



Pontificia **U**niversità **L**ateranense

Questioni di storia del Pensiero Scientifico

Il Novecento della storia della scienza

Facoltà di Filosofia - Corso 50647

M-ST0/05- Storia delle scienze e delle tecniche

Prof.ssa Flavia Marcacci – www.flaviamarcacci.it

Probabilità e statistica.

Cenni storici ed elementi introduttivi
di calcolo.

PARTE II

Breve premessa storica

Il termine “statistica” fu introdotto nel XVII secolo col significato di “scienza dello stato”, per raccogliere e ordinare informazioni utili all’amministrazione pubblica: dati sui commerci, sulla distribuzione della ricchezza, etc...

Scopi della statistica: raccolta, rappresentazione ed elaborazione dei dati

In svariate discipline (medicina, economia, sociologia,...) la statistica viene usata come strumenti utili all’elaborazione di ipotesi.

Un singolare caso storico: le assicurazioni di rendita

- Situazione generale: A e B si accordano perché A paghi a B una somma forfettaria e B rimborsi A in rate annuali.
 - Prestito con interesse = A pretende di ricavare un utile dal prestito fatto a B
 - Assicurazione di rendita = A vuole un reddito sicuro per un periodo stabilito. Questa soluzione era usata dai governi per assicurarsi denaro in contante, evitando di praticare l'usura. Ce ne sono molti tipi, da quelle perpetue (prestiti per cui si paga al titolare un interesse annuo) a quelle limitata a n anni e vitalizie.

- Le tavole di Ulpiano (III sec. d. C.) fanno vedere che l'Impero romano aveva capito che nel caso di rendita vitalizia il costo doveva dipendere dall'età (con un tasso massimo di 30 a 1 se assunta la rendita a 20 anni d'età) era per evitare l'usura
- L'Europa del 500 proponeva soluzioni molto più vantaggiose: una legge inglese del 1540 dichiara una rendita governativa (per cui si paga S e si corrisponde per tutta la vita un guadagno annuo A) equivalente a un prestito settennale. Considerando l'alta mortalità infantile e che molti contraenti erano anziani, nel suo complesso c'erano le condizioni per lo stato per guadagnare.

- Nel 1687 Mabbut, sotto la supervisione di Newton, elaborò delle tavole per le rendite vitalizie congiunte dove k vite varranno quanto un prestito di $1+k(n-1)$ anni.
- Su calcoli di questo genere incidevano fortemente due fattori: la curva di mortalità relativa alla fascia d'età del contraente la rendita e il tasso di interesse sui prestiti a lungo termine.

- Il problema del calcolo delle rendite aveva attratto nomi illustri come Euler o Halley.
- John de Witt, 1671: propone una sua soluzione agli Stati generali d'Olanda e Frisia Occidentale, in collaborazione con Hudde. Di fondo usò la probabilità tenendo conto della curva di mortalità (considerata uniforme tra i 4 e i 54 anni), basate su rilevazioni statistiche. E su queste rilevazioni si avviò un profondo processo di studio e comprensione.

Cenni di statistica

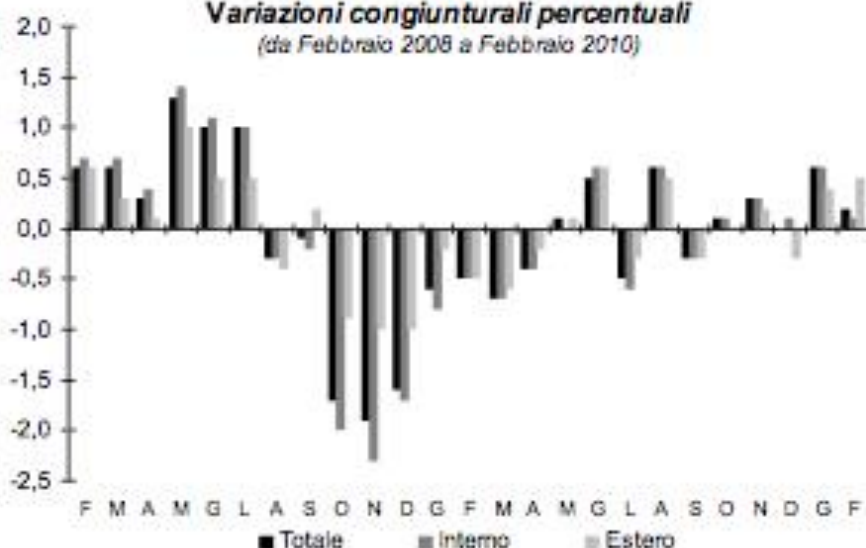
- **Unità statistica** minima unità della quale si raccolgono i dati
- **Popolazione** insieme delle unità statistiche oggetto di studio
- **Caratteri** proprietà che sono oggetto di rilevazione. Possono essere qualitative (es. stato civile, colore dei capelli,...) o quantitative (discrete o continue), anche se spesso è difficile precisare le distinzioni.
- **Campione** parte della popolazione, utile a impostare problemi di **statistica induttiva**. Si comprende bene la difficoltà di costruire campioni attendibili.
- **Statistica descrittiva** applicata all'intera popolazione

Tabella 1. Indice dei prezzi alla produzione dei prodotti industriali (base 2005=100).
Febbraio 2010 (a)

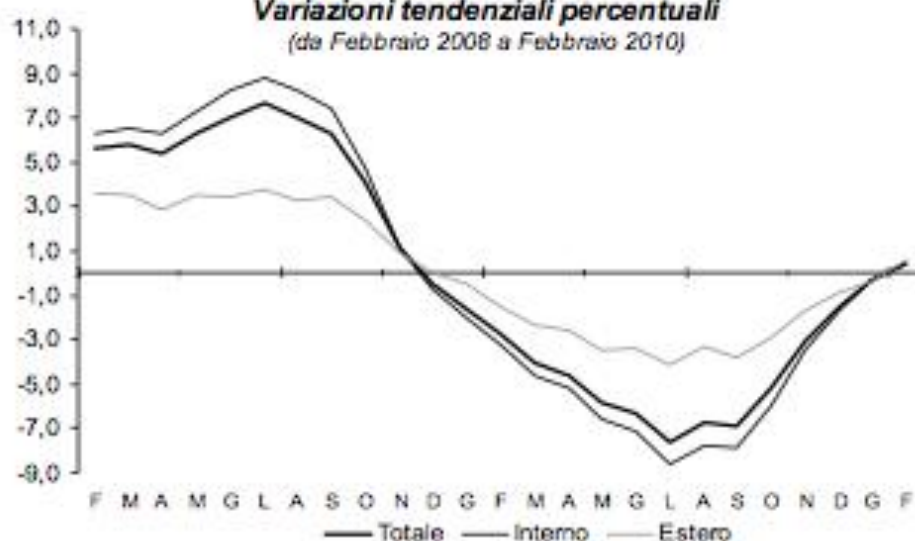
	INDICI	VAR. CONGIUNTURALI		VAR. TENDENZIALI	
		Feb 10	Dic 09-Feb 10	Feb 10	Gen-Feb 10
		Gen 10	Set-Nov 09	Feb 09	Gen-Feb 09
Totale	108,8	+ 0,2	+ 0,6	+ 0,4	0,0
<i>Interno</i>	109,9	+ 0,1	+ 0,7	+ 0,4	+ 0,1
<i>Estero</i>	105,5	+ 0,5	+ 0,2	+ 0,6	+ 0,1

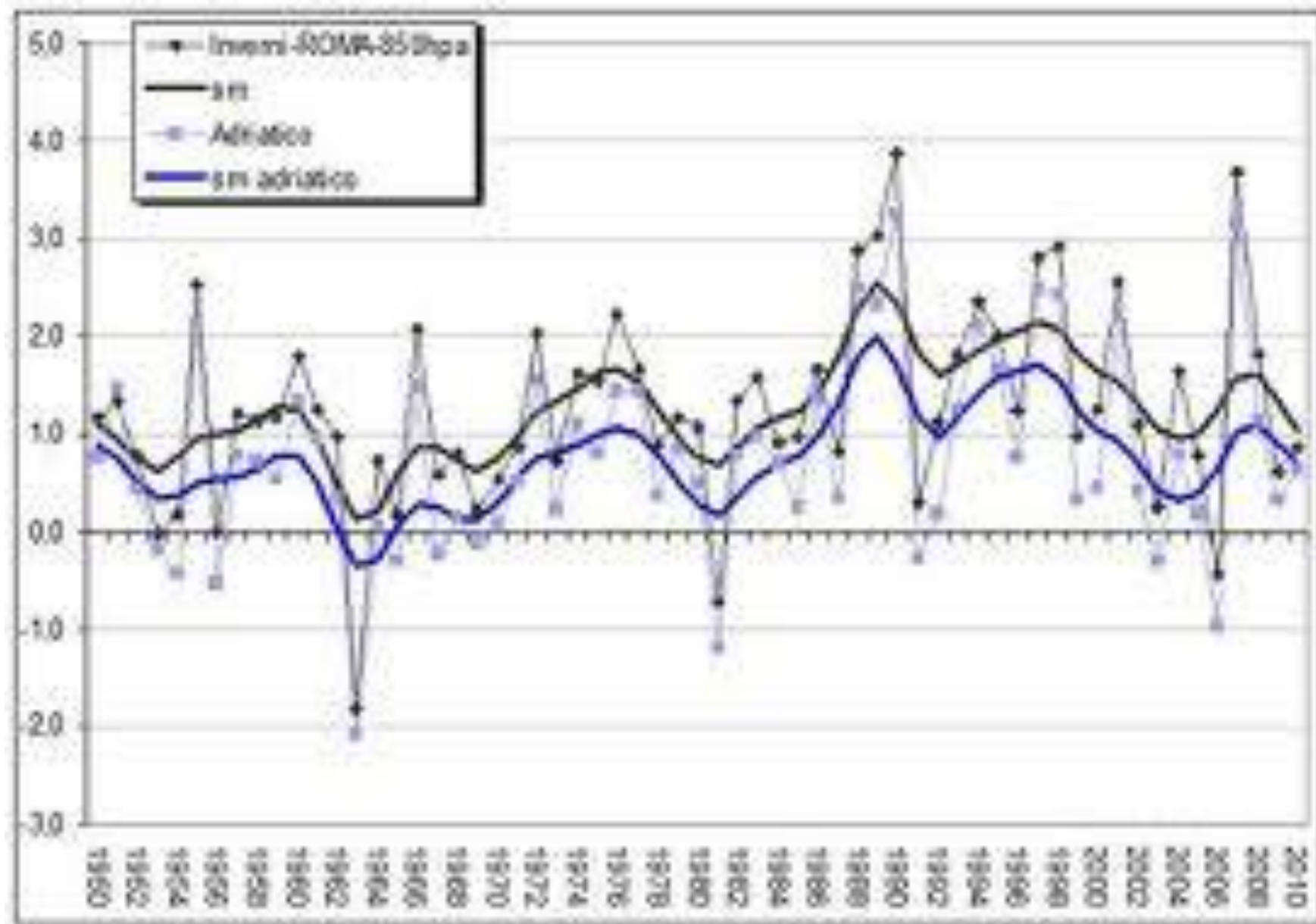
(a) Dati provvisori

Indice Totale, Interno e Estero
Variazioni congiunturali percentuali
(da Febbraio 2008 a Febbraio 2010)



Indice Totale, Interno e Estero
Variazioni tendenziali percentuali
(da Febbraio 2008 a Febbraio 2010)





Raccogliere i dati, costruire grafici

- La raccolta dei dati facilita di molto le operazioni statistiche se ben organizzata. Vediamo subito un esempio.

Un esempio

Il sistema dei pediatri sentinella è stato lanciato nel 2000, e ha coperto circa il 4% della popolazione nazionale sotto i 15 anni di età. Nonostante l'adesione volontaria dei pediatri alla rete di sorveglianza, questo sistema si è rivelato molto più sensibile della rete di notifiche: l'incidenza stimata di rosolia nei bambini secondo la rete di pediatri nel 2000 era dalle 5 alle 6 volte più alta rispetto alle notifiche. Nel 2002 la rete ha stimato un'incidenza elevata con circa 300 casi per 100mila bambini, soprattutto tra i 10 e i 14 anni di età. Questa rete, per sua natura, non dà alcuna informazione sui casi di rosolia tra gli adulti.

Sono disponibili i dati epidemiologici sui casi di rosolia in Italia dal 1996 al 2006: ogni anno vengono aggiornati in base ai casi notificati. L'andamento mostra una serie di oscillazioni, con un massimo di quasi 35 mila casi riportati nel 1997. A partire dal 2002, il trend vede una diminuzione del numero dei casi di rosolia, fino al minimo rappresentato dai 257 casi del 2006.

Rosolia in Italia



*Andamento del numero dei casi di rosolia in Italia dal 1996 al 2006
(fonte: ministero della Salute)*

Casi di rosolia in Italia dal 1996 al 2007, divisi per sesso. I dati del 2007 sono da considerarsi provvisori.

Anno	M	F	n.i.	Totale
1996	14146	7669	63	21878
1997	21037	13464	111	34612
1998	1820	1498	0	3318
1999	615	514	0	1129
2000	1506	1099	0	2605
2001	3169	1979	3	5151
2002	3824	2383	17	6224
2003	877	736	2	1615
2004	238	223	0	461
2005	136	160	1	297
2006	148	109	0	257
2007	286	236	0	522

Le medie

- Per riassumere un determinato insieme di dati occorrono strumenti capaci di offrire facili rappresentazioni. Si tratta delle medie, capaci di offrire dati riassuntivi. Si tratta di un'operazione che appartiene all'uomo in maniera innata, in quanto per natura siamo portati a cercare rappresentazioni sintetiche e unitarie.

Medie mobili

- **Moda:** è il carattere o il valore cui corrisponde la massima frequenza. Es. tra 2,3,4,4,4,5,6,6,6 ci sono due mode, ovvero 4 e 6
- **Mediana:** è il valore che occupa il posto di mezzo, quando i dati sono disposti in ordine crescente. Es.: sufficiente, buono, buono, ottimo. La mediana è buono
- **Moda e mediana** non sempre indicano in maniera efficiente i dati. Ad es. se si applicasse alle pagelle scolastiche

Medie ferme

- **Dati n dati dal valore X_1, X_2, \dots, X_n si dice:**
- **Media aritmetica semplice:** valore che si ottiene sommando i valori dei singoli dati diviso il numero dei dati $M_a = (X_1 + X_2 + \dots + X_n) / n$ Questo tipo di media ha senso quando si eseguono diverse misurazioni per la stessa grandezza e quando serve il valore tipico di una popolazione omogenea.
- **Media aritmetica ponderata:** la somma dei valori tiene conto dell'importanza di ogni valore.

$$M_p = (p_1 X_1 + p_2 X_2 + \dots + p_n X_n) / n$$

- **Media quadratica:** radice quadrata della media aritmetica dei quadrati dei vari valori.

$$M_q = \sqrt{\frac{X_1^2 + X_2^2 + \dots + X_n^2}{n}}$$

- **Media geometrica:** radice ennesima del prodotto

$$M_g = \sqrt[n]{X_1 X_2 \dots X_n}$$

- **Media armonica:** inverso della media aritmetica dei valori inversi

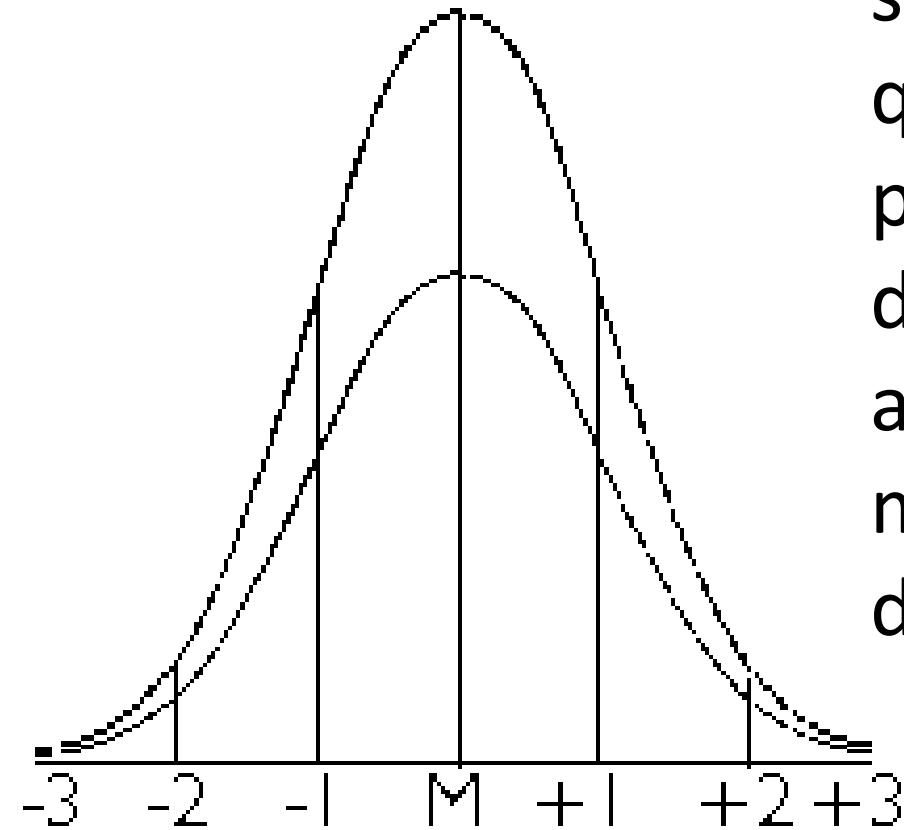
$$M_A = \frac{n}{\frac{1}{X_1} + \dots + \frac{1}{X_n}}$$

Rappresentare la variabilità dei dati

- Le medie forniscono un valore sintetico che però non è indicativo circa la variabilità dei dati. Per rappresentare tale variabilità occorrono degli indicatori:
- **Campo di variazione:** differenza fra il minimo e il massimo dei valori osservati. Si tratta di una misurazione molto poco precisa.
- **Scarti dal valore medio:** distanza tra il singolo valore e la media utilizzata. Si introduce anche lo **scarto semplice medio** dato dalla media dei valori degli scarti in valore assoluto.
- **Scarto quadratico medio:** media quadratica degli scarti

Distribuzione dei dati

- I più importanti indici di dispersione sono lo scarto semplice medio e lo scarto quadratico medio. A tal proposito è possibile avere distribuzioni di Gauss attorno allo stesso valore medio ma diverse per la differenza tra gli scarti:



Bibliografia

- I. Hacking, *L'emergenza della probabilità*, Il Saggiatore, Milano 1987
- M. Trovato – R. Manfredi, *Calcolo delle probabilità e statistica inferenziale*, Ghisetti & Corvi Editori, Milano 1997
- N. Doderò – P. Baroncini – R. Manfredi, *Elementi di matematica. Vol. 5*, Ghisetti & Corvi Editori, Milano 1991
- D. Palladino – S. Scotto – M. Frixione, *Matematica. 2A*, Principato, Milano 2000