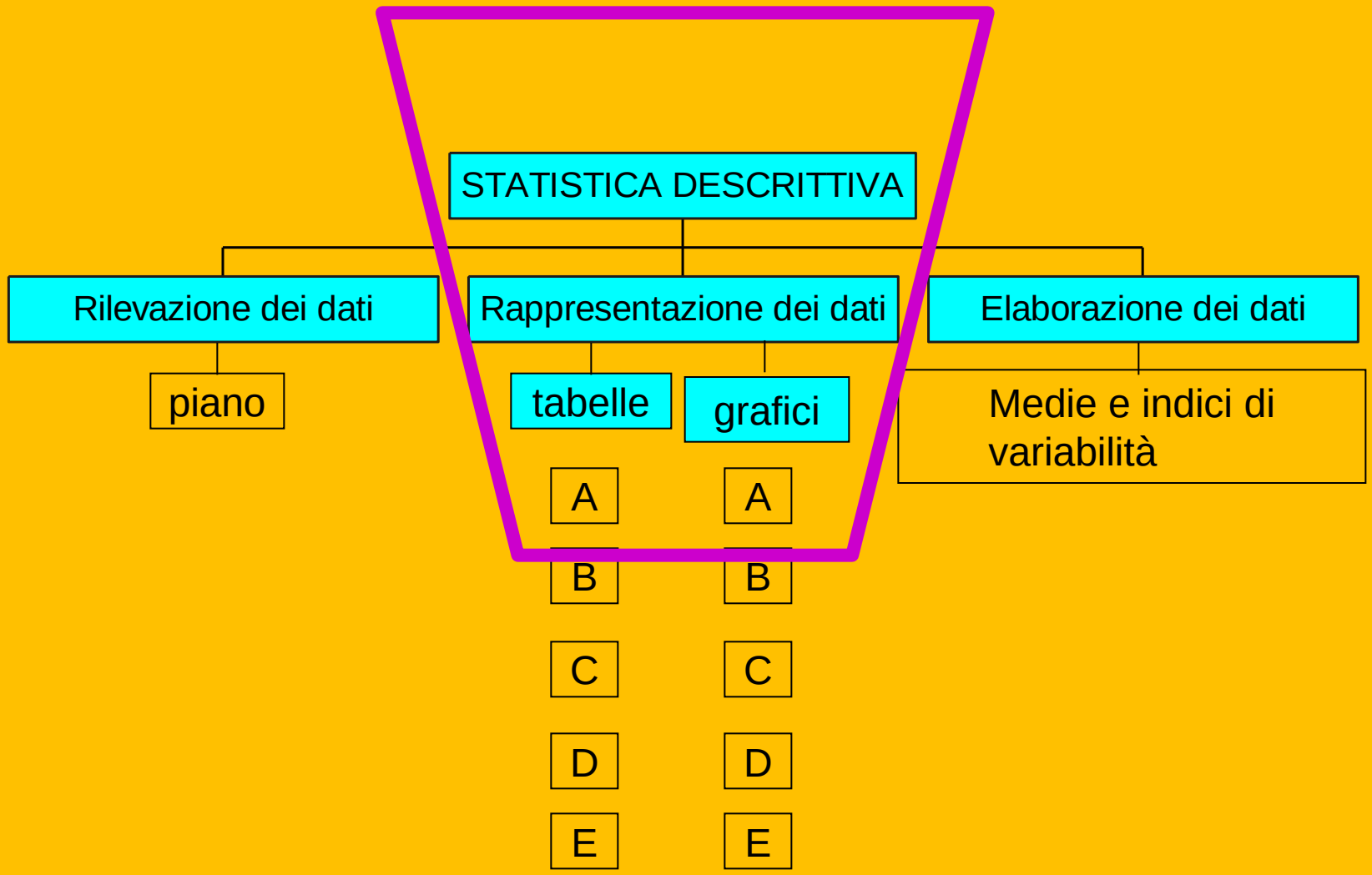
- Per condurre in modo corretto una ricerca scientifica, cioè per raccogliere un campione con un numero sufficiente di dati, tenendo in considerazione sia le condizioni esistenti nella popolazione, sia la successiva applicazione dei test, occorre seguire alcuni passaggi metodologici, riassumibili in 4 fasi:
- - il disegno sperimentale,
- **- il campionamento,**
- - la descrizione statistica,
- - la scelta dei test per l’inferenza.

Campionamento

- 2 - Il **campionamento permette di raccogliere i dati in funzione dello scopo della ricerca,** rispettando le caratteristiche della **popolazione o universo dei dati.**
- Uno dei **problemi fondamentali della statistica è come raccogliere solamente un numero limitato di dati (per motivi economici, di tempo, di oggetti effettivamente disponibili, cioè per** limiti oggettivi che quasi sempre esistono in qualsiasi ricerca sperimentale), **ma attraverso la loro analisi pervenire ugualmente a conclusioni generali, che possano essere estese a tutta la** popolazione.

- Per condurre in modo corretto una ricerca scientifica, cioè per raccogliere un campione con un numero sufficiente di dati, tenendo in considerazione sia le condizioni esistenti nella popolazione, sia la successiva applicazione dei test, occorre seguire alcuni passaggi metodologici, riassumibili in 4 fasi:
- - il disegno sperimentale,
- - il campionamento,
- - **la descrizione statistica**,
- - la scelta dei test per l’inferenza.

Statistica descrittiva



Puntualiziamo alcuni termini

- Un po’ di vocabolario



Definizioni

- ***Collettivo statistico (o popolazione): insieme di riferimento del fenomeno*** oggetto di studio (esempi: residenti nella regione Marche, insieme delle famiglie italiane, studenti dell'Università di Urbino).
- ***Campione statistico***. E' un sottoinsieme di una definita popolazione (universo), individuato in essa in modo da consentire, con margini di errori contenuti, la stima di determinati valori dell'intera popolazione (esempi: un certo numero di residenti nella regione Marche, delle famiglie italiane, degli studenti dell'Università di Urbino). .

Definizioni

- ***Campione statistico***. E' un sottoinsieme di una definita popolazione (universo), individuato in essa in modo da consentire, con margini di errori contenuti, la stima di determinati valori dell'intera popolazione. (esempi: un certo numero di residenti nella regione Marche, delle famiglie italiane, degli studenti dell'Università di Urbino)
- ***Unità statistica: singolo caso individuale che compone il collettivo statistico***

  (esempi: ogni residente della regione Marche, ogni famiglia italiana, ogni studente dell'Università di Urbino).

Definizioni

- ***Carattere: ogni caratteristica elementare oggetto di rilevazione presso le unità***
  - statistiche che formano il collettivo (esempi: età, sesso, titolo di studio, peso).
- ***Modalità di un carattere: diversi modi in cui il carattere si manifesta;***
  - modalità del carattere *sesso: M o F;*
  - modalità del carattere *peso: 65Kg., 82Kg.*

- ## Variabili qualitative o categoriche

  sono quantificate con **conteggi, ossia con numeri interi e discreti. Ad esempio, per valutare gli effetti di un tossico è possibile contare quante cavie** muoiono o sopravvivono; con un farmaco, quanti pazienti guariscono o restano ammalati, entro un tempo prefissato; con esperimenti sulle leggi di Mendel, quante piante hanno fiori rossi o bianchi.

- ## Variabili quantitative o Numeriche

  richiedono **risposte numeriche, espresse su una scala continua. Ad** esempio, per un'analisi del dimorfismo animale, dopo la separazione in maschi e femmine, si possono misurare il peso e l'altezza di ogni individuo.

Scale di misurazioni

- Variabili categoriche
  - Nominali (sesso)
  - Ordinali numeri ordinati in senso crescente o decrescente
- Variabili Numeriche
  - Discrete (numeri interi che si possono ricavare da operazioni o conteggi)
  - Continue (tutti i valori possibili di una scala: Temperatura ecc)

SCALE di misura

- Una **scala di misura** è quindi una funzione che mette in corrispondenza gli stati di oggetti su determinate proprietà e i numeri reali. I numeri, o simboli, che costituiscono il sistema numerico possiedono proprietà diverse. In questo modo, possiamo distinguere tra diversi livelli di scale di misura.
- Stevens parla di livelli di scala poiché i quattro tipi da lui distinti stanno in una precisa gerarchia, la prima rappresenta il livello più basso della misurazione, l'ultima è invece il livello più alto

SCALE di misurazioni

- Le **misure possono essere raggruppate in 4 tipi di scale, che godono di proprietà formali**
- differenti; di conseguenza, esse ammettono operazioni differenti. Come per tutte le discipline, **una scala di misurazione dei fenomeni biologici ed ambientali può essere:**
- 1) **nominale o classificatoria;**
- 2) **ordinale o per ranghi;**
- 3) ad **intervalli;**
- 4) di **rapporti.**

Scale di misura

- **scala nominale:** È il livello più basso di misurazione, sola relazione di identità
- sono dati di tipo classificatorio (es. gruppi sanguigni, tipo di patologia, ecc); la scala nominale piu' semplice e' quella dicotomica, cioe' quella che prevede solo due possibilita' (vivi o morti, maschi o femmine, con o senza un certo attributo, ecc);
- L'attribuzione di numeri per identificare categorie nominali, come avviene per individuare i giocatori nei giochi di squadra, è solamente un artificio che non può certamente autorizzare ad elaborare quei numeri come se fossero reali, ad esempio calcolandone la media.
- L'**operazione ammessa è il conteggio degli individui o dei dati presenti in ogni categoria.**
- I **quesiti statistici che possono essere posti correttamente riguardano le frequenze, sia assolute che relative.**

Scale di misura

- **scala ordinale:** o per **ranghi rappresenta una misurazione che contiene una quantità di** informazione immediatamente superiore a quella nominale, all'**equivalenza tra gli individui della stessa classe*, si aggiunge una* graduazione tra le classi o tra individui di classi differenti.**
- Prevede un ordine predeterminato, anche se non è prevista un'uguale distanza fra le modalità (es. indice di Apgar, punteggi di gravità di una malattia, ecc). La scala a ranghi e' quella che ordina gli elementi dal minore al maggiore assegnando i numeri d'ordine e trascurando le distanze tra gli elementi ordinati;
- Questa misura ha un limite fondamentale. **In una scala ordinale, non è possibile quantificare le differenze di intensità tra le osservazioni.**
- Alle variabili così misurate è possibile applicare una serie di **test statistici non parametrici; ma non quelli parametrici.**

Scale di misura

- **scala numerica ad intervalli: aggiunge la proprietà di misurare le distanze o differenze tra tutte le coppie di valori.**
- La scala ad intervalli misura quei sistemi empirici che, oltre a possedere le caratteristiche rappresentate dalle scale nominali e ordinali, consentono di definire degli intervalli costanti e uniformi tra le intensità della proprietà misurata; in una scala ad intervalli, l'intervallo ha il valore "1″ e viene denominato unità di misura.

Scale di misura

- **scala numerica ad intervalli: aggiunge la proprietà di misurare le distanze o differenze tra tutte le coppie di valori.**
- La differenza tra livello di scala ordinale e ad intervalli consiste nel fatto che con una scala ad intervalli sappiamo non soltanto se, sulla base di quella variabile, una persona si colloca più in alto rispetto ad un'altra, ma anche "di quante unità di misura più in alto" si colloca.
- **In particolare:**
- Come per la scala nominale, è possibile stabilire se due modalità sono uguali o diverse (30 ≠ 20); come per la scala ordinale è possibile mettere due modalità in una relazione d'ordine (30 > 20); in più è possibile definire una unità di misura per cui è possibile dire che tra 30 e 20 c'è una differenza di 10 (30- 20).

Caratteristiche della scala ad intervalli equivalenti

- **La scala a intervalli** consente di effettuare operazioni algebriche basate sulla differenza tra i numeri associati ai diversi punti della scala.
- Agli elementi del sistema empirico vengono assegnati dei numeri tali per cui le differenze tra i numeri riflettono le differenze d'intensità tra le modalità della proprietà. E' possibile stabilire se, per esempio, la distanza tra il caso a e b è la metà della distanza tra io casi c e d, oppure che la distanza tra gli individui a e b è uguale alla distanza tra gli individui c e d.

Caratteristiche della scala ad intervalli equivalenti

- **La scala ad intervalli non ha uno zero assoluto** e non è possibile dire se un valore sia in relazione di multiplo o di rapporto rispetto ad un altro, non è possibile, quindi, stabilire i rapporti diretti tra le misure ottenute, ad esempio, non è possibile stabilire se il caso a possiede la proprietà misurata in quantità doppia rispetto al caso b questo perché lo zero non viene attribuito all'elemento del sistema empirico in cui la proprietà misurata risulta assente lo zero della scala è, cioè, arbitrario (non reale). **Nella scala ad intervalli la posizione dello zero è arbitraria**, non corrisponde all'intensità nulla della proprietà misurata e cambia nel passare da una scala ad un'altra, entrambe misura dello stesso sistema empirico.

Scale di misura

- **scala di rapporti: ha il vantaggio di avere un'origine reale. Sono tipiche scale di** rapporti l'altezza, la distanza, la velocità, l'età, il peso, il reddito, la temperatura in gradi Kelvin; più in generale, tutte quelle misure in cui **0 (zero) significa quantità nulla.**
- Alle variabili misurate con una **scala di rapporti, cioè il tipo di misurazione più sofisticato e completo, può essere applicato qualsiasi test statistico.**

Scala di rapporti

- La scala a rapporti equivalenti è simile alla scala ad intervalli. Le due scale differiscono solo per il diverso significato che lo zero possiede nei due tipi di scala: zero relativo (scala ad intervalli equivalenti) o zero assoluto (scala a rapporti equivalenti).
- **Nella scala a rapporti equivalenti la posizione dello zero non è arbitraria** dato che corrisponde all'elemento dotato di intensità nulla rispetto alla proprietà misurata. Lo zero non viene assegnato ad elementi diversi nel passare da una scala ad un'altra, entrambe misura dello stesso sistema empirico.

Scala di rapporti

- Agli elementi di un insieme empirico vengono assegnati dei numeri tali per cui le differenze e i rapporti tra i numeri riflettono le differenze e i rapporti tra le intensità della proprietà misurata.
- Le operazioni aritmetiche sono dunque possibili sia sulle differenze tra i valori della scala (come per la scala a intervalli equivalenti) sia sui valori stessi della scala. L'unica arbitrarietà riguarda l'unità di misura che si utilizza: l'unità di misura può cambiare, ma qualsiasi unità di misura si scelga, lo zero indicherà sempre l'intensità nulla della proprietà considerata (ovvero, l'assenza della proprietà).

##

Tavola riassuntiva delle caratteristiche delle scale di misura Stevens

	Nominale	**Ordinale**	**Intervalli**	**Rapporti**
Tipo di variabile	Qualitativa	Qualitativa rettilinea o ciclica	Quantitativa	Quantitativa
Proprietà	Classificazione Non c'è ordinamento tra le modalità	Ordinamento Le modalità presentano un ordinamento	Conservazione delle relazioni tra le distanze Esiste un unità di misura	Costanza del rapporto tra i valori utilizzati Zero non Arbitrario
Operazioni matematiche di base	Uguaglianza /Disuguaglianza =;	Maggiore di / Minore di >; <; =;	Somma e Sottrazione +; -	Moltiplicazione e Divisione
Indicatori statistici	Numero dei casi Frequenze Moda	Mediana Percentili Quartili	Media Varianza Correlazione	Coeff. di Variazione Concentrazione

Riassumendo i concetti fondamentali esposti

- nella **scala nominale, esistono solo relazioni di *equivalenza;***
- in quella **ordinale, alla precedente si aggiungono relazioni di *minore o maggiore di;***
- in quella **ad intervalli alle due precedenti si aggiunge la relazione di *rapporto tra ogni coppia d'intervalli;***
- nella **scala di rapporti si ha anche la quarta relazione di *rapporto conosciuto tra ogni coppia di valori.***

- Si possono avere numeri che apparentemente hanno le stesse caratteristiche, ma che in realtà richiedono elaborazioni diverse ed impongono il ricorso a test differenti, per rispondere ai medesimi quesiti.
- Per esempio, **i grammi** di una determinata sostanza inquinante sciolta in un litro d'acqua, **la percentuale** di questa sostanza sul peso complessivo, **il punteggio della qualità dell'acqua** determinata dalla presenza di quella sostanza sono misure che utilizzano scale diverse.
- - **Nel primo caso**, si ha una classica scala di rapporti ed è possibile usare test parametrici, se la distribuzione dei dati è normale;
- - **nel secondo caso**, è possibile utilizzare le stesse procedure statistiche e gli stessi test parametrici, solamente dopo apposita trasformazione dei valori;
- - **nel terzo**, si ha una scala di ranghi, poiché la reale informazione fornita da questa serie di punteggi è solo quella di una graduatoria della qualità, nella quale non hanno reale significato né i rapporti né le differenze tra loro.

Presentazione dei dati

- **Un insieme di misure è detto serie statistica o serie dei dati.**

Peso 100 neonati

3,2	3,0	2,9	2,8	3,1	3,1	3,1	3,1	3,1	3,4
2,9	3,4	2,5	3,0	3,5	3,1	3,0	3,2	3,6	3,3
3,1	3,3	3,1	3,4	2,9	3,2	2,8	3,3	3,2	2,8
2,9	2,6	3,0	3,1	3,1	3,0	3,0	2,9	2,9	3,1
2,8	3,1	3,3	3,0	2,8	3,1	3,5	3,2	3,3	2,8
3,0	3,3	3,1	2,7	3,0	3,2	3,6	2,9	2,9	3,4
3,3	3,1	3,1	3,2	2,9	3,0	3,4	2,7	3,2	3,5
3,4	3,2	3,0	2,6	3,3	3,5	3,1	2,8	3,4	3,0
2,7	3,0	3,2	3,2	2,9	3,1	2,7	3,1	3,0	3,3
3,3	3,0	2,9	3,2	3,3	3,2	3,7	3,2	3,1	3,2

Tabelle di frequenza

- Una sua prima ed elementare elaborazione può essere una distribuzione ordinata di tutti i valori, in modo crescente o decrescente, detta **seriazione.**

- Il **valore minimo e il valore massimo insieme permettono di individuare immediatamente il campo (od intervallo) di variazione.**

Tabelle di frequenza

2,5	2,6	2,6	2,7	2,7	2,7	2,7	2,8	2,8	2,8
2,8	2,8	2,8	2,8	2,9	2,9	2,9	2,9	2,9	2,9
2,9	2,9	2,9	2,9	2,9	3,0	3,0	3,0	3,0	3,0
3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0	3,0
3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1
3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1	3,1
3,2	3,2	3,2	3,2	3,2	3,2	3,2	3,2	3,2	3,2
3,2	3,2	3,2	3,2	3,2	3,3	3,3	3,3	3,3	3,3
3,3	3,3	3,3	3,3	3,3	3,3	3,4	3,4	3,4	3,4
3,4	3,4	3,4	3,5	3,5	3,5	3,5	3,6	3,6	3,7

Presentazioni dei dati

- Tabelle di frequenza
- Aggregazione di classi di frequenza
- Distribuzione e frequenza delle osservazioni di conteggio
- Dati bivariati

- Successivamente, la serie può essere raggruppata in classi, contando quanti valori od **unità statistiche appartengono ad ogni gruppo o categoria.**
- Si ottiene una **distribuzione di frequenza o di intensità, detta anche semplicemente distribuzione.**

Peso 100 neonati

2,45-2,54	2,5	I	1
2,55-2,64	2,6	II	2
2,65-2,74	2,7	IIII	4
2,75-2,84	2,8	IIIIIII	7
2,85-2,94	2,9	IIIIIIIIII	11
2,95-3,04	3,0	IIIIIIIIIIIIII	15
3,05-3,14	3,1	IIIIIIIIIIIIIIIIII	20
3,15-3,24	3,2	IIIIIIIIIIIIII	15
3,25-3,34	3,3	IIIIIIIIII	11
3,35-3,44	3,4	IIIIIII	7
3,45-3,54	3,5	IIII	4
3,55-3,64	3,6	II	2
3,65-3,74	3,7	I	1

- ***Frequenza assoluta*** è il numero di individui il cui carattere assume un determinato valore.
- ***Frequenza relativa*** è il rapporto tra la frequenza assoluta e la totalità della popolazione statistica su cui si sta svolgendo l'indagine.
- ***Frequenza percentuale*** è semplicemente la frequenza relativa espressa in termini percentuali.

La trasformazione da frequenza assoluta a frequenza relativa risulta utile quando si vogliono confrontare due o più distribuzioni, che hanno un differente numero complessivo di osservazioni.

La frequenza cumulata offre informazioni importanti quando si intende stimare il numero totale di osservazioni inferiore (o superiore) ad un valore prefissato.

La distribuzione dei dati e la distribuzione delle frequenze cumulate forniscono informazioni **non dissimili**, essendo possibile passare con facilità dall'una all'altra. **Sono diverse nella loro forma**, come si vedrà con maggiore evidenza nelle rappresentazioni grafiche.

La prima ha una forma a campana, la seconda una forma a S, di tipo asintotico; si prestano ad analisi differenti e la scelta è fatta sulla base del loro uso statistico.

Frequenze

Voti	Allievi (frequenza)	Frequenza relativa	Frequenza relativa %
4	*2*	*0.09*	*9*
5	*4*	*0.18*	*18*
6	*8*	*0.36*	*36*
7	*5*	*0.23*	*23*
8	*3*	*0.14*	*14*
Totale	***22***	***1***	***100***

$2 / 22 \approx 0.09$
$4 / 22 \approx 0.18$
….

- **Frequenza relativa, o frequenza** di un evento è il rapporto fra la sua frequenza assoluta e il numero totale delle unità rilevate, se moltiplicata per 100 è detta *frequenza relativa percentuale*