



*Avvio alla
Statística
e alla
Probabilità*

Summer School

San Pellegrino Terme, 7 settembre 2015

Summer School 2015

Valeria Caviezel

Università degli Studi di Bergamo

Dipartimento di Economia

valeria.caviezel@unibg.it



Ponete delle domande... cercherò di rispondere!!

Summer School 2015



La formazione continua...
... dalla scuola superiore
all'università.

Non si modificano gli argomenti trattati,
cambia il livello.

Summer School 2015



La statistica è presente
in tutti gli indirizzi della
scuola secondaria di
secondo grado e nella

maggior parte dei percorsi di studio
universitario (materia fondamentale o
opzionale).

Summer School 2015

Provenienza

- **Licei**
- **Istituti tecnici**
- **Istituti professionali**

Destinazione

- **Ingegneria**
- **Economia**
- **Sociologia**
- **Psicologia**
- **Scienze**
- **Medicina**
- **.....**

Summer School 2015



Come la matematica, la statistica è uno strumento che aiuta nella risoluzione di problemi semplici o complessi.

Consente di analizzare e sintetizzare (riassumere) dati...

e il mondo di oggi è sempre più ricco di informazioni.

Oggi



[...] Quotidianamente siamo bersagliati da dati statistici. Spesso, le informazioni sono riferite nel contesto di una notizia descritta dal quotidiano che leggiamo o fornita in un telegiornale; qualche volta i dati sono offerti sotto forma di tabelle o grafici; altre volte si tratta di percentuali o di medie. Per questo è necessario attrezzarsi in forma adeguata per **valutarli criticamente**. Questi strumenti minimi di analisi dovrebbero far parte del **bagaglio culturale di ogni cittadino**.



La statistica aiuta a
comprendere meglio la realtà
nella quale tutti noi viviamo.

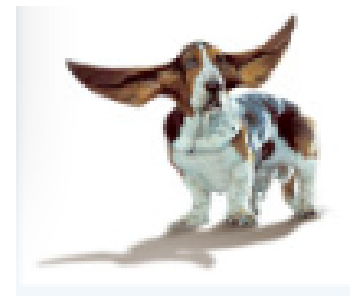
Oggi: notizie

UN SONDAGGIO CONDOTTO SU 2 MILA PERSONE

Giovani inglesi: l'80% non sa leggere una carta stradale

Non sono in grado di capire una mappa e senza i navigatori satellitari si sentono persi. Specie le donne

Dall'interpretazione statistica dei dati si evidenzia che ci si attende che il 30% del latte di pecora sia contaminato in misura superiore al limite di legge». Il rapporto è da alcuni giorni



Elezioni 2013: in Lombardia è testa a testa

Posted on [gennaio 17, 2013](#)

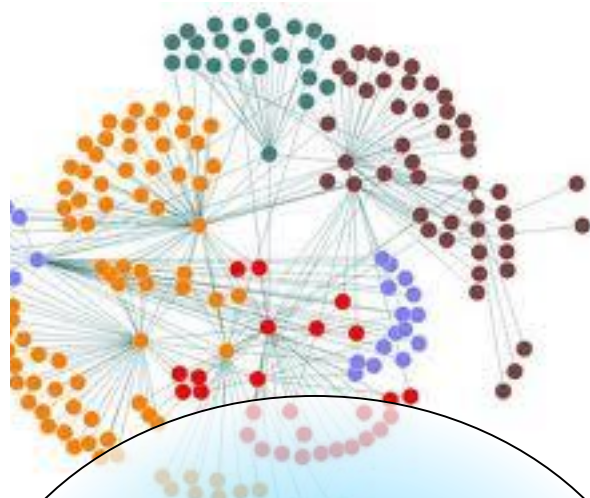
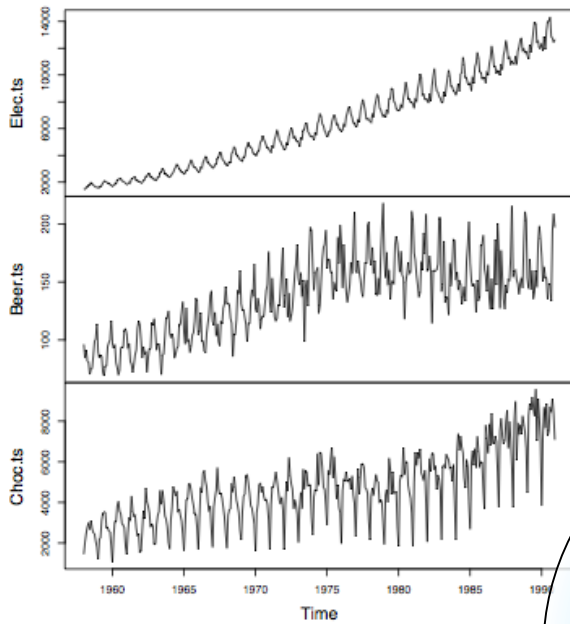
Per molti commentatori, rischia di essere l'Ohio delle **elezioni politiche italiane 2013**. Chi vincerà in **Lombardia al Senato** deciderà infatti le sorti della futura maggioranza di governo. Se a prevalere sarà il **centro-sinistra**, probabilmente riuscirà

Troppe bibite gassate, rischio depressione. Per l'umore è meglio una tazza di caffè

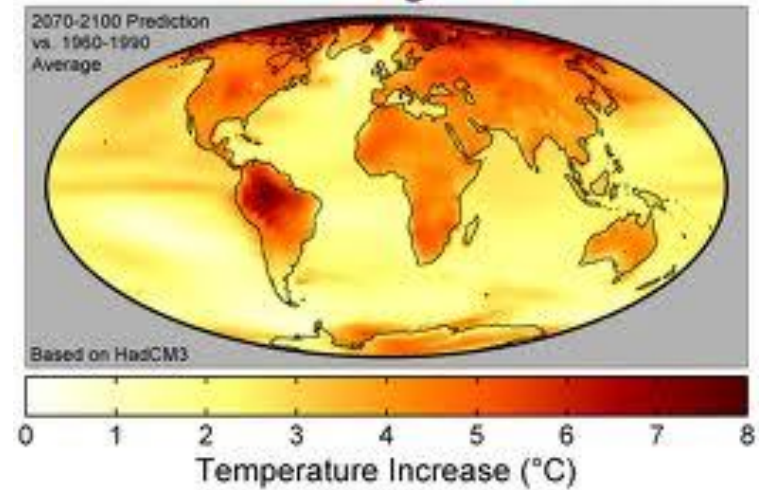
Studio Usa su oltre 250mila adulti seguiti a distanza di dieci anni: i grandi consumatori di bevande zuccherate, specie se in versione light, hanno probabilità fino al 38 per cento maggiori di sviluppare il disturbo. Ma il rischio diminuisce con consumo di caffè

Oggi: applicazioni

Chocolate, Beer, and Electricity Production: 1958–1990

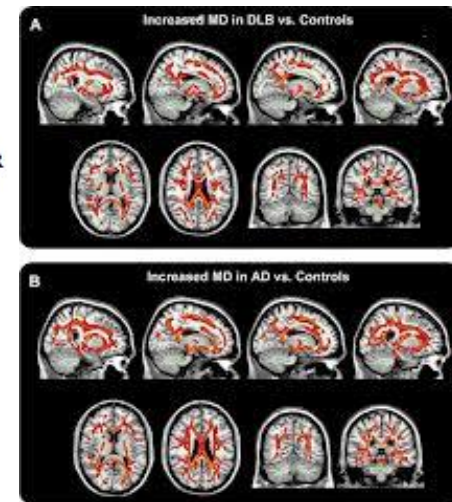
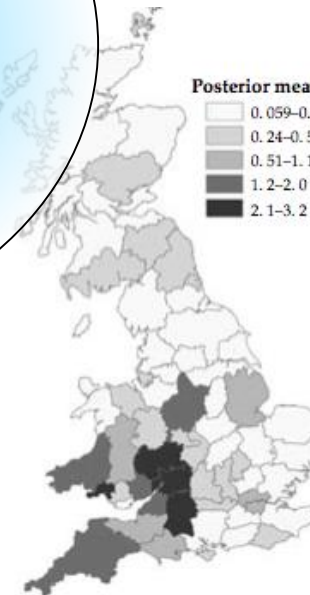
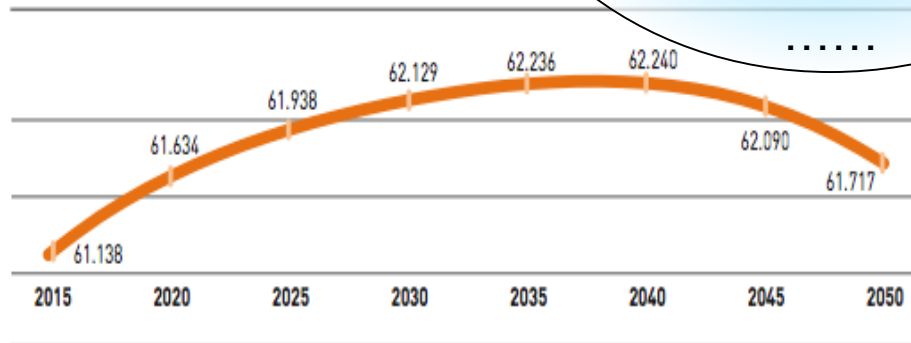


Global Warming Predictions



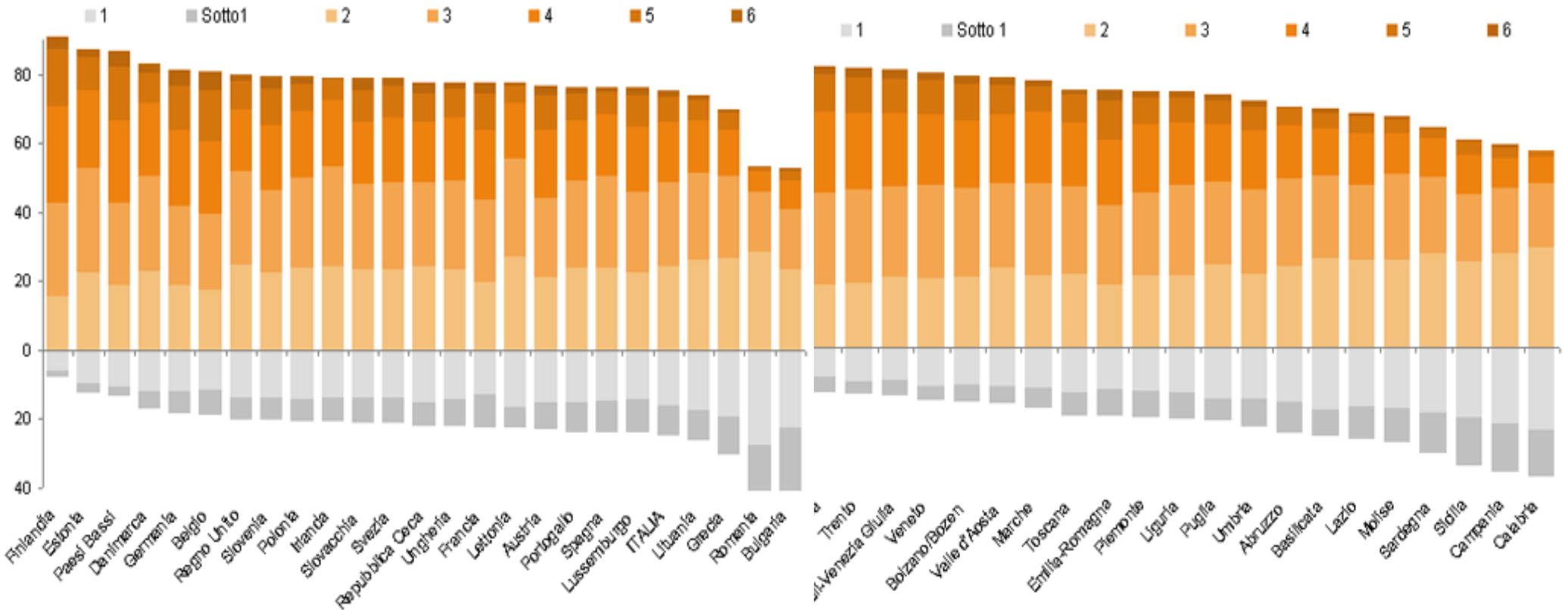
Biostatistica
 Epidemiologia
 Demografia
 Econometria
 Statistica economica
 Statistica ambientale
 Geostatistica
 Psicometria

PREVISIONI DELLA POPOLAZIONE RESIDENTE
Anni 2015-2050, migliaia



Oggi: applicazioni

Noi Italia 2013: 100 statistiche per capire il paese in cui viviamo.
STUDENTI 15ENNI PER LIVELLO DI COMPETENZA IN MATEMATICA (2009)



Quali sono le performance dei nostri studenti?

Oggi: applicazioni



La tua opinione
per noi è
importante

[Vai al questionario](#)



15° CENSIMENTO GENERALE
DELLA POPOLAZIONE
E DELLE ABITAZIONI
2011

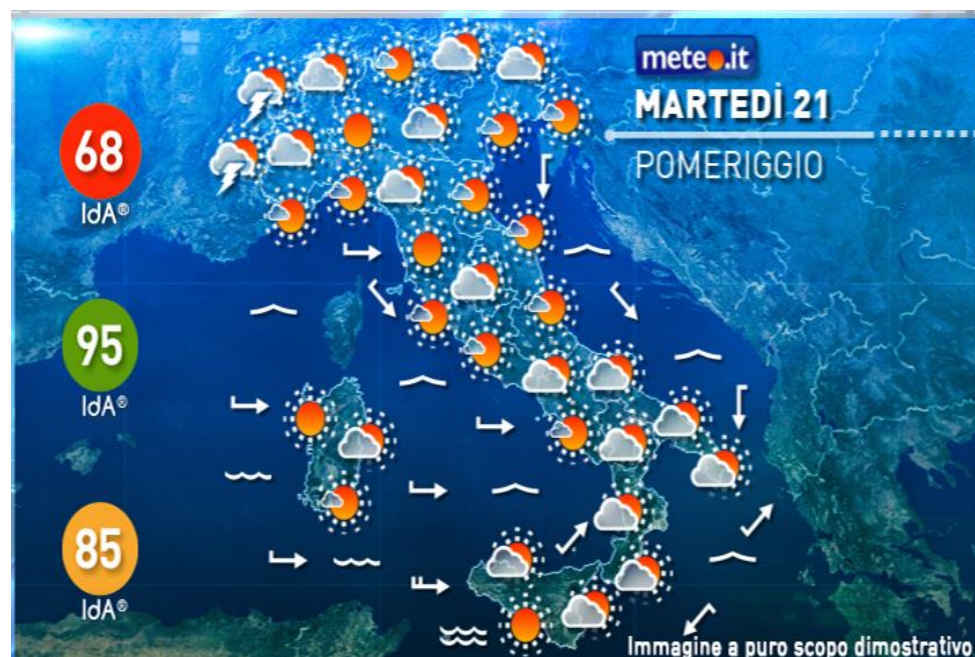
Registrati
accedi all'area riservata



Il questionario è uno strumento
di valutazione / misura.



Oggi: applicazioni



28 luglio 2015
Rivoluzione nelle previsioni
meteo, arriva:

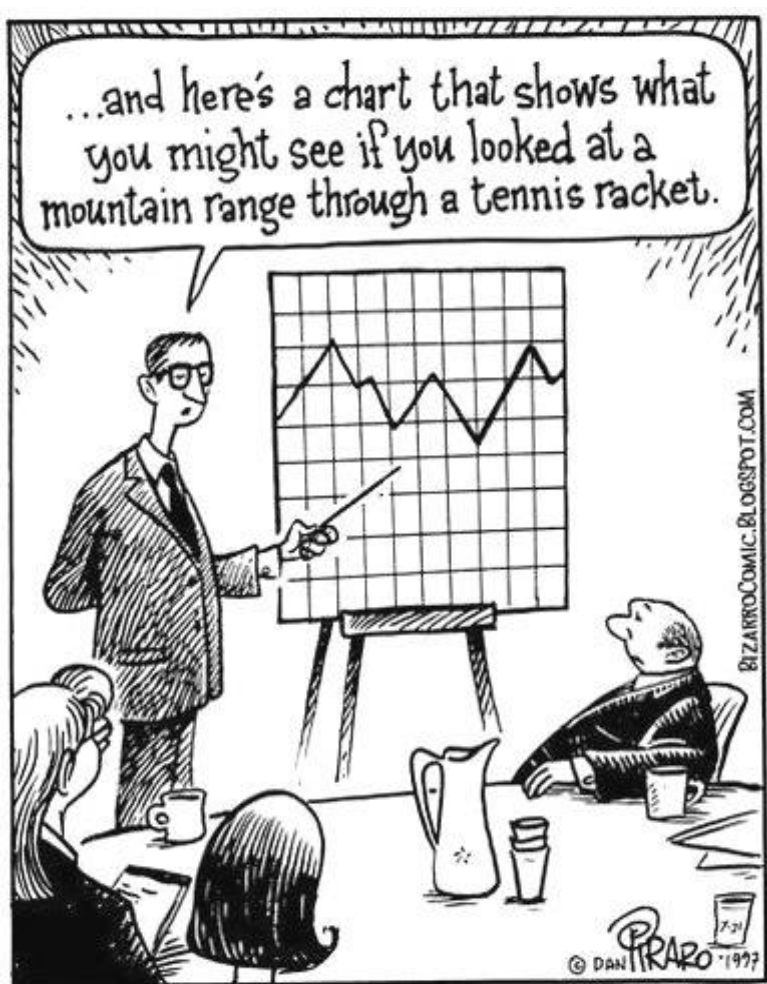
l'Indice di Affidabilità

Per la prima volta in Italia le previsioni diventano complete del livello di attendibilità, per un'informazione scientifica, precisa e onesta.

Di fronte a questo grande insieme di informazioni è importante sviluppare senso critico e saper scegliere!!!

Quindi cosa è la Statistica?

La parola STATISTICA deriva dal latino *Status*. Originariamente, infatti, la statistica si occupava delle rilevazioni ufficiali da parte di istituzioni statali. G. Ghislini nel 1589 indica la statistica come "descrizione delle qualità che caratterizzano e degli elementi che compongono uno Stato".



La STATISTICA è l'insieme delle metodologie finalizzate alla raccolta e all'analisi dei dati per lo studio di fenomeni collettivi che presentano un'attitudine a variare.

Più precisamente la statistica è l'insieme dei metodi per:

- ✓ progettare e pianificare come devono essere raccolti i dati;
- ✓ descrivere, sintetizzare e interpretare i dati (per la comprensione dei fenomeni), anche in condizioni di incertezza (sulla base di un'osservazione incompleta della realtà).

Statistica

- ✧ Statistica Descrittiva
- ✧ Calcolo delle Probabilità
- ✧ Statistica Inferenziale



Obiettivi

- ✧ Comprendere i termini fondamentali
- ✧ Comprendere come/quando gli strumenti statistici possono essere utilizzati di fronte ai dati reali
- ✧ Utilizzo del software (es. Excel)
- ✧ Senso critico rispetto ai risultati ottenuti



Come acquisire i dati



Rilevazione Sperimentale

Usata in medicina, fisica, chimica, *etc.*

Considera un gruppo di lavoro e l'utilizzo di un gruppo di controllo.

Rilevazione Osservazionale

Usata nelle scienze sociali, psicologiche, *etc.* attraverso indagini di mercato, sondaggi.

Indagine statistica.

Indagine statistica



Fenomeno da indagare



Produzione e raccolta di dati



Elaborazione, analisi e sintesi
dei dati

La statistica esiste dove ci sono i dati.

Il suo compito è quello di estrarre dai dati
Informazioni
utili a prendere **decisioni!**

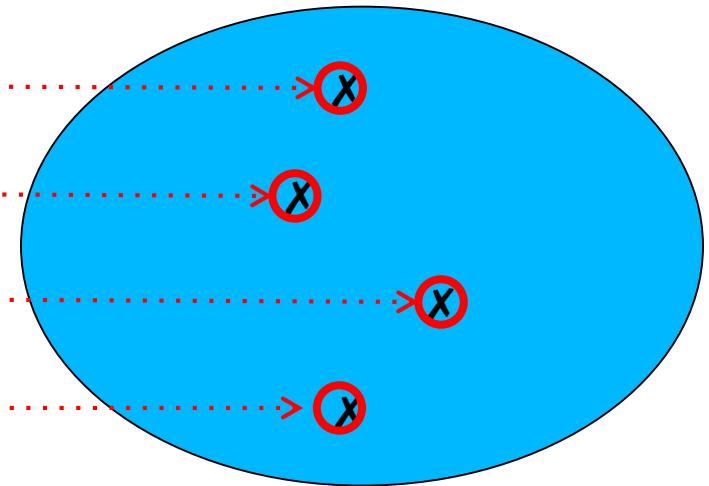
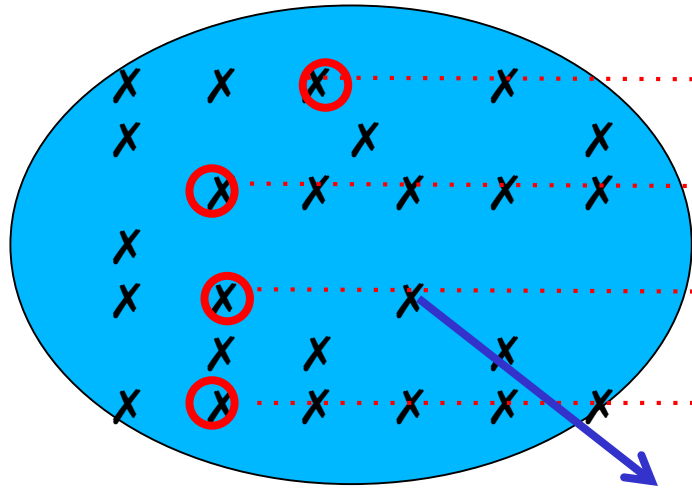
Indagine statistica

Censuaria

Campionaria

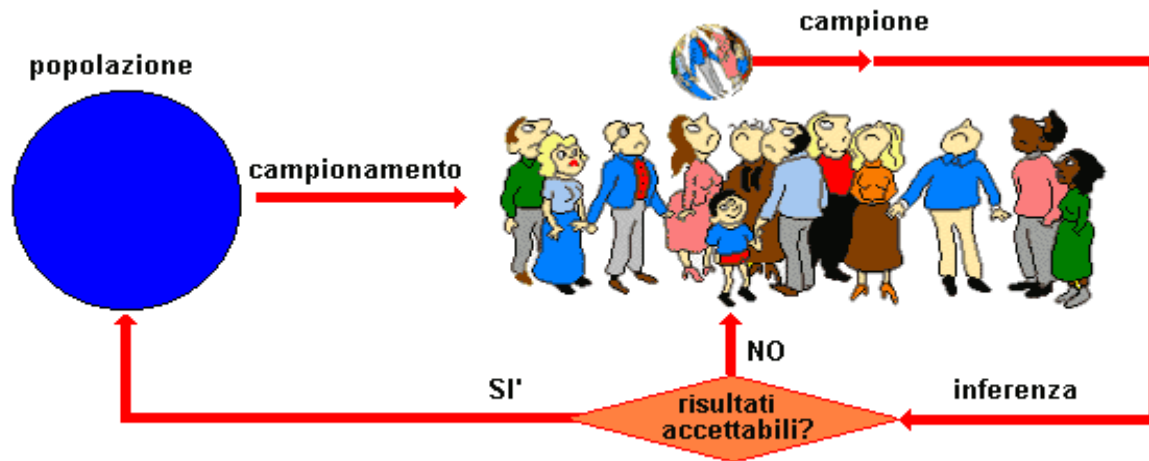
Popolazione su cui è presente il fenomeno X

Campione



Unità statistica su cui è presente la manifestazione x

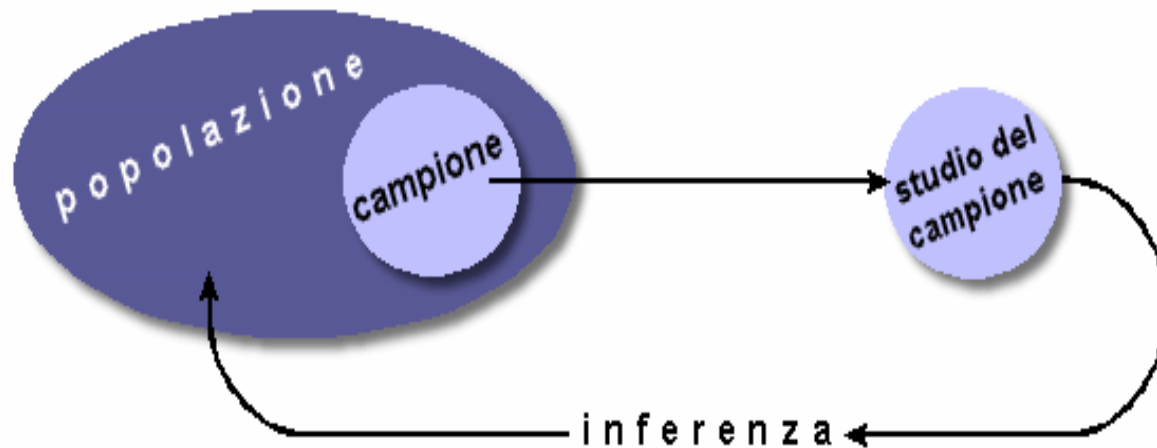
Statistica inferenziale



STATISTICA DESCRITTIVA



INFERENZA STATISTICA



Statistica – Probabilità

Indagine censuaria → la statistica fornisce una sintesi quantitativa dei fenomeni studiati.

Indagine campionaria → la prospettiva dell'analisi muta: non interessa il campione in sé, ma la popolazione che esso rappresenta.

La statistica fornisce i metodi con cui estendere le informazioni raccolte sul campione alla popolazione.

Qui entra in gioco il **CALCOLO DELLE PROBABILITA'**

Statistica – Probabilità

Teoria della probabilità → permette di risolvere il problema diretto: conoscendo la struttura della popolazione si deduce la probabile struttura del campione.

Qual è il comportamento potenziale di un campione estraibile da una popolazione nota?

Inferenza statistica → permette di risolvere il problema inverso: descrivere la struttura della popolazione a partire dal campione osservato.

L'inferenza presuppone che sia risolto il problema diretto.

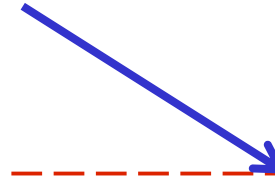
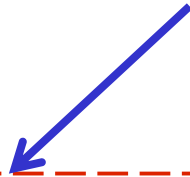
*Con quale precisione si possono descrivere le caratteristiche di una popolazione sulla base delle informazioni osservate sul campione?*²¹

Statistica descrittiva

La statistica descrittiva si articola in:

- **Univariata** → descrive il comportamento di un carattere
- **Bivariata** → studia la relazione che intercorre tra due caratteri
- **Multivariata** → studia la relazione che intercorre tra tre o più caratteri

Statistica Descrittiva



Univariata

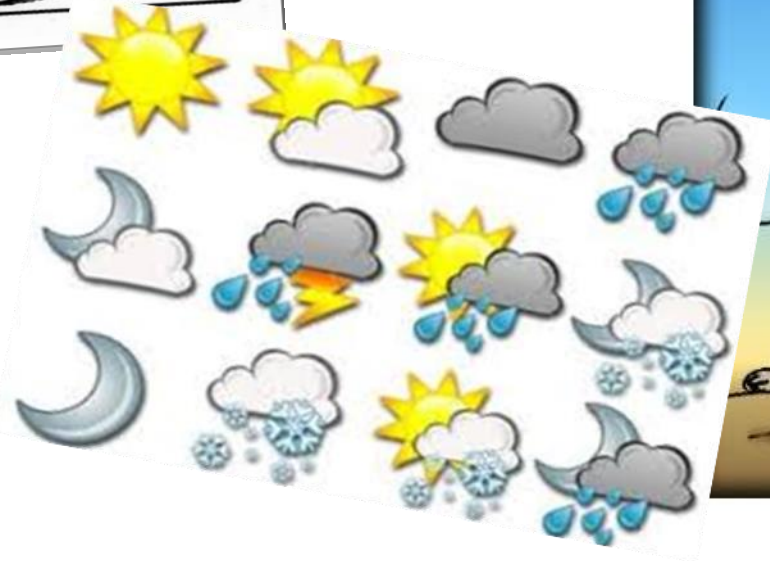
- ➔ Grafici
- ➔ Medie
- ➔ Numeri Indici
- ➔ Variabilità
- ➔ Asimmetria
- ➔ Concentrazione

Bivariata

- ➔ Associazione
- ➔ Correlazione
(non è fissata la relazione di dipendenza)
- ➔ Regressione
(è fissata la relazione di dipendenza)

Errore

Calcolo delle probabilità



Concetti base delle probabilità

- ✧ Qualche cenno storico
- ✧ Esperimento casuale
- ✧ Spazio campionario, evento elementare, evento
- ✧ Definizione di probabilità
- ✧ Concezione classica (a priori), frequentista (a posteriori) e soggettivista della probabilità
- ✧ Teoremi della probabilità
- ✧ Probabilità condizionata ed eventi indipendenti
- ✧ Teorema delle probabilità totali e di Bayes

Concezioni della probabilità

Quale significato dare alla parola **PROBABILITÀ**?

Concezione classica

La probabilità di un evento A è il rapporto tra il numero dei casi favorevoli al verificarsi di A e il numero totale dei casi possibili, ammesso che questi siano equiprobabili. L'equiprobabilità dei risultati possibili è molto importante, infatti se questi non hanno tutti la medesima probabilità la definizione non può essere applicata. Per cui:

$P(A) = \frac{n(A)}{n(\Omega)}$ dove $n(A)$ indica il volte in cui si presenta il risultato A

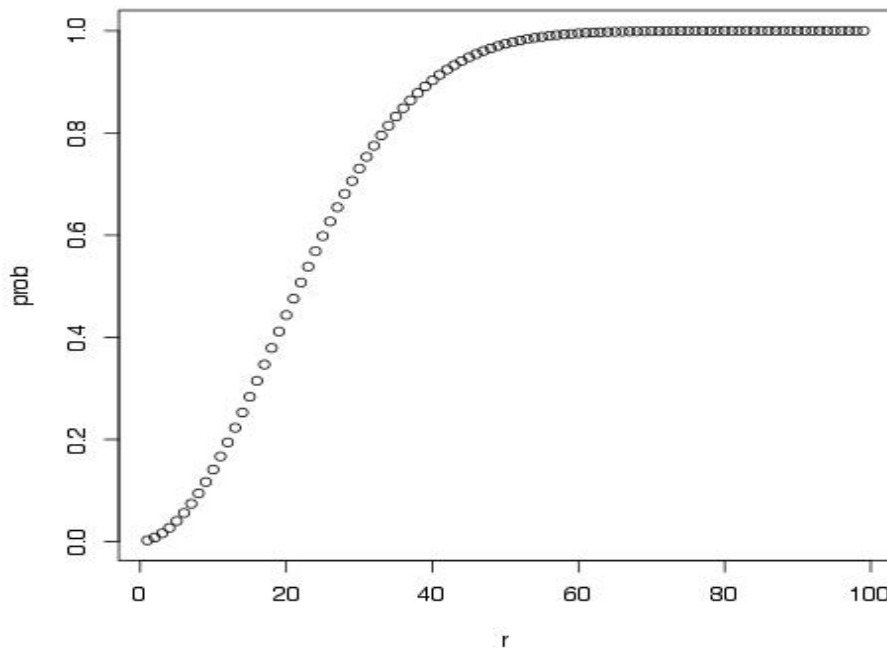
nell'insieme Ω dei risultati possibili.

Il paradosso del compleanno (Richard Von Mises, 1939)

Quale è la probabilità che almeno due persone (qualsiasi) in un gruppo di r persone compiano gli anni nello stesso giorno (evento A)? Si consideri un anno composto da 365 giorni.

Ragiono sull'evento complementare "tutte le r persone hanno compleanni in giorni diversi".

$$p(A) = 1 - p(\overline{A}) = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - r + 1)}{365^r}$$



Bastano 23 persone affinché la probabilità sia maggiore di 0.5.
E con 41 persone si supera 0.9.

Concezioni della probabilità

Concezione frequentista

Secondo questa interpretazione, la probabilità di un evento A è il limite della frequenza relativa con cui A si verifica in una lunga serie di prove ripetute sotto condizioni simili. Per cui:

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n} \quad \text{dove } n(A) \text{ indica il numero di volte in cui si verifica il}$$

risultato A in un insieme molto grande n di prove ripetute in modo indipendente.

Alla base di questa teoria si ha la *legge empirica del caso*, secondo la quale in una serie di prove di un dato esperimento, ripetuto un grande numero di volte in circostanze il più possibile simili, ciascuno degli eventi possibili si manifesta con una frequenza che è pressappoco uguale alla sua probabilità. L'approssimazione ovviamente cresce col crescere del numero delle prove.

Concezioni della probabilità

Problema:

Qual è la probabilità che una squadra di calcio di serie A, ad esempio la Juventus, vinca il campionato di calcio?

La serie A è formata da 20 squadre.

La **concezione classica** presuppone che tutte le squadre abbiano la medesima probabilità di vittoria, quindi $P(\text{Juventus}) = P(\text{Atalanta}) = 1 / 20$!!! **NO**

Secondo la **concezione frequentista** la probabilità di vittoria è data dal rapporto tra il numero di vittorie ottenute (32) e il numero di campionati giocati (111):

$$P(\text{Juventus}) = 32 / 111$$

Ma allora una squadra che non ha ancora vinto il campionato, non lo potrà mai vincere !!! **NO Non possiamo usare nessuna delle due concezioni**

Concezioni della probabilità

Concezione soggettivista

La probabilità è la valutazione che il singolo soggetto può coerentemente formulare, in base alle proprie conoscenze, del grado di avverabilità di un evento.

La probabilità viene pertanto identificata con un prezzo, è il prezzo che un soggetto ritiene equo pagare per:

- ricevere 1 se A si verifica,
- ricevere 0 se A non si verifica;

se A si verifica il guadagno è pari a $1 - P(A)$.

Anche per la concezione soggettivista valgono gli assiomi di Kolmogorov !!!

Concezioni della probabilità

Padre della concezione soggettivista è Bruno de Finetti (1906 - 1985) secondo il quale:

«non ha senso parlare della probabilità di un evento se non in relazione all'insieme di conoscenze di cui una persona dispone. La probabilità soggettiva è quindi un aiuto per dare un'attendibile misura di ciò che non si può misurare oggettivamente.»

La concezione è basata sui **principi di coerenza**, per cui:

- le probabilità devono essere attribuite in modo tale che non sia possibile ottenere una perdita certa o una vincita certa;
- le probabilità associate agli eventi non possono essere modificate;
- il soggetto deve attribuire la stessa probabilità ad eventi simili.

Probabilità condizionata e indipendenza

Se A e B sono due eventi dello spazio campionario Ω e $P(A) > 0$, allora la probabilità condizionata di B dato (rispetto ad) A è pari a:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Se lo spazio campionario è finito, la definizione di probabilità condizionata trova una piena giustificazione intuitiva: poiché si assume che l'evento A si sia già verificato, si tratta di determinare la probabilità dell'evento $A \cap B$ prendendo come nuovo spazio campionario l'evento A . Agli eventi elementari di A si devono allora attribuire delle nuove probabilità la cui somma sia 1.

Difficoltà per gli studenti

- ✧ Unione ed intersezione,
- ✧ Eventi incompatibili (disgiunti) ed eventi indipendenti,
- ✧ Probabilità dell'evento A e probabilità dell'evento $A|B$.

Teoremi: probabilità totali e Bayes

Teorema delle probabilità totali

Dati un evento A ed un insieme finito di eventi E_i (ad esempio k) a due a due incompatibili, tali che la loro unione dia Ω , allora

$$P(A) = \sum_{i=1}^k P(A \cap E_i) = \sum_{i=1}^k P(A|E_i) \cdot P(E_i).$$

Teorema di Bayes

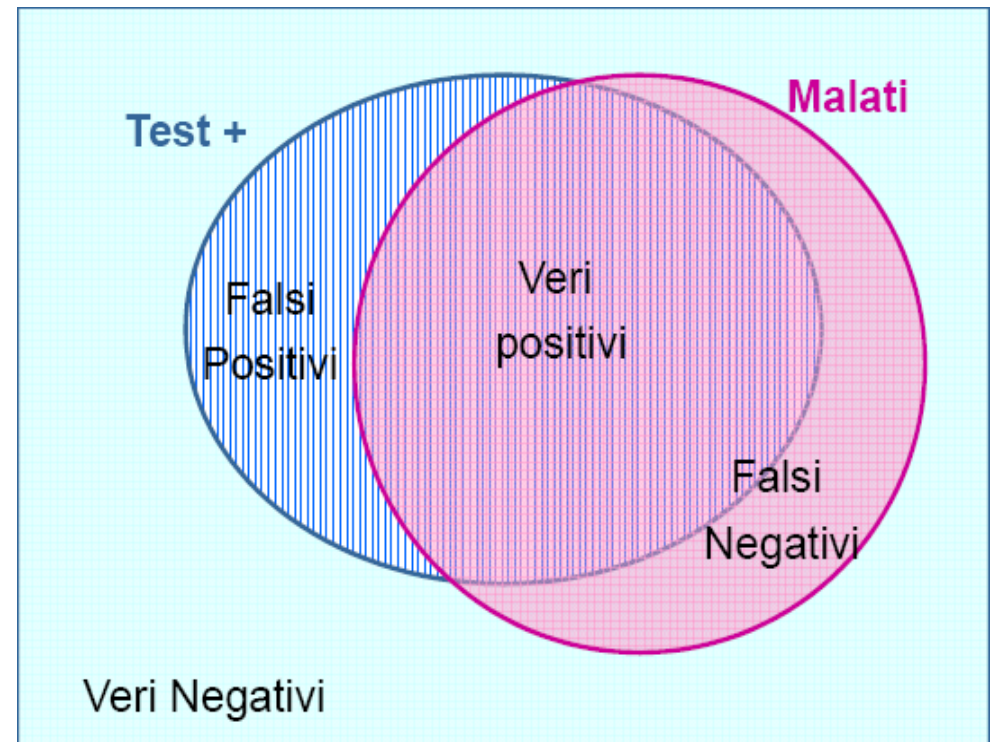
Dati un evento A ed un insieme finito di eventi E_i (ad esempio k) a due a due incompatibili, tali che la loro unione dia Ω , allora

$$P(E_i|A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(A|E_i) \cdot P(E_i)}{\sum_{i=1}^k P(A|E_i) \cdot P(E_i)}.$$

Esempio teorema di Bayes: test diagnostico

Si supponga che in una data popolazione il 10% degli individui sia affetto da una determinata patologia. Per diagnosticare la presenza della patologia viene effettuato un test ematico. E' noto che il test risulta negativo anche per il 5% dei malati (falsi negativi), mentre risulta positivo nel 9% dei sani (falsi positivi). Se un individuo risulta positivo al test, quale è la probabilità che esso sia effettivamente malato?

	malati	sani
Test +	VP	FP
Test -	FN	VN



Esempio teorema di Bayes: test diagnostico

Sensibilità del test:

probabilità che un test fornisca esiti positivi (T+) nei soggetti malati (M+):

$$p(T+|M+) = \frac{p(T+, M+)}{p(M+)}$$

La sensibilità indica la bontà del test nell'identificare i soggetti malati; più il test è sensibile, minore è la probabilità di ottenere un falso negativo.

Specificità del test:

probabilità che un test fornisca esiti negativi (T-) nei soggetti sani (M-):

$$p(T-|M-) = \frac{p(T-, M-)}{p(M-)}$$

La specificità indica la bontà del test nell'identificare i soggetti sani; più il test è specifico, minore è la probabilità di ottenere un falso positivo.

Esempio teorema di Bayes: test diagnostico

Valore predittivo di un esito positivo al test:

probabilità di osservare la presenza della malattia in un soggetto con esito positivo:

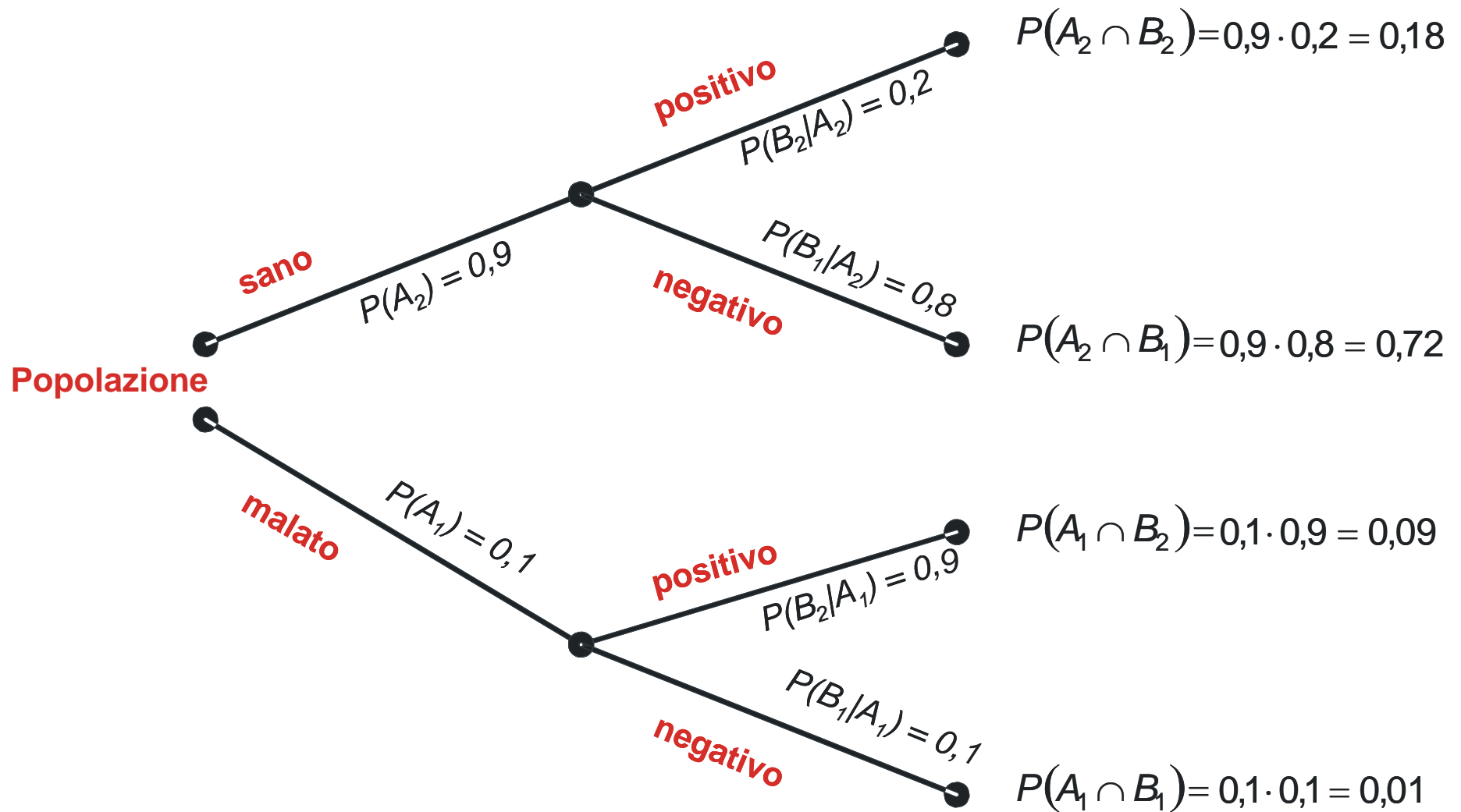
$$p(M^+|T^+) = \frac{p(T^+|M^+)p(M^+)}{p(T^+|M^+)p(M^+) + p(T^+|M^-)p(M^-)}$$

Valore predittivo di un esito negativo al test:

probabilità di osservare l'assenza della malattia in un soggetto con esito negativo:

$$p(M^-|T^-) = \frac{p(T^-|M^-)p(M^-)}{p(T^-|M^-)p(M^-) + p(T^-|M^+)p(M^+)}$$

Esempio teorema di Bayes: test diagnostico

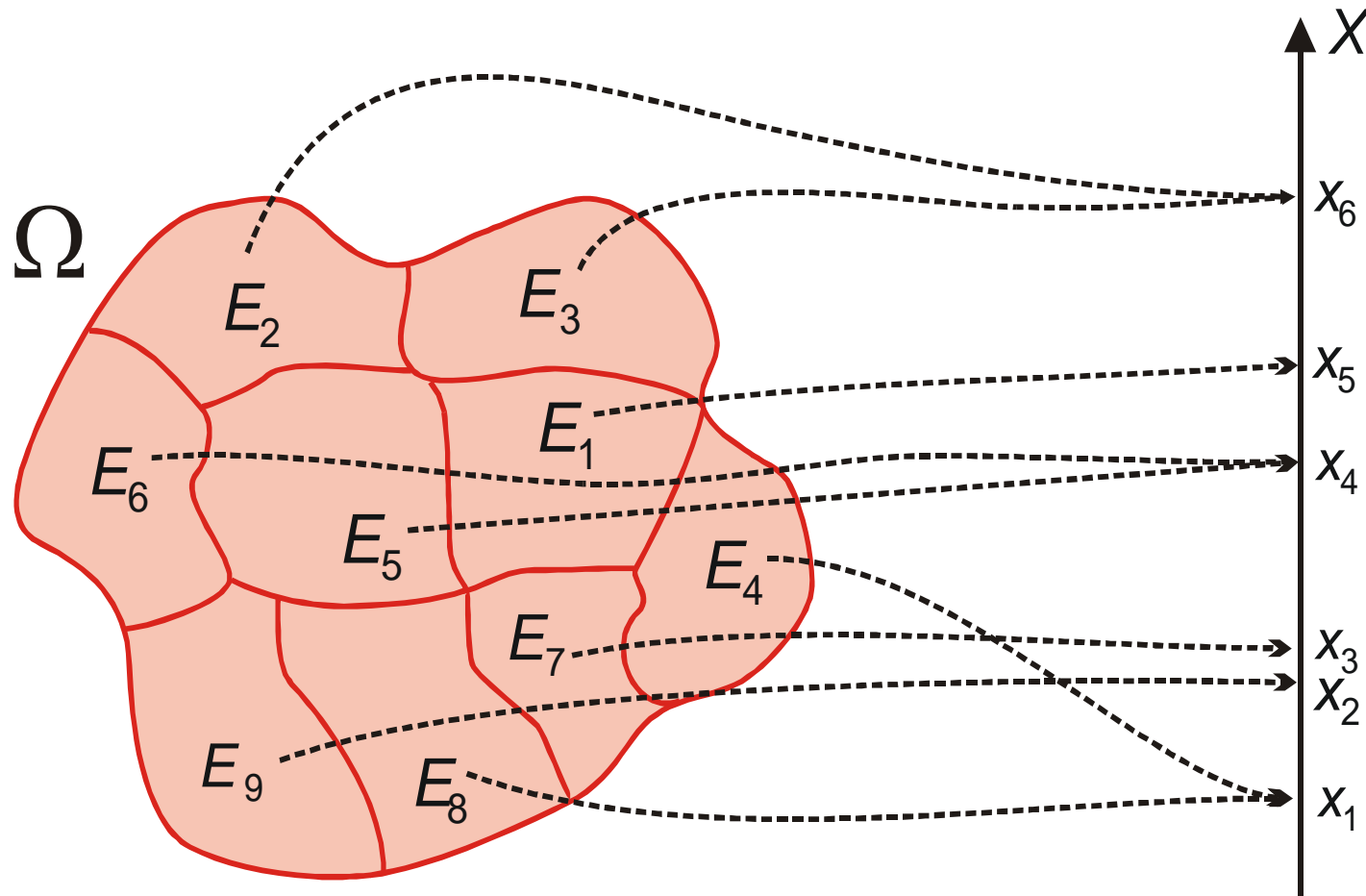


Variabili Casuali o Numeri Casuali

- ✧ Variabili casuali discrete e continue
- ✧ Funzione di ripartizione, di probabilità, di densità
- ✧ Valore atteso e varianza
- ✧ Variabile casuale di Bernoulli
- ✧ Variabile casuale Binomiale
- ✧ Variabile casuale di Poisson
- ✧ Variabile casuale normale
- ✧ Teorema del limite centrale con applicazione alla v.c. binomiale





































Definizione di variabile casuale

Una variabile casuale X è una funzione definita sullo spazio campionario Ω che associa ad ogni evento $E_i \subset \Omega$ uno e un solo numero reale.



Esempio: v.c. discreta

Costruzione della v.c. $X =$ "somma dei punteggi nel lancio di due dadi".

Ω		$\rightarrow X = 2$	X
			$\rightarrow X = 3$	
				$\rightarrow X = 4$	
					$\rightarrow X = 5$	
						$\rightarrow X = 6$	
							$\rightarrow X = 7$	
						$\rightarrow X = 8$	
					$\rightarrow X = 9$	
				$\rightarrow X = 10$	
			$\rightarrow X = 11$	
		$\rightarrow X = 12$	

Esempio: v.c. discreta

X	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

$$E(X) = 2 \frac{1}{36} + 3 \frac{2}{36} + 4 \frac{3}{36} + 5 \frac{4}{36} + 6 \frac{5}{36} + 7 \frac{6}{36} + 8 \frac{5}{36} + 9 \frac{4}{36} + 10 \frac{3}{36} + 11 \frac{2}{36} + 12 \frac{1}{36} = 7$$

$$\text{Var}(X) = 2^2 \frac{1}{36} + 3^2 \frac{2}{36} + 4^2 \frac{3}{36} + 5^2 \frac{4}{36} + 6^2 \frac{5}{36} + 7^2 \frac{6}{36} +$$

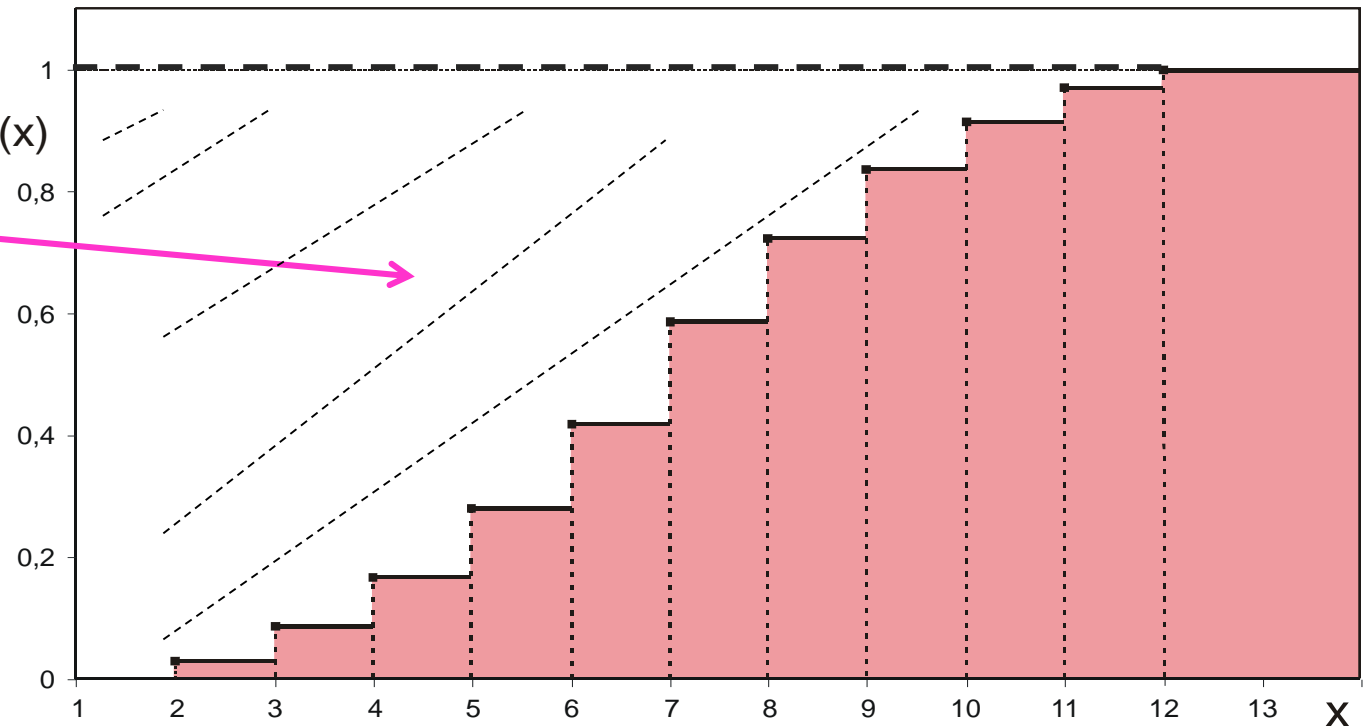
$$+ 8^2 \frac{5}{36} + 9^2 \frac{4}{36} + 10^2 \frac{3}{36} + 11^2 \frac{2}{36} + 12^2 \frac{1}{36} - 7^2 = 5,83$$

$$\sigma(X) = \sqrt{5,83} = 2,415$$

Esempio: v.c. discreta

X	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	1

Area
tratteggiata = ?



L'area tratteggiata è pari al valore atteso della v.c. X .

Variabili casuali discrete in Excel

- `distrib.binom(num_successi; prove; probabilità_s; cumulativo)`

Se `cumulativo=FALSE` ottengo la funzione di probabilità, altrimenti la funzione di ripartizione.
Se `prove = 1` ho la v.c. di Bernoulli.

- `distrib.poisson(x; media; cumulativo)`

Se `cumulativo=FALSE` ottengo la funzione di probabilità, altrimenti la funzione di ripartizione.

V.c. Normale in Excel

- `distrib.norm(x; media; dev_standard; cumulativo)`

Se `cumulativo=FALSE` ottengo la funzione di densità in un punto x , altrimenti la funzione di ripartizione $p(X \leq x)$.

- `inv.norm(probabilità; media; dev_standard)`

Restituisce il valore x tale per cui:
 $p(X \leq x) = \text{probabilità}$.

Variabili casuali con PQRS

PQRS è un software, scaricabile all'indirizzo:

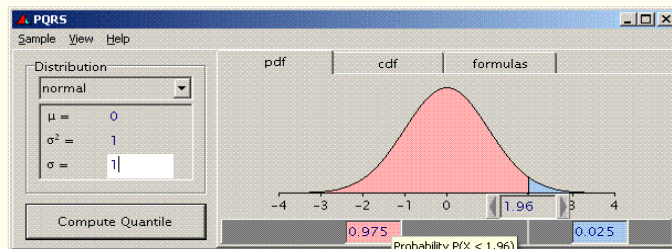
<http://members.home.nl/sytse.knypstra/PQRS/>

che permette di eseguire le operazioni con le variabili casuali.

PQRS

(Probabilities, Quantiles and Random Samples)

PQRS is a tool for calculating probabilities and quantiles associated with a large number of probability distributions. In addition, random samples can be drawn and stored to a file. Quantiles and probabilities are displayed and edited in their natural position relative to the probability (density) graph. This makes PQRS very easy to use.



PQRS was originally designed as a tool for students using computer aided instruction of statistics in order to make printed tables obsolete.

Compared to printed statistical tables PQRS gives you:

- more distributions,
- higher precision (up to 12 decimals),
- results for any parameter value (within a certain range) and for any specified x-value or probability.

Compared to similar computer programs PQRS gives you in addition:

- the probability (density) graph,
- the graph of the cumulative distribution function,
- a clear design: edit fields and results are displayed in their natural position relative to the probability (density) graph,
- formulas for the probability (density) function, mean and variance,
- greek symbols for parameters like alpha, beta, mu and sigma.

Additional advantages:

- quantiles are also given for discrete distributions,
- left side probabilities $P(X < \text{value})$, right side probabilities $P(X > \text{value})$ as well as $P(X = \text{value})$ are simultaneously displayed (the latter only if non-zero),
- random samples of any size can be drawn and written to a file.

PQRS may be freely copied and distributed, provided that files have not been edited or altered in any way, and provided that no charge is made for distributing PQRS. Requests to use PQRS for any commercial purpose must be directed to the author, Sytse Knypstra. His e-mail address can be found in the Help -> About window of PQRS.

To get a free copy of PQRS you can download the compressed file [PQRS.zip](#) (0.7 MB), unzip it and start PQRS.exe. PQRS runs under Windows. Under Linux you can run it through Wine.

Ed ora...

