



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**



Dipartimento di scienze economiche,
aziendali, matematiche e statistiche
"Bruno de Finetti"

Statistica

Introduzione

Francesco Pauli - Francesco Santelli

A.A. 2023/2024

Indice

Dettagli sul corso

Programma del corso

La statistica

Docenti

Francesco Pauli

Orario di ricevimento

si veda il [sito](#).

**Usate liberamente la chat di
teams o se preferite la mail.**

Il ricevimento è possibile in altri orari
su appuntamento.

Francesco Santelli

Prime 20 ore

Libri di testo e riferimento

Giuseppe Cicchitelli

Statistica: principi e metodi

Edizioni Pearson



È uscita recentemente la seconda edizione, che in più rispetto alla precedente ha la parte su web.



Libri di testo e riferimento

Pauli F., Torelli N., Trevisani M.

Statistica: esercizi ed esempi

Pearson Education, Italia



Molti esercizi sono comunque presenti nel testo di Cicchitelli.



Libri di testo e riferimento

David M. Diez, Christopher D. Barr,
Mine Çetinkaya-Rundel

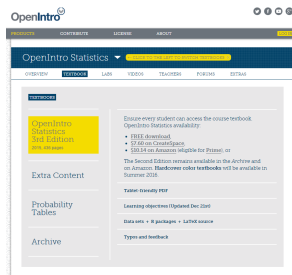
OpenIntro Statistics

Gratis in rete:

<https://www.openintro.org/stat/textbook.php>



Un bel testo, però in inglese.



Altre letture

Leonard Mlodinow

La passeggiata dell'ubriaco. Le leggi scientifiche del caso

Editore Rizzoli

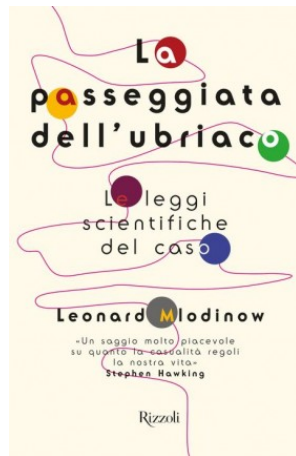


“Un saggio molto piacevole su quanto la casualità regoli la nostra vita”

Stephen Hawking



Riguarda più la probabilità che la statistica, è un'introduzione divulgativa (senza formule) ma con esempi reali per niente banali.



Altre letture

Taleb Nassim Nicholas

Il Cigno nero

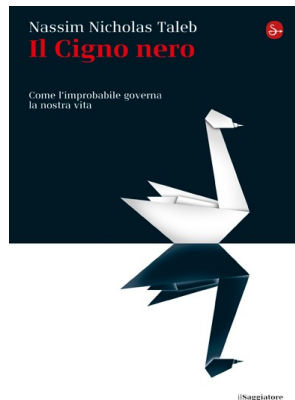
Il Saggiatore



Più complesso (comunque senza dettagli tecnici), e incentrato su fenomeni legati ai mercati finanziari, campo di cui l'autore è esperto, considerazioni molto attuali in tempi di crisi.



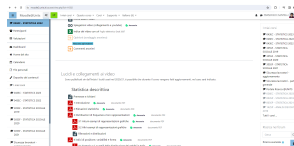
L'autore ha scritto altri tre libri, anche questi interessanti.



Pagina Moodle

Su moodle2.units.it trovate

- ▶ lucidi delle lezioni
- ▶ collegamenti ai video dell'anno scorso
- ▶ Esami passati
- ▶ un modulo per inviare anonimamente **commenti** al docente (critiche, suggerimenti, domande, richieste).



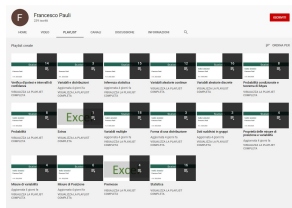
Spiegazioni video

Su [Moodle](#) trovate anche i collegamenti alle spiegazioni video che costituiscono parte integrante del corso.

- ▶ coprono quasi tutti gli argomenti che tratteremo in aula e alcuni che non tratteremo in aula.
- ▶ sono brevi e diretti al punto (è facile ritrovare un argomento specifico)
- ▶ tre tipi di video: teoria, esercizi carta e penna, esercizi in Excel

Usare video e lezioni

- ▶ Video come materiale di ripasso, in qualche caso approfondimento/integrazione
- ▶ lezioni in diretta con spazio all'interazione



A proposito di Excel

L'uso di Excel nel corso ha un'accoglienza mista: ad alcuni piace molto ad alcuni proprio per niente.



Excel come strumento è pervasivo, prima o dopo usarlo per lavoro sarà inevitabile.



La statistica si fa col computer, Excel non è il miglior software per la statistica, tuttavia per cose relativamente semplici è uno strumento utile.



Infine, da un punto di vista puramente didattico, permette di illustrare alcune procedure di calcolo in maniera più efficiente e precisa di quanto si possa fare con "carta e penna".

Esame

L'esame consiste di una prova scritta che consta di due parti

- ▶ La prima parte contiene domande più semplici, per un totale di 30 punti, risolvendo solo questa parte si ottiene un punteggio massimo pari a 21 (cioè se si risolvono correttamente domande per 21 o più punti, il punteggio ottenuto è 21), salvo bonus di cui a lucido successivo.
- ▶ La seconda parte contiene esercizi più difficili, il punteggio ottenuto in questa seconda parte si aggiunge al punteggio ottenuto nella prima parte (purché questa sia superata).

Modalità didattiche

Le lezioni privilegeranno l'interazione, useremo come canale Wooclap, che permette di creare presentazioni interattive.



Durante le lezioni vi proporrò alcuni (semplici) quiz, ma non solo, a cui potrete rispondere seduta stante collegandovi a una pagina web (con smartphone, tablet, computer)



Scopi/vantaggi:

- ▶ per voi un riepilogo immediato delle cose più importanti
- ▶ opportunità di ragionare subito sulle cose (non solo ascoltare)
- ▶ per me un riscontro su eventuali aspetti non chiari

Quiz in itinere **FORSE**

Verranno proposte periodicamente, durante le lezioni, quiz.

In base al punteggio ottenuto vi sarà una maggiorazione sul voto dell'esame fino a un massimo di quattro/sei punti (anche oltre i 21 in caso di svolgimento della sola prima parte del compito).

- ▶ La maggiorazione sarà valida per la sola sessione estiva (giugno/luglio)
- ▶ La maggiorazione dipenderà dalle risposte corrette ma anche della percentuale di domande a cui si è data risposta

Indice

Dettagli sul corso

Programma del corso

La statistica

Statistica descrittiva

Contenuti:

- ▶ Cosa sono i dati
- ▶ Rappresentazione dei dati
- ▶ Sintesi di una variabile
- ▶ Analisi della relazione tra variabili

Statistica descrittiva

Contenuti:

- ▶ Cosa sono i dati
- ▶ Rappresentazione dei dati
- ▶ Sintesi di una variabile
- ▶ Analisi della relazione tra variabili

Obiettivi:

- ▶ Saper interpretare delle informazioni di tipo statistico.
- ▶ Saper leggere/costruire una tabella o un grafico.
- ▶ Capire il significato (e soprattutto i limiti) delle informazioni di sintesi (medie, varianze ecc.)

Statistica descrittiva

Contenuti:

- ▶ Cosa sono i dati
- ▶ Rappresentazione dei dati
- ▶ Sintesi di una variabile
- ▶ Analisi della relazione tra variabili

Obiettivi:

- ▶ Saper interpretare delle informazioni di tipo statistico.
- ▶ Saper leggere/costruire una tabella o un grafico.
- ▶ Capire il significato (e soprattutto i limiti) delle informazioni di sintesi (medie, varianze ecc.)



Difficoltà: bassa

- ▶ contenuti matematici semplici: poco più delle 4 operazioni
- ▶ concetti intuitivi

Probabilità

Contenuti:

- ▶ Probabilità elementari, teorema di Bayes
- ▶ Variabili aleatorie, definizione, distribuzione di probabilità di una variabile aleatoria, v.a. discrete e continue, esempi notevoli (Binomiale, Poisson, Normale)

Probabilità

Contenuti:

- ▶ Probabilità elementari, teorema di Bayes
- ▶ Variabili aleatorie, definizione, distribuzione di probabilità di una variabile aleatoria, v.a. discrete e continue, esempi notevoli (Binomiale, Poisson, Normale)

Obiettivi:

- ▶ Ragionare in condizioni di incertezza quantificando il grado di incertezza
- ▶ Interpretare la quantificazione un guadagno (perdita) incerto.
- ▶ Prendere decisioni
 - ▶ mi conviene giocare al lotto?
 - ▶ ... o al casinò?
 - ▶ mi conviene giocare in borsa?
 - ▶ mi conviene assicurarmi contro ...

Probabilità

Contenuti:

- ▶ Probabilità elementari, teorema di Bayes
- ▶ Variabili aleatorie, definizione, distribuzione di probabilità di una variabile aleatoria, v.a. discrete e continue, esempi notevoli (Binomiale, Poisson, Normale)

Obiettivi:

- ▶ Ragionare in condizioni di incertezza quantificando il grado di incertezza
- ▶ Interpretare la quantificazione un guadagno (perdita) incerto.
- ▶ Prendere decisioni
 - ▶ mi conviene giocare al lotto?
 - ▶ ... o al casinò?
 - ▶ mi conviene giocare in borsa?
 - ▶ mi conviene assicurarmi contro ...



Difficoltà: medio-alta

- ▶ contenuti matematici più sofisticati: insiemistica, integrali
- ▶ concetti poco intuitivi

Inferenza statistica

Contenuti:

- ▶ Popolazione e campione: variabilità campionaria
- ▶ Il concetto di inferenza statistica
- ▶ Stima intervallare per media e proporzione
- ▶ Verifica d'ipotesi per media e proporzione
- ▶ Inferenza per le relazioni tra variabili: tabelle di contingenza e test di indipendenza, regressione.
- ▶ Inferenza per il modello di regressione lineare

Inferenza statistica

Contenuti:

- ▶ Popolazione e campione: variabilità campionaria
- ▶ Il concetto di inferenza statistica
- ▶ Stima intervallare per media e proporzione
- ▶ Verifica d'ipotesi per media e proporzione
- ▶ Inferenza per le relazioni tra variabili: tabelle di contingenza e test di indipendenza, regressione.
- ▶ Inferenza per il modello di regressione lineare

Obiettivi:

- ▶ Saper interpretare un'analisi statistica, valutandone la generalizzabilità, l'affidabilità, le implicazioni pratiche

Inferenza statistica

Contenuti:

- ▶ Popolazione e campione: variabilità campionaria
- ▶ Il concetto di inferenza statistica
- ▶ Stima intervallare per media e proporzione
- ▶ Verifica d'ipotesi per media e proporzione
- ▶ Inferenza per le relazioni tra variabili: tabelle di contingenza e test di indipendenza, regressione.
- ▶ Inferenza per il modello di regressione lineare

Obiettivi:

- ▶ Saper interpretare un'analisi statistica, valutandone la generalizzabilità, l'affidabilità, le implicazioni pratiche



Difficoltà: alta

- ▶ contenuti matematici sofisticati: analisi matematica
- ▶ concetti non intuitivi
- ▶ si basa sui concetti, ancora nuovi e poco assimilati, di probabilità

Inferenza statistica

Contenuti:

- ▶ Popolazione e campione: variabilità campionaria
- ▶ Il concetto di inferenza statistica
- ▶ Stima intervallare per media e proporzione
- ▶ Verifica d'ipotesi per media e proporzione
- ▶ Inferenza per le relazioni tra variabili: tabelle di contingenza e test di indipendenza, regressione.
- ▶ Inferenza per il modello di regressione lineare

Obiettivi:

- ▶ Saper interpretare un'analisi statistica, valutandone la generalizzabilità, l'affidabilità, le implicazioni pratiche



Difficoltà: alta

- ▶ contenuti matematici sofisticati: analisi matematica
- ▶ concetti non intuitivi
- ▶ si basa sui concetti, ancora nuovi e poco assimilati, di probabilità

Indice

Dettagli sul corso

Programma del corso

La statistica

Cos'è la statistica?

La statistica è la scienza che si occupa di raccogliere dati e trarre da essi informazioni/conoscenza.

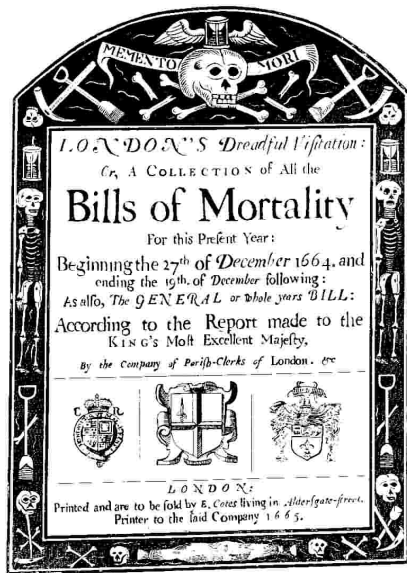


I dati possono aiutare a capire i fenomeni, occorre però raccogliere i dati che servono, e farlo bene; i dati vanno poi esaminati in modo da isolare e evidenziare le informazioni che si cercano.



La statistica si occupa, a partire da una domanda su un fenomeno, di stabilire quali dati possano essere usati per rispondere a quella domanda, e, se i dati non sono già disponibili, come debbano essere raccolti. Segue poi la fase in cui si analizzano i dati per estrarre le informazioni cercate.

I primi dati



- ▶ I dati sulla popolazione e la mortalità sono forse tra i primi a essere raccolti.
- ▶ È con esse che i primi "statistici" (in realtà sono personaggi che hanno varie occupazioni, alcuni, ma non tutti, sono scienziati) si cimentano.
- ▶ A Londra per esempio nel 1600 si compilano e pubblicano settimanalmente i *Bills of mortality*, che riportano il numero di decessi per varie cause.

I primi dati

The Diseases and Casualties this Week,



Bortive	5	Imposthume	11
Aged	43	Infants	16
Ague	2	Killed by a fall from the Bell-frey at Alhallows the Great	1
Apoplexie	1	Kingsevil	2
Bleeding	2	Lethargy	1
Burnt in his Bed by a Candle at St. Giles Cripplegate	1	Palie	1
Canker	1	Plague	7165
Childbed	42	Rickets	17
Chriofomes	18	Rifing of the Lights	11
Consumption	134	Scowring	5
Convulsion	64	Scurvy	2
Cough	2	Spleen	1
Drople	33	Spotted Fever	101
Fever	309	Stillborn	17
Flox and Small-pox	5	Stone	2
Frighted	3	Stopping of the Stomach	9
Gowr	1	Strangury	1
Grief	3	Suddenly	1
Gripping in the Guts	51	Surfeit	49
Jaundies	5	Teeth	125
		Thruith	5
		Timpany	1
		Tiffick	11
		Vomiting	3
		Winde	3
		Wormes	15

Christened	Males — 95	Buried	Males — 4095	Plague — 7165
	Females — 81		Females — 4202	
	In all — 176		In all — 8297	

Increased in the Burials this Week 607

Parishes clear of the Plague	4	Parishes Infected	126
------------------------------	---	-------------------	-----

*The Assize of Bread set forth by Order of the Lord Mayor and Court of Aldermen,
A penny Wheaten Loaf to contain Nine Ounces and a half, and three
half-enny White Loaves the like weight.*

- I dati sulla popolazione e la mortalità sono forse tra i primi a essere raccolti.
- È con esse che i primi “statistici” (in realtà sono personaggi che hanno varie occupazioni, alcuni, ma non tutti, sono scienziati) si cimentano.
- A Londra per esempio nel 1600 si compilano e pubblicano settimanalmente i *Bills of mortality*, che riportano il numero di decessi per varie cause.

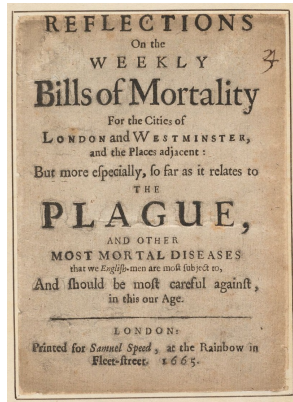
I primi dati

<i>The Diseases and Casualties this Week,</i>	
	Imposthume 11 Infants 16 Killed by a fall from the Bell-frey at Alhallows the Great 2 Kingfevil 1 Lethargy 1 Palfie 1 Plague 7165 Rickets 17 Rising of the Lights 11 Scouring 5 Scurvy 2 Spleen 1 Spotted Fever 108 Stillborn 17 Stone 2 Stopping of the Stomach 9 Strangury 1 Suddenly 1 Surfeit 49 Teeth 124 Thrush 1 Timpany 1 Tiffick 11 Vomiting 3 Winde 3 Wormes 15
A Bortive 5 Aged 43 Ague 2 Apoplexie 1 Bleeding 2 Burnt in his Bed by a Candle at St. Giles Cripplegate 1 Canker 1 Childbed 42 Chirfomes 18 Consumption 134 Convulsion 64 Cough 2 Droptie 33 Fever 309 Flux and Small-pox 5 Frighted 1 Gown 1 Grief 3 Gripping in the Guts 51 Jaundies 5	{ Males 95 { Females 81 } { In all 176 } Increased in the Burials this Week 607 Parishes clear of the Plague 4 Parishes Infected 126
The Assize of Bread set forth by Order of the Lord Mayor and Court of Aldermen, A penny Wheaten Loaf to contain Nine Ounces and a half, and three half-tenny White Loaves the like weight.	

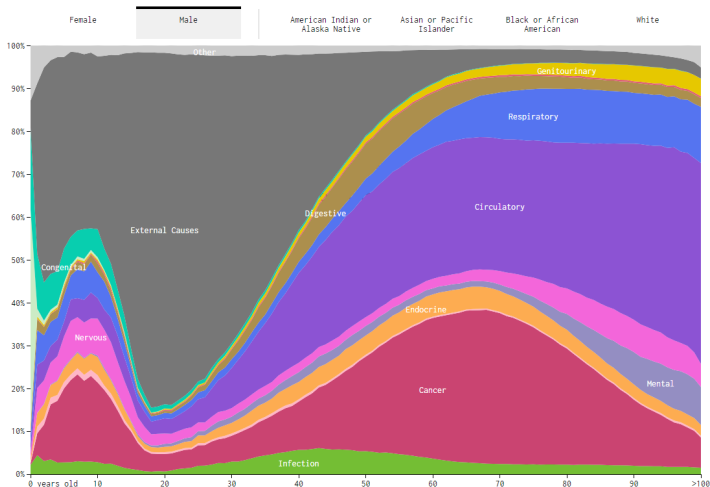
<i>A general Bill for this present year,</i> ending the 19 of December 1665, according to the Report made to the KINGS most Excellent Majesty. By the Company of Parish Clerks of London, &c.	
	
<i>The Diseases and Casualties this year.</i>	
A Bortive and Stillborn 619 Aged 1545 Ague and Fever 557 Appoplex and Suddenly 116 Bedrid 10 Blasted 16 Bleeding 16 Bloody Flux, Scouring & Flux 185 Burnt and Scalded 8 Calenture 3 Cancer, Gangrene and Fiftula 56 Canker, and Thrush 111 Childbed 625 Chirfomes and Infants 1258 Cold and Cough 68 Collick and Winde 134 Consumption and Tiffick 4808 Convulsion and Mother 2030 Diltracted 5 Droptie and Timpany 1478 Drowned 50	Executed 21 Flux and Small Pox 655 Found dead in Streets, fields, &c. 20 French Pox 86 Frighted 23 Gout and Sciatica 23 Grief 46 Gripping in the Guts 1288 Hang'd & made away themselves 7 Headmouldit & Mouldfallen 14 Jaundies 110 Imposthume 227 Kid by severall accidents 46 Kings Evil 26 Lethargy 14 Livergown 20 Measles 23 Murthered and Shot 5 Overlad & Starved 45
Christened Females 48569 Buried Females 48737 In all 9967	Of the Plague 68596 Increased in the Burials in the 130 Parishes and at the Pest-house this year 79000 Increased of the Plague in the 130 Parishes and at the Pest-house this year 68596

Previsione di epidemie: John Graunt e la peste

- ▶ Tra i primi troviamo **John Graunt** (1620-1674), di professione negoziante a Londra.
- ▶ Studia i *bills of mortality* con l'obiettivo, piuttosto moderno, di rilevare l'inizio di epidemie di peste in città.
- ▶ Sarebbe (stato) utile, ad esempio per decidere la quarantena.
- ▶ Sebbene non riesca nell'intento, il suo lavoro è comunque rilevante come studio della popolazione.

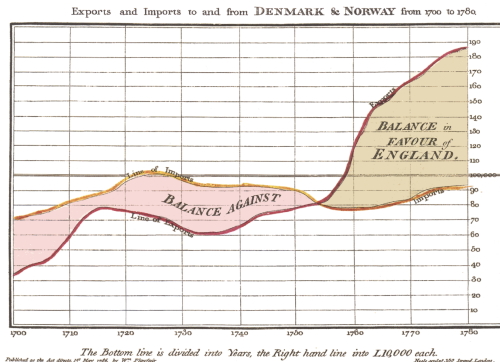
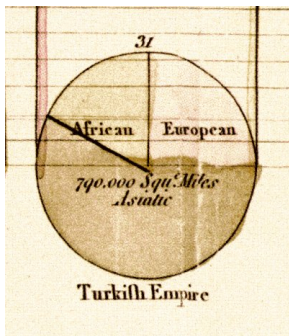


Una rappresentazione più moderna della mortalità per causa



Playfair

L'altro argomento "caldo" dopo la morte sono i soldi, e in effetti c'è una lunga tradizione anche nella raccolta di dati economici e sociali.



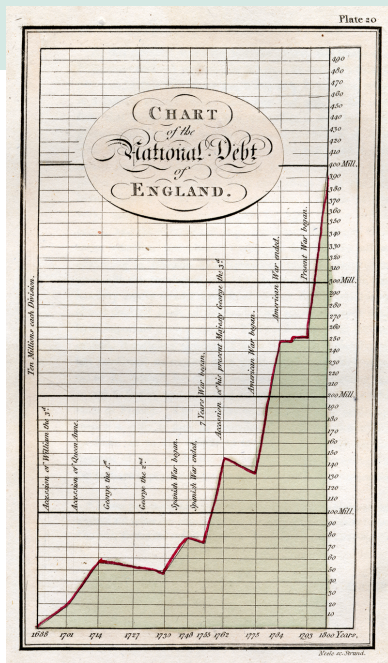
Questi esempi di rappresentazione di dati economici si devono a **William Playfair** (1759-1823), ingegnere e economista politico scozzese.

Playfair

Playfair è tra i primi a impiegare metodi grafici per la rappresentazione di dati, in particolare relativi al commercio o più in generale all'economia politica.



A destra, la serie storica del debito pubblico inglese.



Altro, da Google

Google Public Data
www.google.com/publicdata/directory

Google
Search public data
francesco.pauli@gmail.c...

Public Data
Language

Datasets
Metrics

Any data provider (136)
Eurostat (10)
Destatis (7)
Statistics Iceland (6)
U.S. Bureau of Labor Statistics (5)
Central Statistics Office, Ireland (5)

My Datasets

Fertility rate ? **Countries** ?

Living longer with fewer children

This chart correlates life expectancy and number of children per woman for each country in the world. The bubbles are sized by population and colored by region. Over time, most countries have moved towards the bottom right corner of the chart, corresponding to long lives and low fertility. Note the progression of the bubble for China- in the late 60's and 70's life expectancy rose quickly, then the implementation of the one-child policy caused a drop in the number of children per woman.

Explore the data

Dataset: World Development Indicators
Source: World Bank

World Development Indicators
World Bank
This dataset contains the World Development Indicators (WDI).

Statistica ed economia: applicazioni

Dal macro

- ▶ stima degli aggregati macroeconomici (PIL, inflazione, disoccupazione)
- ▶ previsioni macroeconomiche
- ▶ verifica empirica di relazioni economiche

al micro

- ▶ previsioni di vendita
- ▶ profilazione della clientela (pubblicità, offerte mirate)
- ▶ credito
 - ▶ banche che decidono a chi prestare denaro
 - ▶ valutazione dei crediti a bilancio
 - ▶ *rating* di istituzioni pubbliche e private
- ▶ finanza
 - ▶ valutazione dei rischi di attività finanziarie
 - ▶ ottimizzazione dei portafogli (strategie di investimento mobiliare)
- ▶ assicurazioni

Tassi di disoccupazione e occupazione

5. OCCUPATI E TASSO DI DISOCCUPAZIONE

(migliaia di unità; valori %)



Fonte: Istat

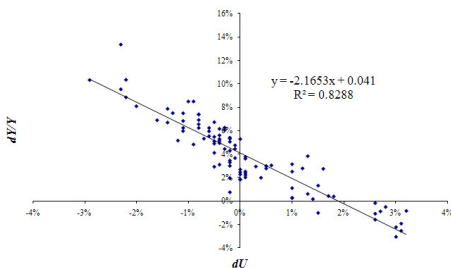
Il tasso di disoccupazione viene stimato dall'ISTAT intervistando un campione di individui.

Verifica empirica di relazioni economiche

Il tasso di crescita del prodotto (PIL) è inversamente proporzionale al tasso di crescita della disoccupazione.

$$\Delta(\text{PIL}) = \alpha + \beta \Delta(\text{Disocc})$$

Estimation of Potential GDP Growth, 1948-1972



La relazione teorica può essere confrontata con le osservazioni empiriche (variazioni dei tassi effettivamente verificatisi) per

- ▶ confermare/smentire la teoria
- ▶ dare un valore a α e β per fare previsioni

Assicurazioni

Premio

Indennizzo

Variabile aleatoria

500€



{
0€
1 000€
2 000€
:
:
10 000€
:
:
100 000€
:
:
1 000 000€
}

Assicurazioni

Premio

Indennizzo

Variabile aleatoria

500€



0€	molto probabile
1 000€	possibile
2 000€	possibile
⋮	⋮
10 000€	possibile
⋮	⋮
100 000€	improbabile
⋮	⋮
1 000 000€	molto improbabile

Assicurazioni

Premio

Indennizzo

Variabile aleatoria

500€



0€	molto probabile	→ p_0
1 000€	possibile	→ p_1
2 000€	possibile	→ p_2
⋮	⋮	
10 000€	possibile	→ p_j
⋮	⋮	
100 000€	improbabile	→ p_k
⋮	⋮	
1 000 000€	molto improbabile	→ p_m

probabilità

stimate

Assicurazioni

Premio

Indennizzo

Variabile aleatoria

500€

← *MEDIA* →
(*speranza matematica*)

0€	molto probabile	→ p_0
1 000€	possibile	→ p_1
2 000€	possibile	→ p_2
⋮	⋮	
10 000€	possibile	→ p_j
⋮	⋮	
100 000€	improbabile	→ p_k
⋮	⋮	
1 000 000€	molto improbabile	→ p_m

probabilità

stimate

Assicurazioni

Premio

Indennizzo

Variabile aleatoria

500€

← *MEDIA* →
(*speranza matematica*)

0€	molto probabile	→ p_0
1 000€	possibile	→ p_1
2 000€	possibile	→ p_2
⋮	⋮	
10 000€	possibile	→ p_j
⋮	⋮	
100 000€	improbabile	→ p_k
⋮	⋮	
1 000 000€	molto improbabile	→ p_m

probabilità

stimate

Tanti premi

Tanti indennizzi

Assicurazioni

Premio

Indennizzo

Variabile aleatoria

500€

← *MEDIA* →
(*speranza matematica*)

0€	molto probabile	→ p_0
1 000€	possibile	→ p_1
2 000€	possibile	→ p_2
⋮	⋮	
10 000€	possibile	→ p_j
⋮	⋮	
100 000€	improbabile	→ p_k
⋮	⋮	
1 000 000€	molto improbabile	→ p_m

probabilità

stimate

Tanti premi

$n \times 500\text{€}$

$\approx$
legge dei grandi numeri

Tanti indennizzi

$\sum_{i=1}^n \text{Indennizzo}_i$

Assicurazioni

In sintesi, in un contratto di assicurazioni si scambiano un importo certo, il premio, con un importo aleatorio, l'indennizzo.



Se il premio è pari alla speranza matematica dell'indennizzo l'affare è equilibrato e, in virtù della legge dei grandi numeri nel lungo termine nessuno perde e nessuno guadagna.



Per determinare la speranza matematica dovremo stimare le probabilità dei possibili indennizzi sulla base dei dati passati.



(In realtà, il premio è maggiore della speranza matematica per garantire la profittabilità dell'impresa assicurativa, di quanto maggiore è legato alla varianza dell'indennizzo.)

Assicurazioni: rendite vitalizie (pensioni)

Premio Pagamento
 x $\longleftrightarrow$ 1000€ al mese a vita

Legare il premio al pagamento probabile non era scontato: nel 1600 il governo britannico offriva questo tipo di rendite e il pagamento x era indipendente dall'età dell'assicurato.

Fu **Edmond Halley** (1656-1742), (l'astronomo, quello della cometa) che, fondandosi sul lavoro precedente di Graunt, analizza le tavole di mortalità di Breslau, consentendo al governo britannico di calcolare il prezzo corretto per le rendite vitalizie sulla base dell'età dell'individuo.



Cos'è la statistica? (ripreso)

Schematicamente le fasi di un'analisi statistica sono

- ▶ *formulare una domanda*, tradurre un'esigenza conoscitiva in modo che sia suscettibile di una risposta in termini statistici;
- ▶ *individuare o raccogliere i dati*, è un ambito vasto, che va sotto il nome di disegno sperimentale e campionamento; i fondamenti poggiano sul calcolo delle probabilità;
- ▶ *organizzare e guardare i dati*, dalla massa di dati così com'è non è immediato estrarre le informazioni che servono, si possono però sintetizzare opportunamente e/o rappresentarli graficamente, in funzione delle informazioni cercate;
- ▶ *modellare* si formula (specifica) un modello probabilistico che possa spiegare i dati osservati in base alle ipotesi fatte sul fenomeno, si stima il modello usando i dati. Il modello potrà essere usato per confermare o smentire delle ipotesi fatte sul fenomeno e/o per previsioni su future istanze.