Variabilità.

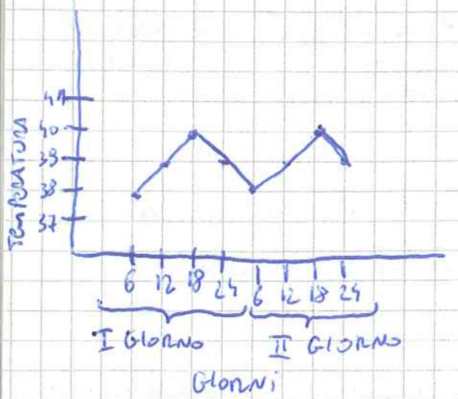
- Fenomeni che si modificano in relazione a differenti avvenimenti (variabili)
- In natura non esistono fenomeni costanti
  → sul piano ~~concettuale~~ concettuale
- Si definisce variabilità la proprietà di alcuni fenomeni di assumere valori (fenomeni quantitativi) o modalità (fenomeni qualitativi) diverse
- I fenomeni che posseggono tali proprietà si definiscono variabili, quelli che non le posseggono si definiscono costanti. Tutti i fenomeni biologici sono variabili
- La variabilità può essere studiata a livello individuale (nel singolo individuo) o collettivo (in un determinato collettivo)
  → in periodi diversi
  → nello stesso periodo
- Inoltre si può studiare la forma della variabilità e la misura della variabilità.
- Per studiare la forma della variabilità si ricorre alla rappresentazione grafica dei fenomeni (diagrammi). Le caratteristiche delle linee (curve) che rappresentano i fenomeni esprimono la forma della variabilità.

ANALISI STATISTICA SI BASA SU 2 PRINCIPI:

1) LA VARIABILITÀ DEI FENOMENI BIOLOGICI

2) LA TEORIA PROBABILISTICA

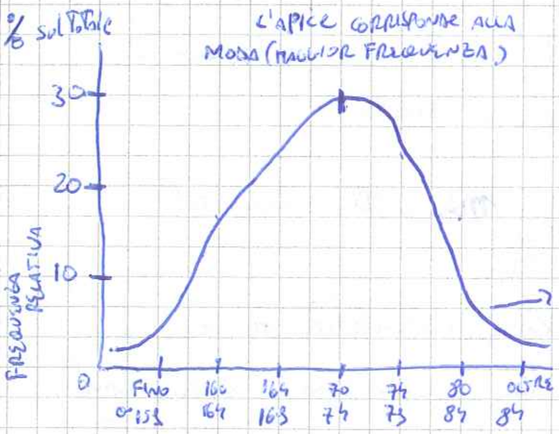
Forme della variabilità nell'individuo o distribuzione di valori



CURVA DI DISTRIBUZIONE DEI VALORI

GRAFICO DELLA TEMPERATURA FEBBRILE

Forme della variabilità nel collettivo



L'APICE CORRISPONDE ALLA MODA (MAGGIOR FREQUENZA)

CURVA DI GAUSS O CURVA NORMALE

STATURA DEGLI ISCRITTI DI LEVA NATI NELL'ANNO '63

→ PUNTO DI FLESSO

→ LA CURVA CAMBIA ORIENTAMENTO

- curva caratteristica di tutti i fenomeni naturali
- curva simmetrica rispetto alla media
- la moda corrisponde alla mediana ed alla media aritmetica

Misura della variabilità

I DATI POI VENGONO SOTTOPOSTI AD ANALISI STATISTICA X ACCERTARNE LA VALIDITÀ ED IL SIGNIFICATO IN RAPPORTO AI FINI SPECIFICI DELL'INDAGINE

- lo scarto è la misura di variabilità

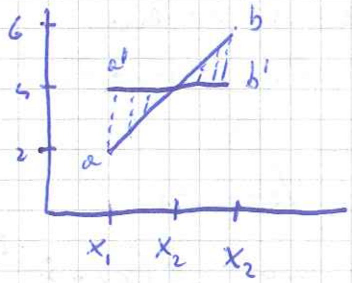
  ↳ DISTANZA FRA I SINGOLI VALORI E LA LORO MEDIA

- Maggiori sono gli scarti maggiore è la variabilità

- In una serie di valori si definiscono scarti (o scostamenti) le differenze tra i singoli valori della serie ed un valore di riferimento

SCARTO $x_i - \overline{x}$

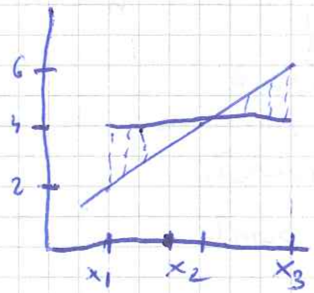
La somma algebrica degli scarti rispetto alla media aritmetica è sempre uguale a zero.



$2-4=-2$

$4-4=0$

$6-4=0$



$1-4=-3$

$4-4=0$

$7-4=3$

Misure di Variabilità

$(2-4)^2 = 4$	$(1-4)^2 = 9$
$(4-4)^2 = 0$	$(4-4)^2 = 0$
$(6-4)^2 = 4$	$(7-4)^2 = 9$
8	18

DEVIANZA

$$\sum (x_i - \bar{x})^2$$

somma dei quadrati degli scarti

- maggiore è la devianza maggiore è la variabilità della serie di valori
- il problema è che maggiore numero osservazioni, maggiore devianza anche se questo non significa una maggiore variabilità
- Allora si divide la devianza per il numero di osservazioni ed i valori diventano confrontabili

$$\frac{8}{3} = 2{,}666 \qquad \frac{18}{3} = 6$$

VARIANZA

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

al quadrato xché parliamo dei quadrati degli scarti

- inconveniente = se noi vogliamo esprimere la variabilità in unità di misura corrispondenti al fenomeno non corrispondono

⇒ si usa deviazione standard (o scarto quadratico medio)

$$\sqrt{2{,}66} = 1{,}63 \qquad \sqrt{6} = 2{,}44$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

→ radice quadrata varianza

- Tutta la ricerca scientifica si basa su queste misure di variabilità
- Mediante deviazione standard noi disponiamo di una formula molto semplice della variabilità di un fenomeno.

media aritmetica ± variazione standard

$$\bar{x} \pm \sigma$$

- A noi interessa sapere l'intervallo in cui è compreso un fenomeno. Questo intervallo rappresenta la variabilità del fenomeno
- Non è accettabile un limite di sicurezza inferiore al 95%

il Problema dei

- I valori medi non tengono conto della distribuzione di valori e per correggere questo si ricorre alla formula

$$\bar{x} \pm \sigma$$

VARIABILITÀ come

- Intervallo in cui sono compresi la maggioranza dei valori considerati.

N.B

- bisogna tener conto se si tratta di una distribuzione di valori o di frequenze

MISURA DELLA VARIABILITÀ (VALORI RAGGRUPPATI IN CLASSI)

STATURA IN CM	VALORI CENTRALI	N. OSSERVAZIONI	$x_i \cdot n_i$
FINO a 150	145	5	725
150 - 160	155	20	3100
160 - 170	165	50	8250
170 - 180	175	20	3500
OLTRE 180	185	5	925
		100	16500

$$\bar{x}_p = \frac{16500}{100} = 165$$

PONDERATA

- Vado a fare la diff. fra la media ed i singoli valori centrali

- Poi devo tener conto del numero osservazioni di ogni classe

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
145	-20	400	400 · 5 = 2000
155	-10	100	100 · 20 = 2000
165	-	-	-
175	+10	100	100 · 20 = 2000
185	+20	400	400 · 5 = 2000
			8000

DEVIANZA = 8000

DISPERSIONE SERIE DI VALORI

devo tener conto del numero delle osservazioni

$$\text{VARIANZA} = \frac{8000}{100} = 80$$

(NUMERO OSSERVAZIONI)

$$\text{Deviazione standard} = \sqrt{80} = 8{,}94$$

La deviazione standard del collettivo è di 8,94 cm

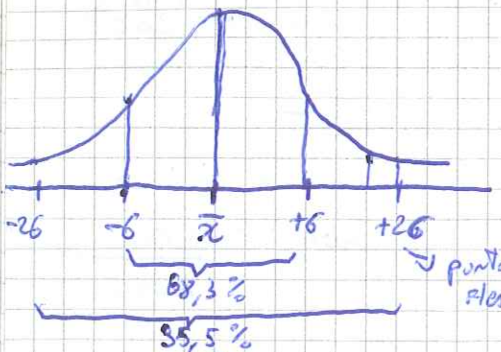
$$\bar{x} \pm \sigma = 165 \pm 8{,}94$$

$156 \longmapsto 174$

La statura del collettivo è compresa tra i 156 e i 174 cm

- Questi concetti trovano una particolare applicazione x i fenomeni che si distribuiscono su una curva gaussiana → Tutti i fenomeni biologici

Legge di Gauss - Laplace



Funzione della curva gaussiana

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}$$

y = singole frequenze corrispondenti ai valori della variabile

σ = deviazione standard (parametro)

π = costante

e = costante di Nepero 2,71828

x_i = singoli valori della variabile dei quali vogliamo conoscere le frequenze

$\bar{x}$ = media ricavata dal collettivo che stiamo esaminando (parametro)

- Se un fenomeno si distribuisce secondo una curva normale o gaussiana, aggiungendo e sottraendo alla media la deviazione standard ($\bar{x} \pm \sigma$) si ottengono 2 valori entro i quali è compreso il 68,3 % delle osservazioni; aggiungendo e sottraendo alla media il doppio della deviazione standard, si ottengono 2 valori entro i quali è compreso il 95,5 % delle osservazioni; aggiungendo e sottraendo alla media il triplo della deviazione standard si ottengono due valori entro i qual. è compreso il 99,7 % delle osservazioni.

$\bar{x} \pm \sigma$ = 68,26 % delle osservazioni
$\bar{x} \pm 2\sigma$ = 95,45 % "
$\bar{x} \pm 3\sigma$ = 99,73 % "

- Così si calcolano i limiti di normalità di un fenomeno.

- Quando studiamo la variabilità la studiamo su un campione:

→ limite = chi ci garantisce che quella è la variabilità dell'universo?

- c'è una diff. fra la variabilità del campione e dell'universo a causa dell'errore di campionamento

dobbiamo ricorrere ai:

GRADI di LIBERTÀ

metodo x ridurre diff. variabilità stimata ed osservata

- Quando si misura la variabilità in un campione e si applica tale misura (= variabilità osservata) all'intera popolazione dalla quale il campione è stato estratto (= variabilità stimata) si commette un errore più o meno grande a seconda della numerosità del campione
- Student ha dimostrato che per ridurre la differenza tra variabilità stimata e variabilità vera, occorre moltiplicare la varianza per un fattore di correzione avente al numeratore il numero delle osservazioni ed al denominatore il numero delle osservazioni −1

$$\sigma^2 = \frac{\sum(x_i-\bar{x})^2}{n}\cdot\frac{n}{n-1} = \sum(x_i-\bar{x})^2\cdot\frac{1}{n}\cdot\frac{n}{n-1} =$$

$$= \frac{\sum(x_i-\bar{x})^2}{n-1}$$

- gli addendi che uno può modificare liberamente si chiamano gradi di libertà
- addendi in questo caso sono le singole osservazioni. se tolgo un addendo devo togliere una osservazione
- Applicando questo metodo aumenta la devianza e cioè la variabilità → è proprio quello che serve
- Se allarghiamo la variabilità è + facile che la variabilità vera della popolazione sia compresa entro quella del campione.
- Se le osservazioni sono 4 il fattore di correzione sarà:

$$\frac{4}{3} = 1{,}33$$

- Se le osservazioni sono 30 il fattore di correzione sarà

$$\frac{30}{29} = 1{,}03$$

Quando il numero delle osservazioni è uguale o minore e 30 dobbiamo considerare i gradi di libertà. Se è superiore si può anche non applicare

Es. data la serie 2, 4, 6 la varianza sarà

$$\sigma^2 = \frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3} = \frac{8}{3} = 2{,}66\ldots;$$

$$\sigma = \sqrt{2{,}66} = 1{,}633$$

applicando i gradi di libertà si avrà

$$\sigma^2 = \frac{8}{3-1} = 4; \quad \sigma = \sqrt{4} = 2$$

Teoria della complessità (o del caos) di Rico Gin

- Fenomeni complessi = fenomeni che sono determinati da una molteplicità di fattori che interagiscono tra di loro o si ostacolano

- Curva di Gauss = Applicazione del fenomeno della complessità
  → si applica a tutti i fenomeni complessi (fenomeni biologici)

- Noi applichiamo la legge di Gauss anche quando il fenomeno non si distribuisce secondo una curva normale o Gaussiana.

- Errori non evitabili (casuali) → anche se non sono dovuti al caso

  → che non riusciamo ad identificare le cause

  → Laplace: Nulla in natura avviene per caso, ma ogni fenomeno risponde a leggi ben precise come quelle che regolano la rotazione della Terra intorno al sole

- La legge di Gauss si applica anche in presenza di errori non evitabili

- Legge di Gauss e Laplace = legge degli errori casuali (o naturali)

- Aspetti pratici = mediante questa legge possiamo intervenire sugli errori casuali

Legge degli errori casuali

1) la frequenza di errori in + o in − rispetto al valore vero è uguale.

2) la frequenza di errori è tanto maggiore quanto minore è la distanza rispetto al valore vero (gli errori casuali sono solitamente piccoli errori)

- A causa degli errori casuali (inevitabili) non siamo mai in grado di conoscere il valore vero

- la moda, media e mediana corrispondono al valore ~~[illegible]~~ che più si avvicina al valore vero

- la media di una serie di misure ripetute ad uno stesso soggetto è il valore che più si avvicina al valore vero.

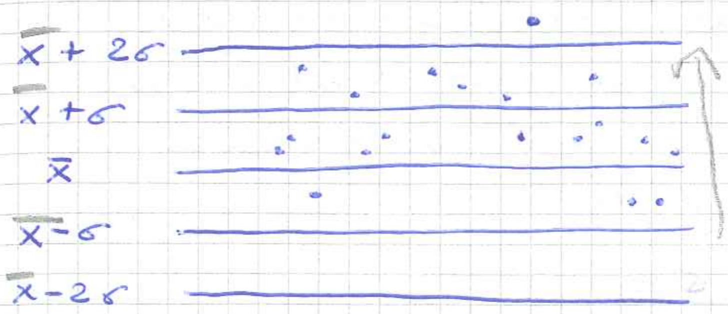
- Come posso distinguere tra ~~[illegible]~~ errori evitabili ed inevitabili?

Applicando la legge di Gauss-Laplace

- Se si tratta di errori casuali il 95% deve essere compreso nell'intervallo tra $\bar{x} \pm 2\sigma$

- Se troviamo che dei valori vanno al di sopra o al di sotto di questi limiti vuol dire che sono errori non casuali

Carta di Controllo per laboratori di Analisi



- Si costruisce esaminando un campione preciso ed ogni giorno si vanno a fare delle misurazioni sempre ~~[illegible]~~ sullo stesso campione

- In questo caso 1 solo errore su 20 va fuori di questi limiti → il 5%

→ è rispettata la legge di Gauss-Laplace

→ non ci sono errori evitabili, però il secondo postulato non è rispettato (perché le misure sono spostate tutte in un verso) quindi vuol dire che l'apparecchio sta per guastarsi.

- la legge di Gauss-Laplace si applica non soltanto ai singoli campioni casuali, ma si applica anche alla media dei campioni casuali estratti.

Errore standard della media

- Le medie di campioni casuali estratti da una popolazione si distribuiscono intorno alla media della popolazione secondo una curva normale (o gaussiana)

- La deviazione standard (= errore standard della media) delle medie dei campioni è data dal rapporto tra la deviazione standard e la radice quadrata del numero di osservazioni del campione

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

All'errore standard della media si applica la legge di Gauss-Laplace. Pertanto aggiungendo e sottraendo alla media del campione il doppio dell'errore standard si ottengono 2 valori entro i quali è compreso il 95% delle medie campionarie (e quindi anche la media della popolazione alla quale è stato estratto il campione)

$$\sigma_{\bar{x}} < \sigma_x$$

→ la curva di Gauss è più stretta e più alta
→ c'è una riduzione di variabilità

$$\bar{x} \pm 2\,\frac{\sigma_x}{\sqrt{n}}$$

media del campione

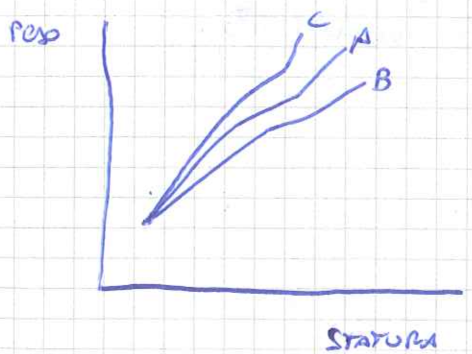
- Spesso ci interessa conoscere la variabilità di più fenomeni
- Spesso dobbiamo considerare la variabilità di più caratteri associati fra loro nella singola persona (es. peso/statura)

Crescita statura ponderale dalla nascita a 24 mesi

ETÀ	Peso A	Peso B	Peso C	STATURA
NASCITA	3500	3500	3500	50
6 mesi	7500	7000	8500	65
12 "	10000	9000	11000	75
18 "	11500	10000	12500	80
24 "	12500	11000	13800	85

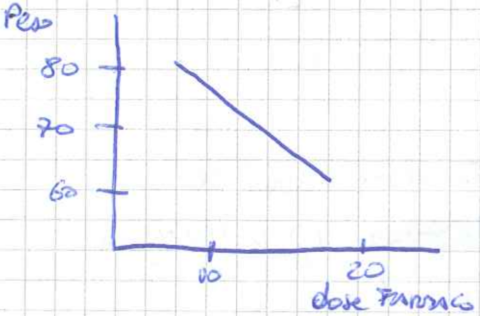
a) bambino con stato di nutrizione regolare
b) Bambino con stato di iponutrizione
c) " " " " ipernutrizione

- Per studiare la forma della variabilità uso un grafico



Rappresentazione grafica della relazione tra Peso e dose di un farmaco ad azione dimagrante

Dose farmaco (in mg)	Peso (in Kg)
5	80
10	75
15	70
20	65
25	60



- Quando le 2 variabili si modificano nello stesso senso abbiamo una linea inclinata verso l'alto e viceversa

Misura di variabilità di 2 serie di valori associati nei singoli termini

$2-4=-2$	$1-4=-3$	$(-2)(-3)=6$
$4-4=0$	$4-4=0$	
$6-4=2$	$7-4=3$	$2\cdot 3=6$
		12

↓

Prima della misura di variabilità associata ai 2 termini → CODEVIANZA

COVARIANZA $\frac{12}{3}=4$ → 3 osservazione (anche se possono sembrare 6 xké sono associate)

scarto negativo ↑	scarto positivo ↑	
$2-4=-2$	$7-4=3$	$(-2)(3)=-6$
$4-4=0$	$4-4=0$	
$6-4=2$	$1-4=-3$	$(2)(-3)=-6$
		-12

↓

codevianza negativa

- Agli scarti negativi della 1° serie corrispondono gli scarti positivi della 2° serie
- Quando le 2 variabili variano in senso opposto la codevianza è negativa e viceversa
- la covarianza ha lo stesso segno della codevianza

Codevianza = $\sum (x_i - \bar{x})(y_i - \bar{y})$

scarti 1° serie — scarti 2° serie

Sommatoria dei prodotti degli scarti associati o combinati delle 2 serie di valori rispetto alle relative medie

Covarianza = $\dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N} = \dfrac{Cd\,xy}{N}$

N = numero osservazioni

~~covarianza = somma~~

Codevianza = quando si vuole misurare la variabilità di 2 serie di valori associati nei singoli termini.

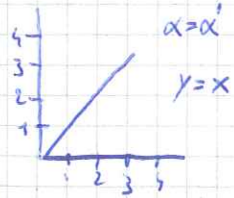
~~somme~~

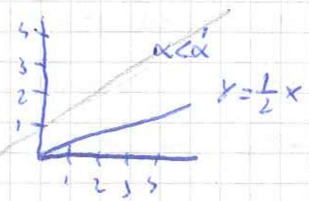
REGRESSIONE

- deve necessariamente esistere un rapporto di causa effetto
- Dobbiamo sapere quale delle 2 variabili determina la variazione dell'altra
- Si definisce regressione la relazione statistica esistente tra 2 variabili legate da un rapporto di causa-effetto tali che le variazioni dell'una determinano le variazioni dell'altra: la variabile che si modifica in conseguenza dell'altra si dice dipendente mentre quella che ne determina le variazioni si definisce indipendente
- Indicando con x la variabile indipendente e con y la var. dip. si dice che y varia in funzione di x poiché esiste una funzione matematica che permette di calcolare i valori che assume la variabile y per ogni determinato valore della variabile x in base alla formula $y = f(x)$
- Regressione lineare = quando alle variazioni di x corrispondono variazioni uguali o proporzionali di y; in tal caso la relazione tra le 2 variabili è rappresentata dalla cosiddetta retta di regressione e la funzione è quella della retta

$$y = ax + b$$

dove a e b sono i 2 parametri che servono ad identificare la retta di regressione; in particolare a definisce l'inclinazione della retta di regressione e b la sua origine o posizione nel piano





a = coefficiente angolare della retta di regressione

$$a = \frac{cod_{xy}}{Dev_x}$$

Il valore di a è dato dal rapporto tra la cod e la dev della variabile indipendente

- Correlazione negativa = le 2 variabili variano in senso opposto

INDIFFERENZA STATISTICA = non variano insieme

INDIPENDENZA STATISTICA = non ci sono relazioni di causa effetto

- Quando il coefficente di correlazione è $= +1$ vuol dire che le 2 variabili sono direttamente proporzionali dal punto di vista geometrico = retta
- Quando il coeff. di corr. è $= -1$ le 2 var. sono inversamente proporzionali → retta
- Se il valore si avvicina di + all'unità vuol dire che c'è una buona relazione statistica. Se si avvicina a 0 è scarsa
- Tabella di correlazione x per caratteri qualitativi

OCCHI	CAPELLI		TOT
	CHIARI	SCURI	
CHIARI	200	150	350
SCURI	150	500	650
TOT	350	650	1000

caratteri concordanti = 200+500 = 700

" discordanti = 150+150 = 300

dividiamo tabella in 4 quadranti

n_1	n_2
n_3	n_4

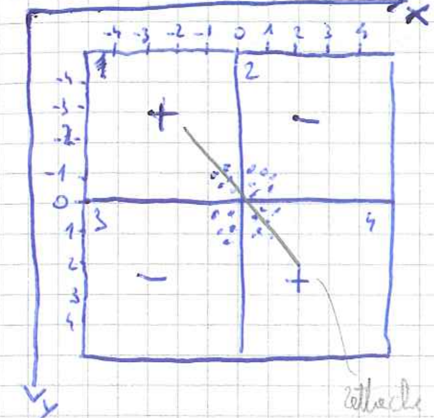
n_1, n_4 = caratteri concordanti

n_2, n_3 = " discordanti

$n_1 + n_4 > n_2 + n_3$ correl. positiva

$n_1 + n_4 < n_2 + n_3$ " negativa

$n_1 + n_4 = n_2 + n_3$ indifferenza statistica



retta che meglio approssima la nostra distribuzione di osservazione

1) Tutte le osservazioni inferiori alla media (entrambi le variabili hanno segno -) che sono inferiori alla media → correl. positiva

2) correl. negativa

3) scarti positivi 1° serie, scarti - 2° serie → correl. negativa

4) correl. positiva

- Si può conferire una rappresentazione visiva alla correl. mediante una tabella a doppia entrata, mettendo nella 1° colonna tutti i valori del fenomeno A e nella 1° riga tutti valori del fenomeno B. All'incrocio si osservano le frequenze x i corrispondenti valori dei 2 fenomeni

CORRELAZIONE

serie x	serie y	Dev x	dev y	cod xy
2-4 = -2	1-4 = -3	4	9	(-2)(-3) = 6
4-4 = 0	4-4 = 0			
6-4 = 2	7-4 = 3	4	9	2·3 = 6
		8	18	12

- correlazione → si divide codevianza x la radice quadrata del prodotto degli scarti

$$\text{Correlazione} = \frac{12}{\sqrt{18 \cdot 8}} = \frac{12}{\sqrt{144}} = \frac{12}{12} = 1$$

- coefficiente di correlazione (BRAVAIS) $= r = \frac{cod_{xy}}{\sqrt{D_x \cdot D_y}}$

$$r = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 \Sigma (y_i - \bar{y})^2}}$$

x	y	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
2	4	-2	4	1	1	-2
4	1	0	0	-2	4	0
6	4	2	4	1	1	2

$\bar{x} = \frac{12}{3} = 4$; $\bar{y} = \frac{3}{3} = 3$, Dev x = 8 , Dev y = 6 ; Cod xy = 0

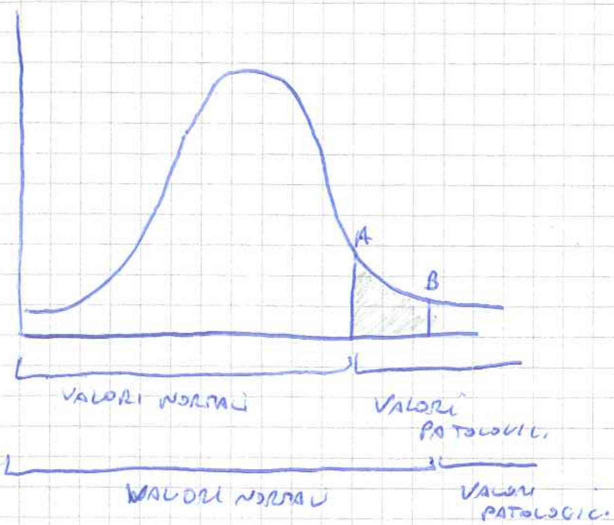
$$r = \frac{0}{\sqrt{8 \cdot 6}} = 0$$

- il coeff. di correlazione può avere un valore compreso tra 0 e 1 e può avere segno + o - che dipende dalla codevianza xchè può essere sia + che -. Segue le stesse regole della codevianza che è + quando le 2 serie di valori variano nello stesso senso, quando vanno in senso opposto è negativo
- la correlazione è la misura della relazione statistica tra 2 serie di valori. A variazioni del 1° corrispondono variazioni dell'altra. Questa relazione può essere + o - quando variazioni sono nello stesso senso e viceversa

→ Esistenza di una relazione statistica tra 2 variabili non dimostra necessariamente l'esistenza tra le 2 variabili di una relazione causa-effetto, questo xchè ci può essere relazione tra le 2 variabili che non dipendono tra loro, ma da una 3° variabile. Si coglie solo il fatto che le 2 variabili si modificano parallelamente, ma non devono essere concatenate

- x fenomeni biologici coefficiente non è mai uguale a 1 a causa della variabilità molto elevata

TEST SENSIBILITÀ E SPECIFICITÀ x UN DATO QUANTITATIVO



A = AUMENTANO i Falsi positivi

- SPECIFICITÀ

DIMINUISCONO i Falsi negativi

+ SENSIBILITÀ

B = DIMINUISCONO i FALSI POSITIVI

+ SPECIFICITÀ

Aumentano i Falsi negativi

- sensibilità

- Nei Test quantitativi la specificità e la sensibilità dipendono dove stabiliscono il valore soglia

$$\boxed{b = y - x}$$

Valutazione dell'azione farmacologica di 2 sostanze antibiotiche mediante prove in vitro

1) Antibiotico A

(DOSE) x	(DIAMETRO DI INIBIZIONE) y	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
3	23	-4	16	-4	16	16
7	27	0	0	0	0	0
11	31	+4	16	+4	16	16
21	81		32 (DEVIANZA x)	0	32 (DEVIANZA Y)	32 (COD)

$\bar{x} = 21:3 = 7$ $\bar{y} = 81:3 = 27$

$Cod\,xy = 32$ $r = \dfrac{32}{\sqrt{32 \cdot 32}} = \dfrac{32}{32} = 1$

correlazione positiva = le 2 variabili variano nello stesso segno

correlazione massima = a variazioni di una variabile corrispondono

CULTURA BATTERICA

DIAMETRO DI INIBIZIONE BATTERIA (DIAMETRO ZONA IN CUI I BATTERI NON CRESCONO)

variazioni uguali o proporzionali dell'altra variabile

$Cod_{xy} = 32$; $Dev\ X = 32$ $a = \dfrac{32}{32} = 1$

$$b = 23 - 3 = 20$$

Calcolo del diametro di inibizione della coltura batterica che si può ottenere con una determinata dose di antibiotico

per $x = 15$ $\quad y = 1 \cdot 15 + 20 = 35$

per $x = 20$ $\quad y = 1 \cdot 20 + 20 = 40$

- X certi valori di medicinale abbiamo la curva dose-~~effetto~~ come * sarebbe tale una retta ~~è infinito~~

Calcolo logaritmico

- uso logaritmi consente di costruire delle rette anche quando dalla distribuzione di valori non si avrebbe una retta e poter applicare così la regressione

$\log 1 = 0$

$\log 10 = 1$

$\log 100 = 2$

$\log 1000 = 3$

- i logaritmi in base 10 dei numeri da 2 a 9 sono numeri decimali inferiori a 1

- i logaritmi con base 10 dei numeri 11 e 99 sono numeri decimali aventi 1 come cifra a sinistra della virgola

- i logaritmi con base 10 dei numeri da 101 a 999 sono numeri decimali aventi 2 come cifra a sinistra della virgola

- si definiscono caratteristica del logaritmo le cifre a sinistra della virgola e mantissa le cifre a destra della virgola

- Teoremi

1) il log di un prodotto è uguale alla somma dei logaritmi dei singoli fattori (quindi x

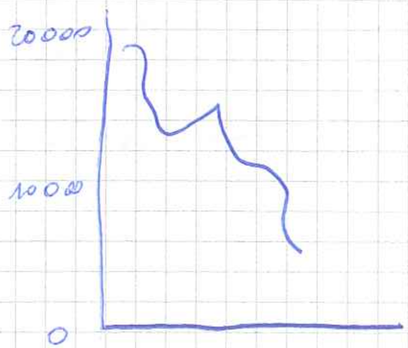
$$\log(20 \cdot 3) = \log 20 + \log = \begin{array}{r} 1,3010 + \\ 0,4771 = \\ \hline 1,7782 \end{array} = 60$$

2) il log di un quoziente è uguale alla differenza dei log del dividendo e del divisore (quindi x dividere un numero x un altro basta sottrarre il log del 2° dal log del 1° e individuare il numero al quale corrisponde)

3) il log della potenza di un numero è uguale al prodotto dell'esponente per il log del numero.

4) il log della radice di un numero è uguale al quoziente del log del numero x l'indice della radice

$$\log \sqrt{3} = \frac{\log 3}{2} = \frac{0,9542}{2} = 0,4771 = 3$$



Fenomeni biologici = variabili probabilistiche

STUDIO FENOMENI

- COSTANTI = SISTEMI DI MISURA

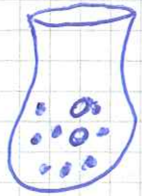
RICREARE ARTIFICIALMENTE GLI FENOMENI X STUDIARLI E RICAVARE DELLE LEGGI

- VARIABILI
  - DETERMINISTICI = SPERIMENTAZIONE (CLAUDE BERNARD 1813-1878)
  - PROBABILISTICI O ALEATORI = ~~calcolo~~ CALCOLO DELLE PROBABILITÀ (LAPLACE 1749-1827)

- ALLA BASE DEL CALCOLO PROBABILITÀ C'È UN RAPPORTO TRA GLI EVENTI FAVOREVOLI E IL TOTALE DEGLI EVENTI EGUALMENTE POSSIBILI

PROBABILITÀ SEMPLICE

Probabilità che si verifichi un determinato evento fra n eventi ugualmente possibili



Probabilità di estrarre dall'urna una pallina bianca è

$P = \frac{2}{10} = 0,2$ (20%)

↓ PROBABILITÀ semplice (che l'evento avvenga)

Q

↓ PROBABILITÀ OPPOSTA (che l'evento non si verifichi)

$0,2+0,8=1$

$\boxed{P+Q=1}$

- Quando la probabilità è 0 l'evento è impossibile
- " " " " 1 " " certo

$0<P<1$

Probabilità composta

probabilità che si verifichi 2 o + volte lo stesso evento

- estrazione con ~~[illegible]~~ ripetizione o senza ripetizione
- dobbiamo considerare metodo di estrazione
- alla estrazione con ripetizione = eventi indipendenti (non si influiscono a vicenda)
- se P_a = I estrazione e P_b = seconda estrazione la probabilità $P_{a \cdot b}$ di ~~est~~ estrarre pallina bianca in 2 estrazioni successive sarà:

1) In caso di estrazione con ripetizione (eventi indipendenti)

$$P_{a \cdot b} = P_a \cdot P_b$$

$$\frac{2}{10} \cdot \frac{2}{10} = \frac{4}{100} = 0{,}04$$

↓

Probabilità estrazione pallina bianca in 2 estrazioni successive con ripetizione

2) in caso di estrazione senza ripetizione (= eventi dipendenti)

$$P_{a \cdot b} = P_a \cdot P_{b/a}$$

$$\frac{2}{10} \cdot \frac{1}{9} = \frac{2}{90} = 0{,}0022$$

NB Se si tratta di 2 estrazioni da urne diverse la probabilità composta è come quella di 2 estrazioni con ripetizione dalla stessa urna.

= **Probabilità condizionata** = qual è la probabilità che si verifichi un evento se se ne è già verificato un altro

(Noi sappiamo già che si è verificato il 1° evento e vogliamo sapere qual'è la probabilità che si verifichi il 2°.)

= Probabilità che un evento si verifichi la seconda volta se si è verificato la prima volta

Se la P_a = I estrazione e P_b = II estrazione qual'è la probabilità ~~di~~ di estrarre pallina bianca la II volta se la prima è stata bianca?

[1] in caso di estrazione con ripetizione (eventi indipendenti)

$$P_{b/a} = P_b = \frac{2}{10} = 0{,}2$$

probabilità che sia uscita la pallina bianca la prima volta se è uscita bianca la seconda

2) in caso di estrazione senza ripetizione (eventi dipendenti)

$$P_{b/a} = \frac{P_b \cdot P_{a/b}}{P_a}$$

P_a → probabilità estrarre pallina bianca 1° volta

$$\frac{\frac{2}{10} \cdot \frac{1}{9}}{\frac{2}{10}} = \frac{\frac{2}{90}}{\frac{2}{10}} = 0{,}111\ldots$$

PROBABILITÀ TOTALE

Probabilità che si verifichi un evento almeno una volta in 2 tentativi successivi
(oppure probabilità che si verifichi uno dei 2 eventi possibili)

1) IN CASO DI EVENTI INCOMPATIBILI
(se si verifica uno non si può verificare l'altro)

Se nell'urna vi sono 2 palline bianche 3 palline rosse e 5 palline nere
la probabilità di estrarre pallina bianca o rossa alla prima estrazione è

$$P_{a+c} = P_a + P_c$$

$$\frac{2}{10} + \frac{3}{10} = \frac{5}{10}$$

2) EVENTI COMPATIBILI INDIPENDENTI (ESTRAZIONE CON RIPETIZIONE)
(verificarsi primo non influisce sul verificarsi del secondo)

Se P_a è la probabilità di estrarre pall. bianca alla I estrazione e P_b la prob. di estrarre pallina bianca alla II estrazione

$$P_{a+b} = P_a + P_b - P_{a.b}$$

$$\frac{2}{10} + \frac{2}{10} - \left(\frac{2}{10} \cdot \frac{2}{10}\right) = \frac{4}{10} - \frac{4}{100} = 0{,}4 - 0{,}04 = 0{,}36$$

Infatti in caso di estrazione con ripetizione si potranno verificare le seguenti possibilità

	I estraz	II estraz.	Prob. composta
1)	bianca	bianca	$\frac{2}{10} \cdot \frac{2}{10} = 0{,}04$
2)	bianca	nera	$\frac{2}{10} \cdot \frac{8}{10} = 0{,}16$
3)	nera	bianca	$\frac{2}{10} \cdot \frac{8}{10} = 0{,}16$
4)	nera	nera	$\frac{8}{10} \cdot \frac{8}{10} = 0{,}64$
			TOT = 1

la possibilità che soddisfano alla probabilità attesa sono la 1, la 2, e la 3 quindi

$$0{,}04 + 0{,}16 + 0{,}16 = 0{,}36$$

3) Eventi compatibili dipendenti
(estrazione senza ripetizione)

Se P_a è la prob. di estrarre pallina bianca alla I estraz.
e P_b è la prob. di estrarre p. bianca alla II estraz.

$$P_{a+b} = P_a + P_b - P_{ab} = P_a + P_b - (P_a \cdot P_{b/a})$$

$$\frac{2}{10} + \frac{2}{10} - \left(\frac{2}{10} \cdot \frac{1}{9}\right) = \frac{4}{10} - \frac{2}{90} = 0{,}3778$$

- P. totale maggiore della P. TOT in caso di estrazione con ripetizione

	I estraz	II estraz	Prob. complessiva
1)	bianca	bianca	$\frac{2}{10} \cdot \frac{1}{9} = 0{,}0222$
2)	bianca	nera	$\frac{2}{10} \cdot \frac{8}{9} = 0{,}1778$
3)	nera	nera	$\frac{8}{10} \cdot \frac{7}{9} = 0{,}6222$
4)	nera	bianca	$\frac{8}{10} \cdot \frac{2}{9} = 0{,}1778$

Le possibilità che soddisfano alla probabilità attesa sono 1, 2 e 4 quindi $0{,}0222 + 0{,}1778 + 0{,}1778 =$
$= 0{,}3778$

- non ci basiamo su un dato del passato x misurare la frequenza futura → purtroppo x i fenomeni biologici mi devo basare sulle frequenze passate

Problemi

- Il calcolo probabilità non può essere applicato ai fenomeni biologici xché non conosciamo la composizione dell'urna della natura (ci manca il denominatore) e non possiamo calcolare la probabilità.
Possiamo basarci sulla frequenza che però si basa sul passato, la probabilità sul ~~f~~ futuro.

- ci troviamo di fronte a una variabilità notevole e non possiamo limitare gli eventi a 2 o 3

Soluzione = variabilità ~~f~~ fenomeni biologici attraverso legge di Gauss e Laplace
possiamo conoscere la frequenza dei singoli valori

Se è frequenza relativa si esprime con la probabilità (es. 70%)

- Noi basiamo le nostre previsioni (probabilità = ~~f~~ futuro) su dati ~~r~~ relativi al passato

3 - Legge Gauss Laplace - calcolare probabilità di un fenomeno senza ricorrere al calcolo delle probabilità

Limiti ~~fiduciari~~ o di sicurezza

- Si basano sulla legge di Gauss e Laplace
- Si aggiunge e sottrae alla media 1,96 volte la deviazione standard si ottiene il 95% delle probabilità x arrotondare

$\bar{x} \pm 1,96\sigma$ (95% popolazione) intervallo in cui c'è il — probabilità sbagliare 5%

$\bar{x} \pm 2,576\sigma$ (99% ") probabilità sbagliare 1%

- Probabilità di errore accettabile

$\bar{x} \pm 3,291\sigma$ (99,9% popolazione) probabilità sbagliare (0,1%)

- Non c'è mai la sicurezza del 100% x chè la curva di Gauss tende all'in~~finito~~

Con curva Gaussiana grazie legge Gauss-Laplace posso conoscere intervallo entro il quale stanno il 95% di probabilità che l'individuo sia a quelle popolazioni

- nel caso in cui considero i singoli individui

Valore Z

dato da un rapporto

$$Z = \frac{x_i - \bar{x}}{\sigma}$$

x_i = valore di cui vogliamo conoscere la probabilità che rientri in una determinata popolazione

(Es. valore pressione 145)

$x_i - \bar{x}$ = intervallo all'interno della curva di Gauss che va o dall'una o dall'altra parte

- Si va a confrontare con delle tavole già note x vedere qual è la probabilità che una persona con quel valore faccia parte della popolazione sana
- se $Z = -1$ vuol dire che il valore trovato si colloca al di sotto della media ad una distanza da essa corrispondente alla d.s. (cioè entro i limiti compresi tra la media e la media $-\sigma$; se la distribuzione è normale questi limiti corrispondono al 34% della pop.)
- un valore di z compreso tra -2 e $+2$ racchiude il 95% della popolazione

Limiti fiduciari della media

1) $\bar{x} \pm 1,96\ \sigma_{\bar{x}}$ (sicurezza al 95%, Probabilità errore = 0,05%)

errore standard

nell'affermare che quel campione che ha quella media appartenga a quella popolazione

2) $\bar{x} \pm 2,576\ \sigma_{\bar{x}}$ (" " 99%, $p = 0,01\%$)

- Applicazioni importanti nella ricerca sperimentale (Es. farmaco nuovo si sperimenta su un certo numero di persone e si vede la media)

a) GRANDI CAMPIONI ($n > 30$) μ = media popolazione

1) $\bar{x} - 1,96\ \sigma_{\bar{x}} < \mu < \bar{x} + 1,96\ \sigma_{\bar{x}}$

media pressione su campione ammalati da non sappiamo l'intera popolazione

2) $\bar{x} - 2,576\ \sigma_{\bar{x}} < \mu < \bar{x} + 2,576\ \sigma_{\bar{x}}$

$\bar{x}$ = media del campione

μ = media della popolazione

$\sigma_{\bar{x}}$ = errore standard della media (calcolato sul campione)

Possiamo dire che la media della popolazione è compresa tra questi limiti, anche se non conosciamo la popolazione

- allargando l'intervallo della popolazione dei malati e riduciamo " " " dei sani

inconveniente pratico

Aumentando i limiti di sicurezza su un gran numero di farmaci che potrebbe essere efficace nel 95% dei pazienti lo rifiutiamo che non è efficace sul restante 4%

- Non è accettabile un limite sicurezza inferiore al 95%

- È il ricercatore che sceglie i limiti fiduciari

- La curva gaussiana la troviamo quando la serie è continua

- Quando abbiamo una variabile discreta o che si manifesta in forma binaria (SI, NO) non è possibile ricorrere alla curva gaussiana

$$\text{Errore standard} = \frac{\text{deviazione standard}}{\sqrt{\text{numero osservazioni}}}$$

- Se estraggo campioni rappresentativi popolazione calcoliamo le medie di questi campioni (medie campionarie) → non saranno tutte uguali ma si distribuiscono secondo curva gaussiana ma invece di essere regolata dalla deviazione standard è regolata dall'errore standard → sarà curva medie + stretta e + alta di quella singoli individui →

Lanciando un dado una volta qual'è la probabilità che esca un determinato numero?

Applicando la probabilità semplice sarà

$\frac{1}{6} = 0,1666\ldots = 16,67\%$

Qual'è la probabilità che il numero non esca? (probabilità opposta)

$\frac{5}{6} = 0,833\ldots = 83,33\%$

$p+q=1$ somma probabilità semplice e opposta è sempre 1

Lanciando un dado 2 volte, qual'è la probabilità che esca un determinato numero?

	I lancio	II lancio	
1)	NO	NO	numero non esce mai
2)	SI	NO	il numero esce una sola volta
3)	NO	SI	
4)	SI	SI	il num. esc. 2 volte

→ xche facendo la media andiamo giù a eliminare le differenze (ho diminuito la variabilità)

In base al calcolo delle probabilità sarà (probabilità composta):

	I lancio		II lancio	
1)	5/6	·	5/6	= 0,6944
2)	1/6	·	5/6	= 0,1389
3)	5/6	·	1/6	= 0,1389
4)	1/6	·	1/6	= 0,0278
				1,000

P_0 (probabilità che il numero non esca mai) = 69,44 %

P_1 (Probabilità che il numero esca 1 volta) = 27,78 %

P_2 (Probabilità che il numero esca 2 volte) = 2,78 %

$P_0 = \binom{2}{0} \left(\frac{1}{6}\right)^0 \cdot \left(\frac{5}{6}\right)^2$

$P_1 = \binom{2}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^1 =$

→ quando si considerano le medie anche le popolazioni che non si distribuiscono secondo una curva normale si distribuiscono così

Distribuzione Binomiale

- si usa quando la variabile è discreta

Curva normale = curva di distribuzione di frequenze

- Alcune di queste distribuzioni di frequenze vengono chiamate notevoli (distribuzione normale
  " binomiale)

- Funzione distribuzione binomiale

$$P_x = \binom{n}{x} p^x \cdot q^{n-x}$$

casi possibili

probabilità opposta (che il numero non esca)

probabilità semplice (che esca il numero che ho puntato)

coefficiente binomiale

casi favorevoli

probabilità che l'evento atteso si verifichi x volte su n tentativi successivi

coefficiente binomiale: rapporto a cui c'è al numeratore il prodotto degli x numeri decrescenti a partire da n

$$\binom{n}{x} = \frac{n(n-1)(n-2)\ldots(n-x+1)}{1\cdot 2\cdot 3 \ldots x}$$

al denominatore c'è il prodotto degli x numeri crescenti a partire da 1

Es) $\binom{4}{2}$ $\quad \frac{4\cdot 3}{1\cdot 2} = \frac{12}{2} = 6$

Es) probabilità che lanciando il dado 1 volta il numero esca 0 volte o 1 volta

$$P_0 = \binom{1}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^1 = 0,8333\ldots \quad (1)$$

$$P_1 = \binom{1}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^0 = 0,1666\ldots \quad (2)$$

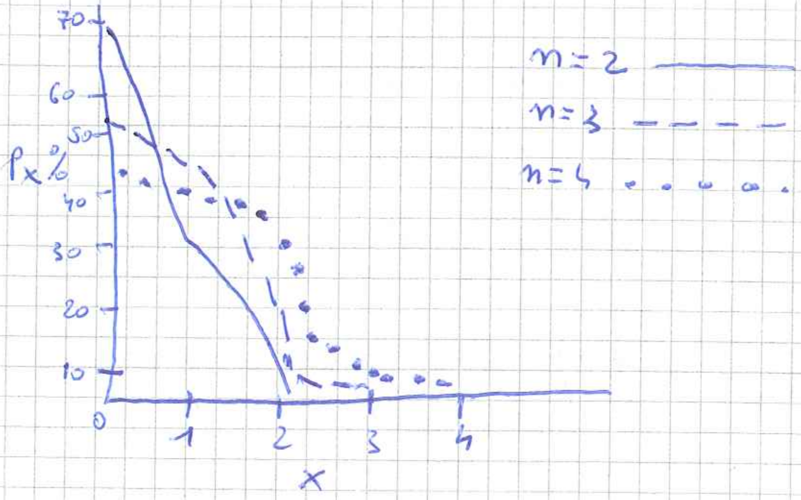
(1) corrispondono alla probabilità opposta $q = \frac{5}{6}$

(2) " " " semplice $p = \frac{1}{6}$

$$P_2 = \binom{2}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 =$$

0,83

Distribuzione binomiale



la frequenza relativa è uguale alla probabilità

- Queste leggi della probabilità si possono dimostrare empiricamente (attraverso la ripetizione dell'esperimento)

Media distribuzione binomiale

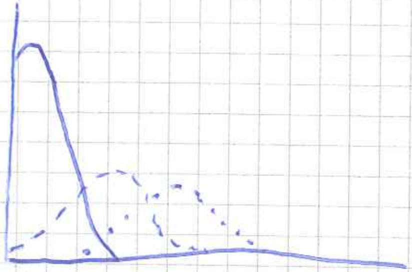
$\bar{x} = np$ numero lanci × probabilità semplice

VARIANZA BINOMIALE

$\sigma_x^2 = npq$ → PROBABILITÀ OPPOSTA

DEVIAZIONE STANDARD

$\sigma = \sqrt{npq}$



se consideriamo un numero elevato di lanci la distribuzione binomiale tende a una distribuzione normale

- legge empirica del caso

leggi che regolano i fenomeni casuali generali

- la distribuzione normale è parametrica xké × definire la funzione della curva c'è bisogno di 2 parametri (media e deviazione standard)

- la distribuzione binomiale è NON parametrica xké media e deviazione standard non sono 2 parametri fra li calcolano

la distribuzione binomiale quando il numero delle osservazioni è elevato assomiglia a una distribuzione normale

⇒ possiamo applicare la legge di Gauss e Laplace

NICOLA

Esempio)

In una popolaz. il 40% si ammala di influenza, prendo a ~~[illegible]~~ un campione di 100 persone e sottoposto a vaccinazione influenzale, 30 persone si ammalano di influenza. Qual'è la probabilità di trovare un medesimo numero di ammalati (in proporzione) nella popolazione non vaccinata?

(Abbiamo preso un campione → ci può essere un errore di campionamento)

probabilità che differenza tra 30 e 40% sia dovuta al caso

(cioè qual'è la probabilità di trovare 3 ammalati su 10 se il vaccino non avesse alcun effetto)

$$P_x = \binom{10}{3} \cdot \left(\frac{4}{10}\right)^3 \cdot \left(\frac{6}{10}\right)^{10-3} =$$

$$= \left(\frac{10\cdot 9\cdot 8}{1\cdot 2\cdot 3}\right)\left(\frac{64}{1000}\right)\left(\frac{279936}{1000000}\right) = 0,2149$$

Probabilità del 21,49% che il vaccino non è efficace

→ Troppo alta → non possiamo utilizzarla

La probabilità di trovare un solo ammalato sarebbe invece

$$P_x = \binom{10}{1}\left(\frac{4}{10}\right)^1\left(\frac{6}{10}\right)^9 = 0,04$$

Probabilità di errore 4%

il vaccino è efficace con il 96% di limiti di sicurezza e il 4% di probabilità d'errore

0,4 · 0,54 · 10

Media e deviazione standard della distribuzione binomiale

Se si lancia una monetina 100 volte la probabilità che esca testa è

$$p = \frac{50}{100} = 0,5$$

quindi la media della distribuzione binomiale sarà

$$n \cdot p = 100 \cdot 0,5 = 50$$

Legge empirica del caso (basata sulle osservazioni, non su un ragionamento)

La varianza è

$$npq = 100 \cdot 0,5 \cdot 0,5 = 25$$

e la deviazione standard è

$$\sigma = \sqrt{npq} = \sqrt{25} = 5$$

Quindi effettuando ripetute volte 100 lanci il numero di volte in cui uscirà testa sarà compreso tra

$$np \pm 2\sigma = 40 \vdash\!\dashv 60$$

Distribuzione di Poisson

– si ha quando sono variabili discrete, ma "p" è molto piccola ed "n" è molto grande, la funzione della distribuzione binomiale può essere sostituita dalla seguente

$$P_x = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{dove } \lambda = np - x$$

$e = 2,71828\ldots$

Probabilità x si verificati in n volte possibilità

media distribuzione Poisson

xchè n è molto grande p è molto piccolo quindi q (molto approssimato)

È come se fosse 1

$$\sigma^2 = \lambda = np$$

$$\sigma = \sqrt{np}$$

media e varianza coincidono

Applicando la funzione è possibile calcolare la probabilità P_x che l'evento si verifichi x volte per $x = 0, 1, 2, 3 \ldots n$

$$P_0 = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-\lambda} = \frac{1}{e^{\lambda}}$$

uguale a 1

$$P_1 = \frac{\lambda^1}{1!} e^{-\lambda} = e^{-\lambda} \cdot \lambda = \frac{\lambda}{e^{\lambda}}$$

$$P_2 = e^{-\lambda} \frac{\lambda^2}{2!} = \frac{\lambda^2}{2!\, e^{\lambda}}$$

$$P_3 = e^{-\lambda} \frac{\lambda^3}{3!} = \frac{\lambda^3}{3!\, e^{\lambda}}$$

$$P_n = e^{-\lambda} \frac{\lambda^n}{n!} = \frac{\lambda^n}{n!\, e^{\lambda}}$$

Esempio

In occasione di una lotteria x la quale sono previsti 100 premi, vengono venduti 1'000'000 di biglietti. La probabilità di vincere sarà:

$$p = \frac{100}{1'000'000} = \frac{1}{10'000} = 0,0001$$

Se un rivenditore ha venduto 10'000 biglietti la probabilità di vincere x coloro che hanno acquistato il biglietto sarà $\frac{1}{10'000}$

$n = 10'000$, $\lambda = np = 10'000 \cdot 0,0001 = 1$

media vincitori x ogni rivenditore sarà 1

Si vuol sapere qual'è la probabilità che presso quel rivenditore non vinca nessuno, vinca uno soltanto ecc....

$P_0 = 2,71828^{-1} = 0,36787$ 36% che nessuno vinca la lotteria di coloro che HA comprato BIGLIETTI PRESSO QUEL RIVENDITORE

$P_1 = 2,71828^{-1} \cdot 1 = 0,36787$

$P_2 = 2,71828^{-1} \cdot \frac{1^2}{2!} = 0,1839$

$P_5 = 2,71828^{-1} \cdot \frac{1^5}{5!} = 0,003$

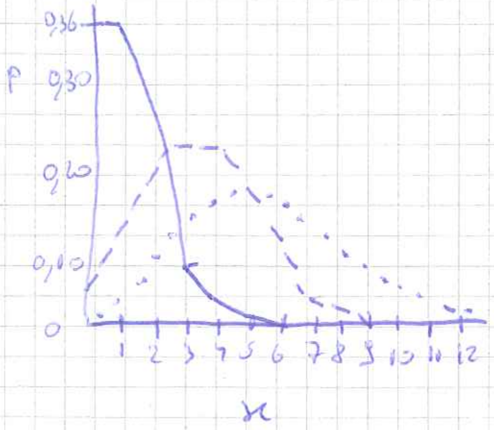
- Usata x probabilità trovare cellule malate in cellule sangue.

Distribuzione di Poisson x diversi valori di λ

—— $\lambda = 1$

– – – $\lambda = 3$

. . . . . $\lambda = 5$



Aumentando λ aumenta il numero delle osservazioni x che dipende da esso essendo p molto piccolo

Funzioni delle diverse distribuzioni

Binomiale $\quad P_x = \binom{n}{x} p^x q^{n-x}$

Poisson $\quad P_x = \frac{\lambda^x}{x!} e^{-\lambda}$

Normale $\quad y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}$

Probabilità a priori o matematica = non cambia fino a che non cambia il numeratore o il denominatore del rapporto

(se non conosciamo la composizione dell'urna della natura)

- Ma se non abbiamo la possibilità del numero di eventi favorevoli o possibili si ricorre alle 3 distribuzioni notevoli. Applichiamo la probabilità statistica o a posteriori (dobbiamo prima conoscere le diverse frequenze relative e poi calcoliamo le probabilità)
- Non potendo conoscere i dati sull'intera popolazione si estrae un campione a cui si applicano le diverse distribuzioni
- Applicando la legge di Gauss e Laplace possiamo conoscere tutte le caratteristiche che ci interessano

NB il campione deve essere abbastanza numeroso se no non posso applicare le distribuzioni normali

- Quando ci troviamo di fronte a campioni ridotti (Es. sperimentazione farmaco su poche persone) non potendo usare le distribuzione notevoli devo ricorrere a degli adattamenti delle distribuzioni notevoli.

→ Metodo dei Tests di significatività

- Si parte da un ipotesi di lavoro ovvero i risultati che mi aspetto
- Si fa l'esperimento x vedere se questi risultati attesi sono confermati da quelli ottenuti
- Se troviamo una differenza tra questi risultati mi devo chiedere se è dovuta al fatto che abbiamo operato su un campione o se è reale.
(Se fosse dovuta a un errore di campionamento non avrebbe nessun significato)

- Non si può avere la sicurezza al 100% xchè la curva di Gauss va da + ∞ a − ∞

- i tests di significatività si utilizzano x verificare se la differenza tra i valori attesi ed i valori ottenuti può essere dovuta al caso

valori ottenuti − valori attesi = → differenza significativa / → " non significativa

differenza significativa: si può escludere che la differenza sia dovuta al caso

" non significativa = non si può escludere che sia dovuta al caso → la escludiamo

- Possiamo stabilire anche il $\boxed{\text{livello di significatività}}$

Probabilità residua di errore nell'affermare che la differenza non può essere dovuta al caso = ripetendo l'esperimento n volte

(La probabilità dev'essere inferiore al 5%)

rappresenterebbe il numero di volte nelle quali si può ottenere la medesima differenza per caso ($p = 0{,}05$ vuol dire che se si ripetesse l'esperimento 100 volte, in 5 di esse la differenza sarebbe dovuta al caso)

Effetto Placebo: autosuggestione dovuto a un farmaco sul paziente
→ potenzia effetto medicina

- Ipotesi zero = si somministra ad alcuni il farmaco ad altri placebo. E si vede l'effetto

Se la differenza è significativa rifiuto l'ipotesi zero (cioè che la diff. sia dovuta al caso).

Tests
- PARAMETRICI = DERIVANO DA UN'ADATTAMENTO ALLA DISTRIBUZIONE NORMALE
- NON PARAMETRICI = DERIVANO DA UNA DISTRIBUZIONE BINOMIALE (si usano nei casi in cui non possiamo essere sicuri che la popolazione a cui si riferisce il campione non si distribuisce lungo una curva normale)

Test di Student (dati appaiati)

- Test parametrico (basato sulla distribuzione normale)

↗ variabile continua

- Si usa quando si confrontano le medie di 2 distribuzioni

2 Applicazioni = RIDUCE AL MINIMO I FATTORI SUBSPERIMENTALI CIOÈ DI AVERE FATTORI CHE NON CI INTERESSANO x esperimento.

1) Applicazione = quando si tratta degli stessi malati sottoposti a 2 diversi trattamenti (Test T per dati appaiati)

- Si usa anche x gemelli omozigoti
- Cavie provenienti da medesima nidiata

RISCHIO CHE QUANDO SOMMINISTRIAMO IL 2° FARMACO NON SONO ANCORA SCOMPARSI EFFETTI DEL 1° FARMACO
→ SI FA "WASH OUT" INTERVALLO FRA UN FARMACO E L'ALTRO NEL QUALE IL PAZIENTE NON PRENDE NIENTE (MA SE C'È UN F. NUOVO NON CONOSCIAMO IL PERIODO DI ATTESA x ELIMINARE EFFETTI E POI x UN PERIODO DI TEMPO IL PAZIENTE NON SUBISCE NESSUNA CURA)

$$t = \frac{\bar{d} - 0}{Es\,d} = \frac{\bar{d} - 0}{\sqrt{\frac{\Sigma(d_i - \bar{d})^2}{n-1} \cdot \frac{1}{n}}}$$

$\bar{d}$ = $\boxed{\text{differenza media}}$. Se somministriamo 2 farmaci diversi allo stesso malato avremo una differenza di azione e calcoliamo la diff. media x tutti gli animali

0 = riferito all'ipotesi zero in ogni ammalato l'effetto dei 2 farmaci è stato perfettamente uguale

$\boxed{E_S}$ = errore standard (definisce la distribuzione delle medie campionarie intorno alla media della popolazione)

in questo caso
Errore standard delle differenze

$$\sqrt{\frac{\Sigma(d_i-\bar{d})}{n-1}} =$$ Sommatoria devianza delle differenze registrate dai 2 farmaci nei pazienti diviso i gradi di libertà
= deviazione standard

- moltiplico per il tutto $\times \frac{1}{n}$ x trovare E_S.

infatti $E_{Sx} = \frac{\sigma x}{\sqrt{n}} = \sqrt{\frac{\sigma^2 x}{n}} = \sqrt{\sigma^2 x \cdot \frac{1}{n}} =$

$$= \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1} \cdot \frac{1}{n}}$$

e quindi $ES_d = \frac{\sigma d}{\sqrt{n}} = \sqrt{\frac{\sigma^2 d}{n}} = \sqrt{\frac{\Sigma(d_i-\bar{d})^2}{n-1} \cdot \frac{1}{n}}$

- il Test t serve per saggiare se è accettabile l'ipotesi zero; il risultato è significativo se possiamo rifiutare l'ipotesi con sufficiente sicurezza

Esempio: Si sottopongono 4 pazienti a 2 diversi trattamenti con ipotensivi (x abbassare ~~xxxx~~ pressione) e si vogliono confrontare i risultati ottenuti

Pazienti	Trattamento A	T. B	Diff Tra i 2 Trattamenti
1)	180 - 160 = 20	180 - 165 = 15	20 - 15 = 5
2)	190 - 150 = 40	190 - 160 = 30	40 - 30 = 10
3)	195 - 140 = 55	195 - 145 = 50	55 - 50 = 5
4)	190 - 140 = 50	190 - 160 = 30	50 - 30 = 20
			TOT 40

$\bar{d} = \frac{40}{4} = 10$

$(5-10)^2 = 25$
$(10-10)^2 = 0$
$(5-10)^2 = 25$
$(20-10)^2 = 100$
150

Dev d = 150

$\sigma^2 d = \frac{150}{3} = 50$

3 → gradi di libertà

$ES_d = \sqrt{50 \cdot \frac{1}{4}} = \sqrt{50 \cdot 0,25} = 3,53$

$$t = \frac{10}{3,53} = 2,83$$ (NON significativo)

valore di t con 3 gradi di libertà (per P=5%) =

= 3,182

- Se diff. non è significativa e non si può respingere l'ipotesi zero

2) APPLICAZIONE DATI NON APPAIATI (APPAIATI DIVERSI e QUINDI LA CONFRONTABILITÀ È MOLTO MINORE)

Confronto tra 2 medie (dati non appaiati)

$$t = \frac{\text{differenza tra le medie campionarie}}{\text{errore standard della differenza tra le medie}} =$$

$$= \frac{\bar{x} - \bar{y}}{\sigma_{\bar{x}-\bar{y}}}$$

- Se il test è significativo vuol dire che le medie appartengono a 2 popolazioni diverse (1 gruppo è preso da popolazione malata e quello sana)

$$\sigma_{\bar{x}+y} = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

~~[illegible]~~ delle differenze tra le medie campionarie

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2}{n_x + n_y - 2}} \cdot \sqrt{\frac{N_x + N_y}{N_x \cdot N_y}}} =$$

osservazioni x + osservazioni y

somma devianze diviso somma gradi di libertà

fattore di dimensione

$$= \frac{\bar{x} - \bar{y}}{\sigma_{xy}} \cdot \sqrt{\frac{N_x \cdot N_y}{N_x + N_y}}$$

il valore di t serve x saggiare le probabilità che le 2 medie si riferiscano ad una stessa popolazione (ipotesi 0)

Es.

Verificare l'efficacia di un farmaco ad azione ipocolesterolizzante, furono confrontate le quantità di colesterina nel tessuto nervoso di un gruppo di cavie sottoposte al trattamento

GRUPPO SPERIMENTO x_i	GRUPPO CONTROLLO y_i
1,64	1,92
1,64	1,92
1,64	1,92
1,64	1,92
1,89	1,94
$\bar{x} = 1,69$	$\bar{y} = 1,924$

$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1,64 - 1,69 = -0,05	0,0025
" -0,05	"
" -0,05	"
" -0,05	"
1,89 - 1,69 = 0,20	0,0400
	0,05

Dev x = 0,05

$y_i - \bar{y}$	$(y_i - \bar{y})^2$
	0,000016
	"
	"
	"
	0,000256
	0,00032

Dev y = 0,00032

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{Dev\,x + Dev\,y}{(n_x-1)(n_y-1)}} \cdot \sqrt{\frac{n_x+n_y}{n_x \cdot n_y}}$$

$$\sqrt{\frac{0,05 + 0,00032}{4+4}} \cdot \sqrt{\frac{5+5}{5 \cdot 5}} = \sqrt{\frac{0,05032}{8}} \cdot \sqrt{\frac{10}{25}} =$$

$$= \sqrt{0,00629} \cdot \sqrt{0,4} = 0,079 \cdot 0,632 = 0,05$$

$$t = \frac{1,924 - 1,69}{0,05} = \frac{0,234}{0,05} = 4,68$$

Per gradi libertà = 8 e P = 0,01 il valore riportato nella tavola del T di student è 3,355 e quindi la differenza è significativa all'1%

La diff. è significativa escludiamo che sia dovuta al caso e affermando ciò la probabilità d'errore è inferiore all'1%

Test χ^2 di Pearson

- Si usa quando si vogliono confrontare 2 distribuzioni di frequenze

$$\chi^2 = \sum \frac{(x_i - y_i)^2}{y_i}$$

valori attesi o teorici

valori ~~osservati~~ OSSERVATI

- test non parametrico (distribuzione ~~discreta~~ binomiale) variabile discreta

x_i = frequenze osservate ; y_i = frequenze teoriche

- Attribuiamo ai singoli gruppi la stessa distribuzione di frequenze del totale

- Mescoliamo i 2 gruppi e andiamo a vedere quanti sono quanti e la percentuale di quanti la applichiamo ai singoli gruppi

→ così calcoliamo le frequenze teoriche

- come si fanno ad avere i valori attesi?

Esempio

Frequenza del sintomo disorientamento spaziale e temporale e durata della degenza in ospedale psichiatrico

	DURATA RICOVERO			
	Fino a 20 g	21 - 30	oltre 30	TOT
PRESENZA SINTOMO	117 (65%)	59 (72%)	49 (86%)	225 (70%)
ASSENZA SINTOMO	63 (35%)	23 (28%)	8 (14%)	94 (30%)
TOT. PAZ.	180	82	57	319

Calcolo delle frequenze teoriche (ipotesi che la durata del ricovero fosse uguale nei 2 gruppi di pazienti) (con o senza il sintomo)

$180 : 319 = x : 225 \qquad x = 126,96$

$82 : 319 = x : 225 \qquad x = 57,84$

$57 : 318 = x : 225 \qquad x = 40,20$

$180 : 319 = x : 94 \quad x = 53,04$

$82 : 319 = x : 94 \quad x = 24,16$

$57 : 319 = x : 94 \quad x = 16,80$

	FINO A 20	21-30	OLTRE 30	TOT
PRESENZA SINTOMO	126,96 $(-8,96)^2$	57,84 $(+1,16)^2$	40,20 $(+8,80)^2$	225
ASSENZA SINTOMO	53,04 $(+8,96)^2$	24,16 $(-1,16)^2$	16,80 $(-8,80)^2$	94
	180	82	57	319

$$x^2 = \frac{(-8,96)^2}{126,96} + \frac{(+1,16)^2}{57,84} + \frac{(+8,80)^2}{40,20} +$$

$$+ \frac{(+8,96)^2}{53,04} + \frac{(-1,16)^2}{24,16} + \frac{(-8,80)^2}{16,8} = 9,15$$

$x^2 = 9,15$; gradi libertà $= (3-1)(2-1) = 2 \cdot 1 = 2$

Quando si tratta di una tabella a doppia entrata i gradi di libertà sono dati dal numero delle colonne − 1 moltiplicato il numero delle righe − 1 (indipendentemente dal n° delle osservazioni)

- La differenza è significativa. Si può escludere che il sintomo sia dovuto al caso con probabilità d'errore inferiore al 5% e superiore all'1%

Calcolo del x^2 in tabelle di contingenza 2x2

Esempio: 2 gruppi di ammalati sono stati sottoposti a 2 diversi trattamenti: nel gruppo

Valori osservati

Terapia	GUARITI	NON GUARITI	TOTALE
Trattamento A	10 (17,86%)	46	56
" B	11 (45,83%)	13	24
TOTALE	21 (26,25%)	59	80

Valori Teorici

Terapia	Guariti	Non guariti	Totale
Trattamento A	14,7	41,3	56,0
" B	6,3	17,7	24,0
Totale	21,0	59,0	80,0

Valori osservati (con correzione di Yates)

- si diminuiscono dello 0,5 le frequenze + ~~basse~~ ALTE
  si aumentano " " " + ~~alte~~ BASSE
- altero i dati così è + difficile che il Test sia significativo

Terapia	Guariti	Non guariti	Totale
A	10,5	45,5	56,0
B	10,5	13,5	24,0
Totale	21	59,0	80,0

- Si cerca di avvicinare i 2 gruppi x rendere meno significativa la differenza

$$x^2 = \frac{(10-14,7)^2}{14,7} + \frac{(10,5-6,3)^2}{6,3} + \frac{(45,5-41,3)^2}{41,3} + \frac{(13,5-17,7)^2}{17,7} = 5,43$$

gradi Libertà = 1 $\qquad x^2$ per $P = 0,05 = 3,841$

x^2 per $P = 0,01 = 6,635$

Con la correzione scende il livello di significatività

Test di Fisher

Serve a confrontare la variabilità di 2 serie di valori onde accertare se la differenza tra la variabilità attesa e variabilità ottenuta può essere dovuta al caso

$$F = \frac{\sigma^2 x}{\sigma^2 y}$$

dove $\sigma^2 x$ è la ~~variazione~~ VARIANZA della serie x

e $\sigma^2 y$ è la " della serie y

avendo l'avvertenza di mettere sempre al numeratore la varianza maggiore.

il Test F di Fisher si usa nell'analisi della varianza (ANOVA)

VRQ = verifica revisione qualità

in tutti i servizi pubblici deve essere fatta periodicamente x vedere la qualità del servizio

ipotesi di lavoro

↓

acquisizione dei dati (seriazioni dati, controllo qualità, errori che si possono commettere, come evitare errori evitabili e ridurre errori inevitabili)

↓

elaborazione dei dati → rendere dati confrontabili con altri dati (es. intervalli di riferimento)

↓

analisi statistica → ipotesi di lavoro

risultati attesi = risultati ottenuti | ipotesi 0

risultati attesi > risultati ottenuti | ipotesi 1
risultati attesi < risultati ottenuti |

- Ipotesi lavoro = Acquisizione di ricerche precedenti o studi teorici (in caso farmaci)

Necessità di procedere sempre attraverso un'ipotesi di lavoro

↓

Così si può verificare perché non si è attuata la previsione che ha fatto e si va ad analizzare le cause (errori o altro ecc.)

- distribuzione in classi e priori (decidiamo prima l'ampiezza delle classi)

- Percentili = si distribuiscono tutte le osservazioni in ordine di grandezza e le si va a suddividere in parti uguali (~~[illegible]~~ quartili = 25%, quintili = 20%, decili = 10%)

↓

si vanno poi a individuare i valori in cui è compreso ogni quartile e si fanno ~~[illegible]~~ i limiti delle classi

→ classi ottenute a posteriori

- i limiti dei percentili non hanno uguale altezza (non c'è intervallo di classe)
- l'uso dei 2 metodi cambia a seconda delle finalità. Uso le classi se devo poi fare ulteriori ~~[illegible]~~ elaborazioni
- Per avere idea di come sono distribuiti i valori di un determinato collettivo e vedere meglio il singolo individuo rispetto al collettivo si usano i percentili.

- Frequenza cumulata = imp. perché ~~ci~~ non ci interessa spesso sapere qual è la frequenza di un ~~[illegible]~~ valore ma di tutti i valori fino a un certo

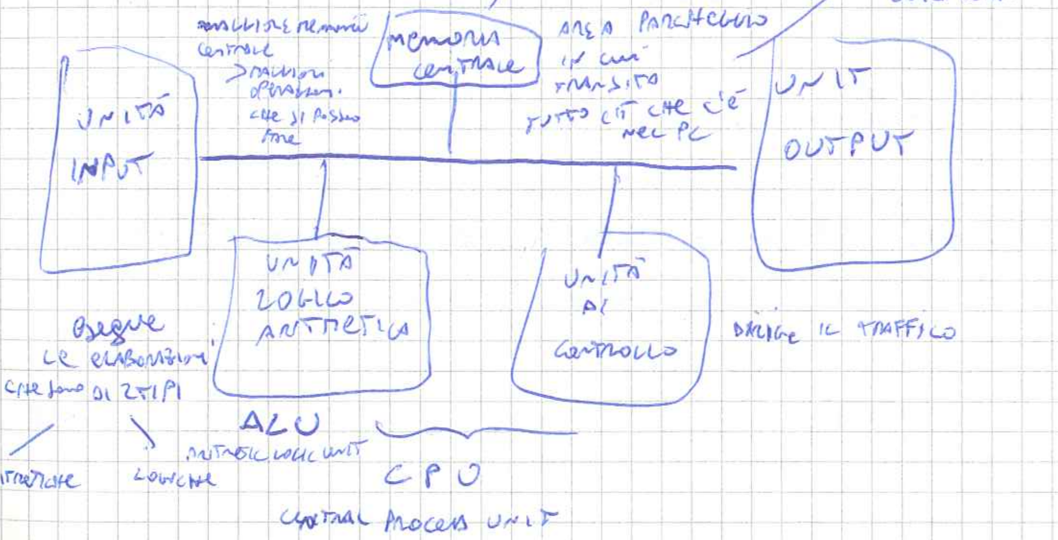
---

- Hardware = non può essere modificato da utente — componenti meccaniche ed elettroniche computer

Software = ogni utente può modificare — programmi

2 tipi software
- SOFTWARE DI BASE: UNO CHE SERVE A FAR FUNZIONARE PC. SISTEMI OPERATIVI. SERVE A GESTIRE LA MACCHINA
- SISTEMA OPERATIVO DI SUPPORTO x UTENTE (UTILITIES)
  SOFTWARE APPLICATIVO

Hardware = diverse unità o componenti



Spazio necessario a calcolare dati
si utilizzano unità di misura

BIT = spazio in memoria ~~x~~ contenere
una cifra del sistema binario

BYTE = 8 BIT x le in maniera x
scrivere una lettera dell'alfabeto occorrono
8 BIT

+ RAPIDA memoria periferiche

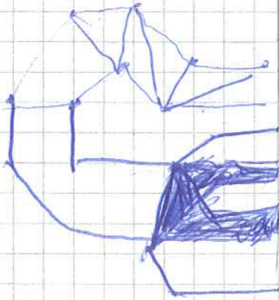
memoria RAM = RANDOM ACCESS MEMORY
c'è la possibilità di accedere a memoria + recuperare dato
proprio dove si trova il dato
~~memoria ...~~

Memoria
centrale
(distinzione fatta all'inizio dal costruttore e non modificabile)

memoria ROM = READ ONLY MEMORY
sono contenute istruzioni
x la gestione delle risorse
interne (funzionamento PC)

memorie periferiche = sequenziale, deve analizzare
uno x 1 tutti i dati.

memoria RAM → DATI da elaborare
→ ISTRUZIONI (programmi)
↓
divisione
a libera scelta
dell'utente



nelle memoria c'è uno spazio
occupato da indirizzi

Locazioni = spazi dove
sono situate determinate
informazioni

A ogni locazione corrisponde un
indirizzo = codice che serve a identificare
spazio memoria dove si trova
informazione

↓ permette a sistema centrale
di andare a recuperare
direttamente dati

Diff. tra memoria centrale e periferica
in struttura fisica.

memorie centrale elettronica
(funzione su passaggio e-)

memorie periferica o ottiche
o magnetiche

(Hard disk memorie periferica)
cd-ROM

↓ raggi laser
imprimo segnali su
particolare sostanza
a base d'argento.

RAGGIO LASER = capacità miniaturizzazione

cd-rom = molte + informazioni e rapidità d'accesso maggiore

- Memoria centrale si annulla da sola
spegnendo il P.C.

→ funzionano mediante e- mediante
CHIPS o circuiti integrati,
piastrine di silicio in cui sono
tracciati circuiti in cui passano
e-

Inserimento dati può avvenire anche da memorie periferiche

- Inserimento dati può avvenire anche attraverso
lettori ottici
Scanner = scansione ~~[illegible]~~
~~[illegible]~~

↓

scomposizione
in + parti, riesce
a riconoscere il tutto anche da piccole parti.

- INCONVENIENTI = anche scanner +
avanzati possono commettere
errori

- c'è anche codice a barre

- banda magnetica

- sistema a microchip
SMART-CARD
permette di proteggere
dati sensibili come dati sanitari

Sistema binario = sistema di numerazione

Linguaggio macchina

Computer tradurre istruzione che

linguaggio simbolico in linguaggio macchina

attraverso programma compilatore
TRADUTTORE

programma sorgente tradotto in linguaggio macchina

Linguaggi simbolici + usati = BASIC

begginers all purpose symbolic instruction code

- linguaggio simbolico * principianti e tutte

le applicazioni (linguaggio universale)

~~[illegible]~~

- Sistemi trasmissione dati e comunicazione

da un PC ad un altro. Si usa linea

telefonica. → compito × trasmettere segnali

analogici (espressi da forme, tracciati, onde)

Segnale digitale = si basa su numeri

MODEM = modulatore - demodulatore di frequenza

↓ trasforma segnale digitale in analogico

e viceversa.

- Rete locale = che non passa attraverso centraline Telecom (non si paga)

- Rete normale = si passa x Telecom (si paga)

- linee ISDN = rete telefonica nella quale i segnali vengono trasmessi in forma DIGITALE

Integrated Services Digital Network

Rete digitale x servizi integrati

Si acquista molto in rapidità

~~Anche~~

- anche TV diventerà digitale

- Si possono utilizzare anche onde radio x trasmissione dati → sistemi satellitari.

Conversione analogico - digitale attraverso diagrammi - cartesiani

ogni punto tracciato corrisponde a coppia di numeri

~~[illegible]~~ ~~[illegible]~~

~~[illegible]~~ ~~[illegible]~~ ~~[illegible]~~ x non perdere dati e occupare poco spazio

procedimento di sostituzione

si sostituisce a tracciato figure geometriche rettangoli che hanno stesse base, ma diversa altezza. X identificare area rettangolo mi basta un singolo numero, altezza e l'area rappresenta spazio sotteso e curva iniziale

Media aritmetica = si calcola media classi (valori centrali) che poi moltiplico x frequenza sommiamo e dividiamo x 100.

=> m_p stature 174,

moda = 175 che ha frequenza maggiore

stature 50 e 51 grasse o poi si divide x 2

- mediana = valore centrale classe dove sono il 50 e 51 grasse

quando i valori sono divisi in classi o corrisponde una frequenza

~~med~~ ~~D~~ Elevare al quadrato scarti e moltiplicare per frequenze, faccio somma = $\frac{\text{Devianza}}{\text{numero osservazioni}} = \sqrt{\text{Varianza}} = \sqrt{\text{scarto quadratico medio}}$

Applicazione pratica limiti fiduciari

Giovane alto 185 cm possiamo

dire che c'è il 95% possibilità

che appartiene a popolazione Lazio

o Piemonte e il 5% di errore

→ Possiamo affermare che non appartiene

a popolazione sardegna

- Se riconosciamo a valore z

a sapere con

sicurezza a quale

popolazione appartiene.

$$\frac{x_i - \bar{x}}{\sigma}$$

valore di cui vogliamo calcolare appartenenza

$$\frac{1{,}85 - \underset{\text{piemonte}}{\text{media}}\ (174)}{7{,}23}$$

$$\frac{1{,}85 - \underset{\text{Lazio}}{\text{media}}\ 174}{7{,}03}$$

Lo andiamo a confrontare con

tavole apposite e sappiamo qual è

la probabilità che il giovane appartenga

a quella regione