

CDL in MEDICINA & CHIRURGIA

Statistica Medica

gbarbati@units.it

A.A. 2025-26



UNITÀ DI BIOSTATISTICA

Dipartimento Universitario Clinico di
Scienze Mediche Chirurgiche e della Salute



CALENDARIO DELLE LEZIONI:



23/09: Intro + Statistica descrittiva I

30/09: Statistica descrittiva II

07/10: Introduzione alla statistica inferenziale

14/10: Test diagnostici & Curva ROC

21/10: Lezione di statistica inferenziale I

28/10: Lezione di statistica inferenziale II

04/11: Correlazione & Regressione

11/11: Dimensione Campionaria

<https://moodle2.units.it/>

Dipartimento Universitario Clinico di Scienze mediche, chirurgi... / Laurea Magistrale Ciclo Unico 6 anni / ME03 - MEDICINA E CHIRURGIA / A.A. 2025 - 2026

894ME-2 - STATISTICA MEDICA 2025

[Corso](#)

[Impostazioni](#)

[Partecipanti](#)

[Valutazioni](#)

[Report](#)

[Altro ▾](#)

Scaricare il materiale didattico da Moodle

Materiale didattico:

Lezioni (.pdf delle slides)

Quiz di esempio

*esempi utilizzando
Jamovi, basato su R*
<https://www.jamovi.org/>

The image shows a screenshot of the Jamovi website and a video playlist. The website is titled "jamovi" and features a navigation bar with links to "docs", "blog", "forum", and "developer's hub". The main content area displays a large image of the Jamovi software interface, which includes a data table and an ANOVA table. Below the image, there are two buttons: "jamovi Cloud" and "jamovi Desktop". The video playlist on the left side of the image contains seven videos, each with a thumbnail and a title. The titles are: "jamovi 1 intro", "jamovi 2 calcola e trasforma", "Jamovi 3-stat descr variabili numeriche", "Jamovi 4 descr variabili categoriche", "jamovi 5 filtro e trasformazione", "jamovi 6 z score e correlazione", and "jamovi 7 variabili aleatorie".

Generale Post File JAMOV! playlist +

+ Aggiungi nuovo elemento Condividi Modifica Elimina Apri in SharePoint ...

Elenchi personali
JAMOV! ☆

- jamovi 1 intro 08:25
- jamovi 2 calcola e trasforma 06:36
- Jamovi 3-stat descr variabili numeriche 09:20
- Jamovi 4 descr variabili categoriche 06:40
- jamovi 5 filtro e trasformazione 08:31
- jamovi 6 z score e correlazione 07:06
- jamovi 7 variabili aleatorie

jamovi - open statistical software

docs blog forum developer's hub

features products about resources contribute

open statistical software for the desktop and cloud

jamovi Cloud Use jamovi in the cloud directly from your browser

jamovi Desktop Download and install Jamovi onto your computer

Help me choose

SUPPORT US

STATS MADE SIMPLE

jamovi is a new "3rd generation" statistical spreadsheet, designed from the ground up to be easy to use, jamovi is a compelling alternative to costly statistical products such as SPSS and SAS.

Modalità di esame

La modalità di esame [quiz] sarà su **Moodle**, attraverso l'utilizzo di un **quiz online**.

L'esame si svolgerà in **aula informatica (SPAZIO ROSA SAN GIOVANNI)**.

Vista la **ridotta capienza** delle aula informatiche, all'occorrenza si svolgeranno su **più turni**.

- Il quiz è composto da **16** domande.
- Ogni risposta corretta vale **2** punti, per ottenere la sufficienza (18/30) occorre rispondere correttamente a 9 domande.
- Per ottenere 30/30 occorre rispondere correttamente a 15 domande.
- Per ottenere **30 e lode (31/30)** occorre rispondere correttamente a tutte e 16 le domande.

Esempi Quiz

Quiz di esame

Iscrizione all'esame

Sede di San Giovanni
Spazio Rosa
Laboratorio Informatico

Sono già inserite le **date** per le **prove parziali di statistica** a cui vi iscrivete in ESSE 3.

Gli esiti del quiz di statistica saranno disponibili in Moodle appena conclusa la prova.

Superata la prova di statistica e quella di informatica, **SE E SOLO SE** accettate il voto finale (**media aritmetica dei due voti, si arrotonda all'intero superiore**) vi iscrivete all' appello per la registrazione.

INDICATE AL MOMENTO DELLA ISCRIZIONE ALL'APPELLO DI REGISTRAZIONE LE DATE IN CUI AVETE SOSTENUTO LE PROVE PARZIALI CON I RELATIVI ESITI

Se **non si supera** una delle due prove (o **si vuole ripetere** la prova) si ha tempo **fino alla sessione straordinaria dell'anno successivo** (**febbraio 2027**) (mantenendo eventualmente valido l'esito dell'altra prova). Altrimenti bisogna ripetere entrambe. Se si decide di **ripetere** la prova parziale, automaticamente si rifiuta il voto della precedente.

DATE APPELLI PROVE PARZIALI DI STATISTICA:

Descrizione Appello		Data ora aula
Prova parziale di Statistica		25/02/2026 09:00 Laboratorio Informatico
Prova parziale di Statistica		11/02/2026 09:00 Laboratorio Informatico
Prova parziale di Statistica		29/01/2026 09:00 Laboratorio Informatico
Prova parziale di Statistica		14/01/2026 09:00 Laboratorio Informatico

L'iscrizione in ESSE 3 alle prove parziali si attivano **dal giorno seguente** alla prova che precede. Per la prima prova del 14 gennaio ci si può iscrivere da metà novembre.

ORGANIZZATEVI **CON CONGRUO ANTICIPO** PER
SCADENZE DI REGISTRAZIONE VOTO LEGATE A BORSE DI
STUDIO O SIMILI !!!!

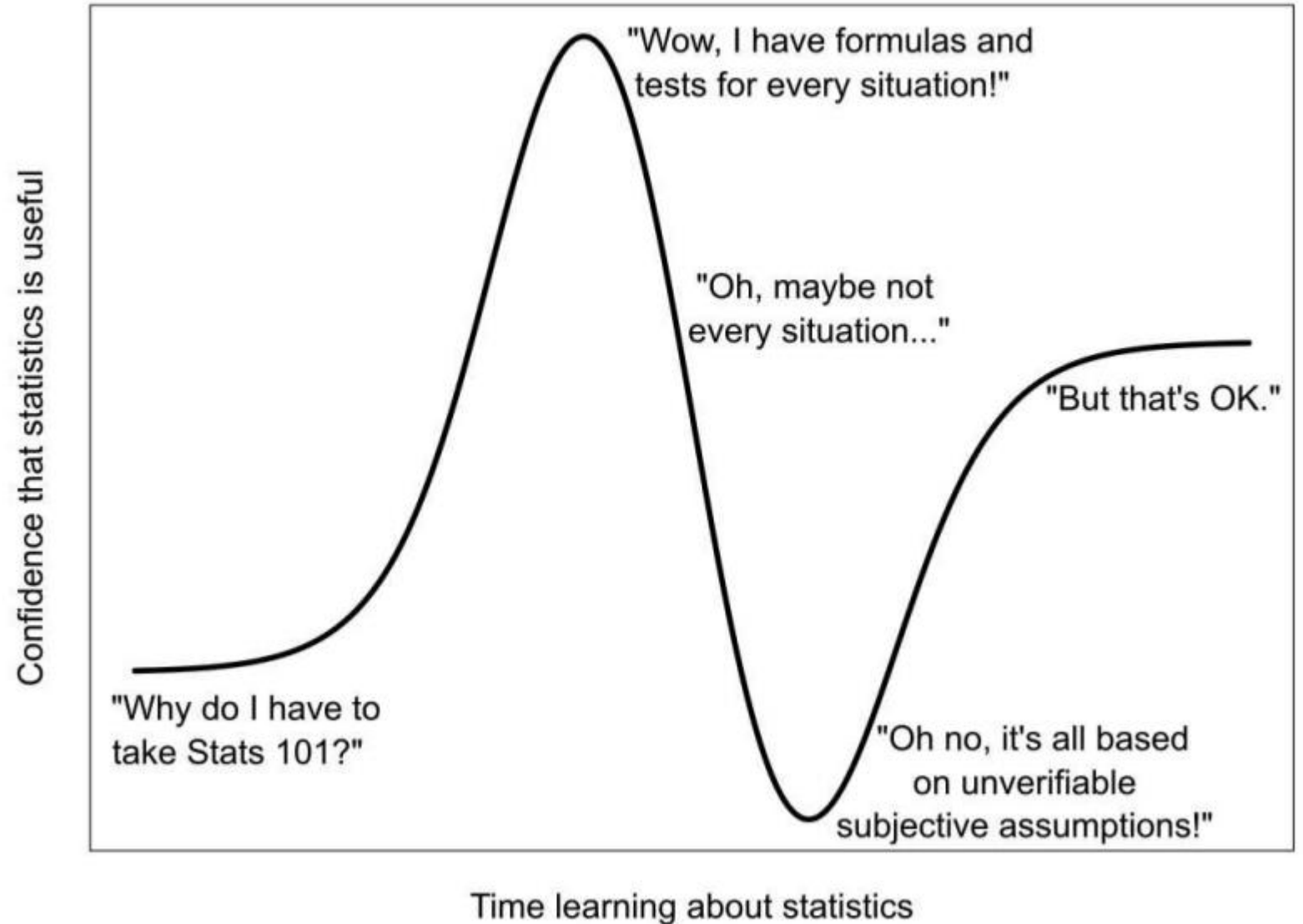
Ricevimento studenti: **al termine** della lezione o tramite **videochiamata** in MS teams o in Università Centrale, Edificio D, studio 4.2 [quarto piano]

Previo avviso tramite e-mail: gbarbati@units.it



ARGOMENTI:

- Introduzione alla Statistica
- Concetti generali e terminologia
- Studi sperimentali e osservazionali



Che cos'è la statistica ?

E' la scienza che utilizza i **dati** per rispondere a specifiche **domande** di interesse. Include:

- metodi per **raccogliere dati** pertinenti alla domanda (disegno dello studio)
- metodi per **riassumere e visualizzare** i dati per far luce sulla domanda (statistica descrittiva)
- metodi che ci consentano di **trarre risposte** alla domanda supportate dai dati (statistica inferenziale)

I dati contengono (quasi) sempre incertezza.

Questa incertezza può derivare dalla **selezione** degli elementi da misurare o dalla **variabilità** del processo di misurazione.

Trarre conclusioni generali dai dati è la base per aumentare la conoscenza del mondo ed è la base di tutta l'indagine scientifica razionale.

L'inferenza statistica ci fornisce **metodi** per farlo **nonostante** l'incertezza nei dati. I metodi utilizzati per l'analisi dipendono *dal modo* in cui i dati sono stati raccolti. Vengono utilizzati gli strumenti della **probabilità** per tenere in conto tali fonti di incertezza.

Che cos'è la statistica [descrittiva]?

La statistica descrittiva ha come scopo generale la conoscenza **quantitativa** dei **fenomeni collettivi**.

Es: Quante persone ci sono in questa classe?

L'operazione che si deve fare è forse la più importante, ma certo la più semplice, operazione della statistica:
il **conteggio**.

In base ad esso si perviene alla conoscenza del numero delle persone che costituiscono il gruppo.

I fenomeni che la nostra mente non può conoscere con una sola osservazione, ma che apprende tramite la sintesi delle osservazioni di fenomeni più semplici, si possono definire

fenomeni di massa o fenomeni collettivi

Essi sono composti da una collettività di osservazioni di fenomeni più semplici

fenomeni individuali o singoli

Tramite la statistica è possibile sostituire ad un'impressione *qualitativa* di un fenomeno la sua misura *quantitativa*.

Che cosa si intende in statistica per 'fenomeni collettivi'?

- Collettività di casi singoli:

Caratteristiche di una popolazione osservando gli individui che compongono.
(Es: nascite, matrimoni, morti...)

- Ripetizioni dell'osservazione di un unico fenomeno singolo:

La lunghezza di un segmento, la massa di un corpo, la temperatura di un materiale ecc...: un'unica misurazione non può fornire l'esatta misura del fenomeno, poiché eseguendo misure della stessa grandezza si ottengono generalmente misure diverse (anche se di poco). Se le differenze sono effetti di cause che sfuggono ad ogni controllo, si può avere una misura verosimilmente esatta **ripetendo** l'osservazione un certo numero di volte e sintetizzando in un unico numero i risultati ottenuti. Alla misura si perviene quindi mediante una **collettività di misurazioni**.

Chi usa la statistica?

- Le scienze che studiano un aspetto di una collettività di individui o enti:
per es: demografia, economia, sociologia, *epidemiologia* (cioè quel settore della medicina che studia il tipo, la frequenza, la modalità di distribuzione e il ruolo di fattori causali di un determinato evento nelle popolazioni)...ecc...
- Le scienze cosiddette 'sperimentali' quelle cioè che devono fornire ed utilizzare dati misurati con molta precisione: la fisica, la chimica, l'ingegneria, la biologia, la ricerca biomedica...ecc...

Qual è il ruolo della **statistica** nella metodologia della ricerca?

Lo studio dei fenomeni collettivi avviene attraverso diverse fasi:

FASE 1: *Schematizzazione*, consiste nella **definizione del fenomeno**, nell'individuazione della **collettività** in cui esso si realizza e nella scelta delle **caratteristiche** della collettività che interessano.

Per esempio: se si intende studiare la *propensione al consumo di alcolici* (= il fenomeno) nella *popolazione europea* per la fascia di età compresa tra i 14 e i 30 anni (= collettività) si potrebbe decidere di rilevare *la spesa media per alcolici in una settimana* (= caratteristica di interesse).

(A) TUTTA LA POPOLAZIONE VIENE STUDIATA:

Se la collettività di individui alla quale è legato il fenomeno in esame è *interamente da osservare* ed allora alla fase di schematizzazione segue la:

FASE 2A: **Osservazione**: raccolta, nell'ordinamento e nella classificazione del materiale di osservazione.

Al termine di queste due fasi, alla collettività di individui si sostituisce un **insieme di dati** (=es. dati sulla spesa media settimanale per alcolici per individuo) su cui si può operare con procedimenti per lo più propri della matematica per la:

FASE 3A: **Descrizione**.

FASE 1: Schematizzazione, consiste nella **definizione del fenomeno**, nell'individuazione della **collettività** in cui esso si realizza e nella scelta delle **caratteristiche** della collettività che interessano.

(B) VIENE STUDIATA UNA PARTE DELLA POPOLAZIONE:

Se della collettività di individui alla quale è legato il fenomeno conviene osservare solo una parte (un *campione*), allora alla fase di *schematizzazione* segue quella di:

FASE 2B: Formulazione di ipotesi sulla collettività più vasta, o di enunciazione di una teoria o di ricorso ad un modello;

FASE 3B: Osservazione del campione, che è una parte della collettività, e a cui poi si sostituisce un insieme di dati, che viene poi

FASE 4B: Descrizione con procedimenti matematici dei dati rilevati sul campione.

A tutte queste fasi segue quella di:

FASE 5B: Induzione o di **Inferenza** che fa risalire dalla descrizione del campione a quella dell'insieme più ampio e che è anche una verifica delle ipotesi formulate nella fase 2b.

(A): TUTTA LA POPOLAZIONE:

Nel primo caso la metodologia impiegata è puramente descrittiva, da cui deriva il termine di “***statistica descrittiva***” dato all’insieme di tecniche relative a questo tipo di indagini. Occorre comunque notare che **anche nella statistica descrittiva si può fare riferimento a dei campioni di popolazione**; l’intera collettività viene generalmente osservata solo in occasioni di indagini particolari, come per esempio il ***censimento***.

(Statistica descrittiva).

(B) UN CAMPIONE DELLA POPOLAZIONE:

Nel secondo caso, la metodologia impiegata può essere detta inferenziale ed il relativo campo della statistica viene definito “***statistica inferenziale***”. Ovviamente nella statistica inferenziale si devono compiere le operazioni di schematizzazione e di descrizione e quindi necessitano tutti gli strumenti che sono propri della statistica descrittiva; ad essi poi vanno aggiunti i procedimenti puramente inferenziali.

(Statistica inferenziale).

La statistica ci aiuta nel delineare il quesito scientifico e le diverse fasi del progetto di ricerca:

quesito scientifico -> ipotesi -> **disegno dello studio** -> analisi dei dati

Raccolta e gestione dei dati & analisi statistica sono delle **componenti cruciali** per la maggioranza degli studi clinici.

Quesiti di ricerca:

- Quali gruppi si vogliono **descrivere/confrontare**?
 - Quali parametri si vogliono **misurare** ?
 - **Quando**/in che occasione/i si vogliono misurare ?
 - Quali parametri si ha bisogno di “**controllare**” per confrontare i gruppi ?
 - ...
-

OBIETTIVI DELLA RICERCA

Descrivere associazioni/*patterns*.

- Moderatamente semplice...
- Occorre compilare un *report* delle analisi statistiche fatte.

—————→ Statistica descrittiva

Inferenza: generalizzare dal campione alla popolazione.

- Moderatamente difficile...
- Richiede un piano di analisi *prima* di cominciare.

—————→ Statistica inferenziale

Prevedere cosa accadrà ai soggetti della popolazione di interesse.

- Descrittiva + Inferenza...(usando i modelli adeguati...)
- Piano di analisi pre/post per pianificare/documentare.

└—————→ Modelli di regressione
[faremo alcuni cenni...]

TERMINOLOGIA NEI DISEGNI DI STUDIO:

Outcome : il risultato di interesse per i soggetti (unità statistiche) dello studio.

- ◀ es : tempo ad un evento (morte/recidiva);
- ◀ es : un parametro di funzionalità fisiologica o uno stato di disabilità;

Predittore/i di interesse : la/le variabile/i che si vogliono studiare come possibili **cause** dell'outcome.

- ◀ es : genotipo/i
- ◀ es : trattamento

Potenziale Confondente: variabile che può essere “**confusa**” con un predittore di interesse

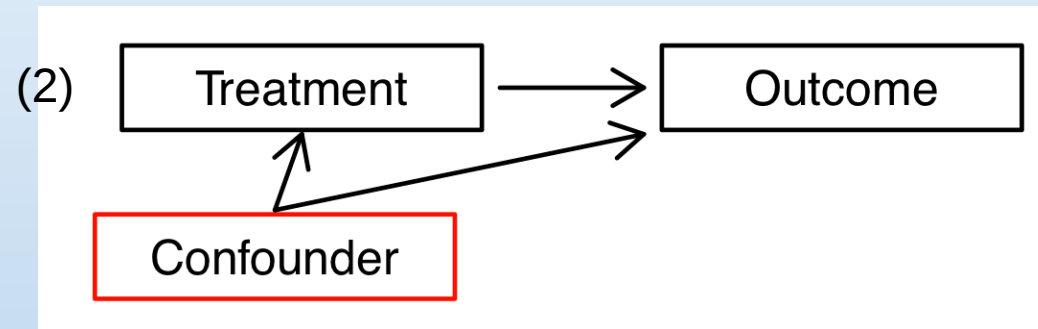
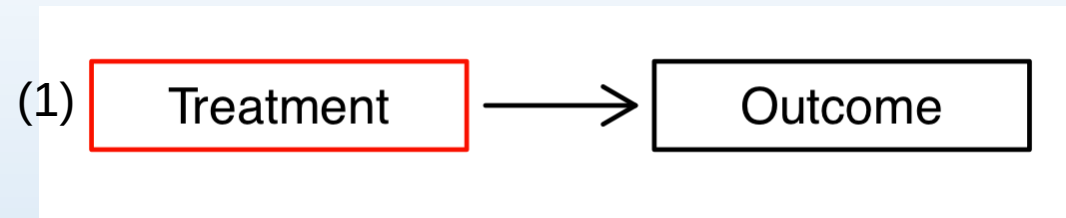
(associata sia all'outcome che al predittore).

- ◀ es: indicazioni cliniche ad un certo trattamento;
- ◀ es: luogo di arruolamento;

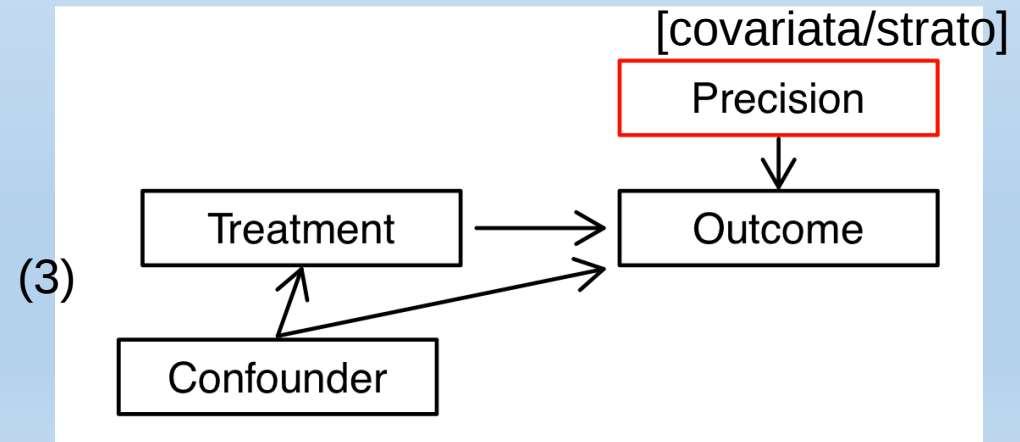
‘Precision’/[covariata/strato]: spiega parte della **variabilità** dell'outcome (non è associata al predittore di interesse).

- ◀ es: baseline (pre-randomization) stato di salute;
- ◀ es: età/sexo...

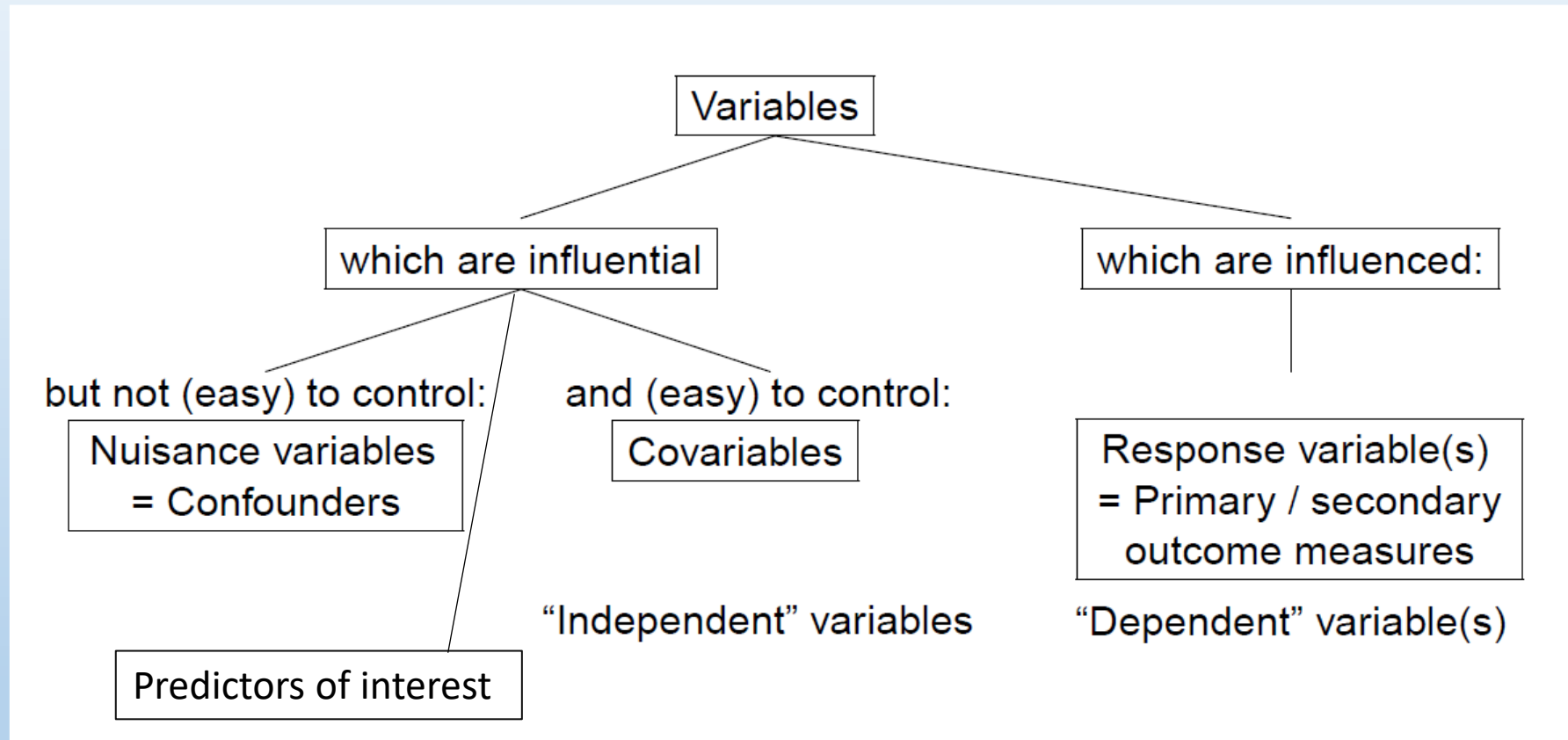




Livello crescente di «complessità»
del disegno dello studio
da (1) a (3)



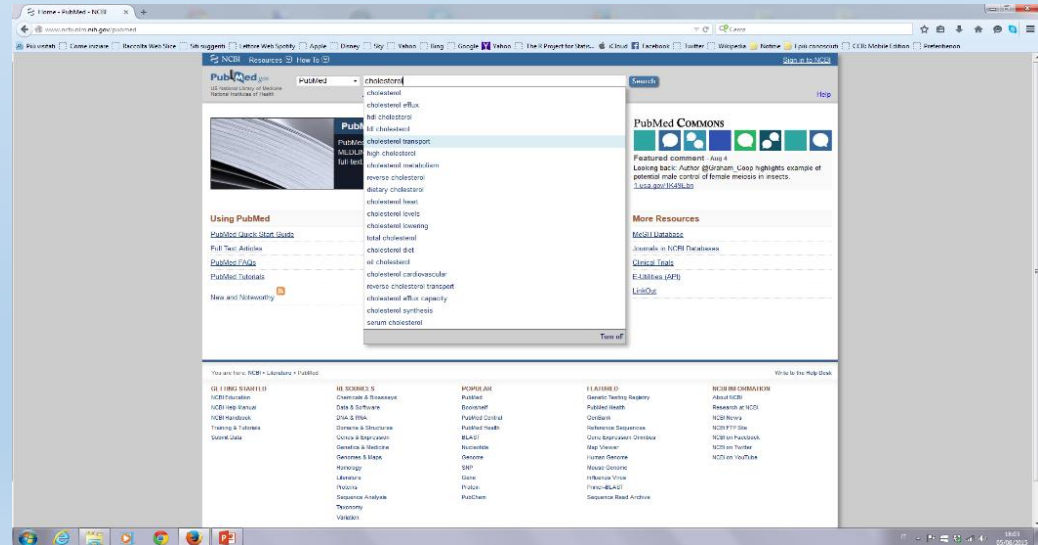
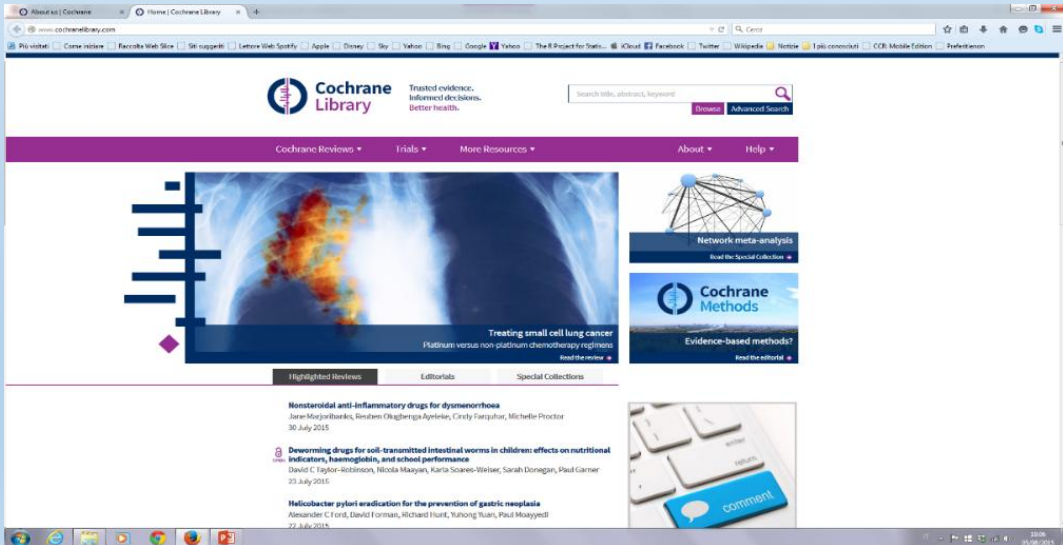
Le 'Variabili' in un Disegno di Studio:



Prima di cominciare uno studio

Revisione **aggiornata** della letteratura sull'argomento:

- **Background**) sullo stato dell'arte della ricerca sul fenomeno, (**bibliografia**) ; discutere **in modo critico** risultati e metodi dei papers riportati.
- Documentare dai risultati di altri studi la efficacia del trattamento/farmaco/intervento oggetto di studio o sul potenziale fattore di rischio, per avere delle **stime quantitative** del suo effetto... (serve per determinare la **dimensione campionaria** opportuna)



Nel pianificare uno studio due aspetti sono particolarmente rilevanti:

-> ***precisione*** dei risultati

-> ***capacità di rispondere alle domande*** che ci si era posti

Questioni più frequenti da risolvere:

(i) come **selezionare** gli individui da osservare;

(ii) come decidere il **numero** di osservazioni da fare su uno o più gruppi di soggetti;

(iii) come **ripartire** le osservazioni tra differenti possibili gruppi

- per es. come suddividere i soggetti che devono ricevere differenti trattamenti-

Il campionamento

Alla base di uno studio statistico è il tipo di *campionamento* che si è condotto; a parte infatti alcuni casi particolari che richiedono **la completa enumerazione** dei soggetti studiati (ex: il *censimento* della popolazione), la maggior parte degli studi richiede la formazione di un *campione*.

Il ***campionamento semplice casuale*** è il modello basilare per la maggior parte delle indagini:

- si definisce una sorta di *lista* che identifichi gli individui della popolazione da cui estrarre il campione;
(ex: in un'indagine ospedaliera i codici che identificano in modo anonimo i pazienti ricoverati)
- si estraggono *un certo numero* di codici *a caso* utilizzando dei 'generatori di numeri casuali' che andranno ad identificare i pazienti selezionati nel campione.

Altri tipi di campionamento:

-campionamento stratificato: la popolazione viene dapprima suddivisa in gruppi (*strati*) (ex: l'Italia suddivisa nelle sue regioni) e si estrae un campione per ogni strato (ex: 100 cittadini per ogni regione);

-campionamento a più stadi: la popolazione viene suddivisa in unità appartenenti ad un primo *stadio di campionamento* (ex: si estraggono *casualmente* 10 scuole in una certa zona) e poi in unità appartenenti ad un *secondo stadio di campionamento* (ex: per ogni scuola estratta nella prima fase si estraggono 2 classi).

Una prima distinzione **tra due tipi** di studi statistici:

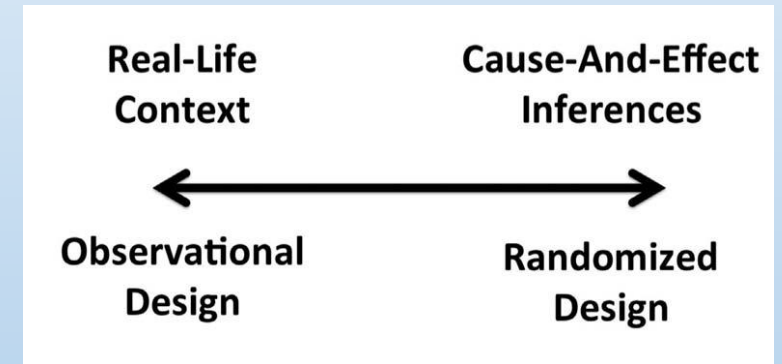
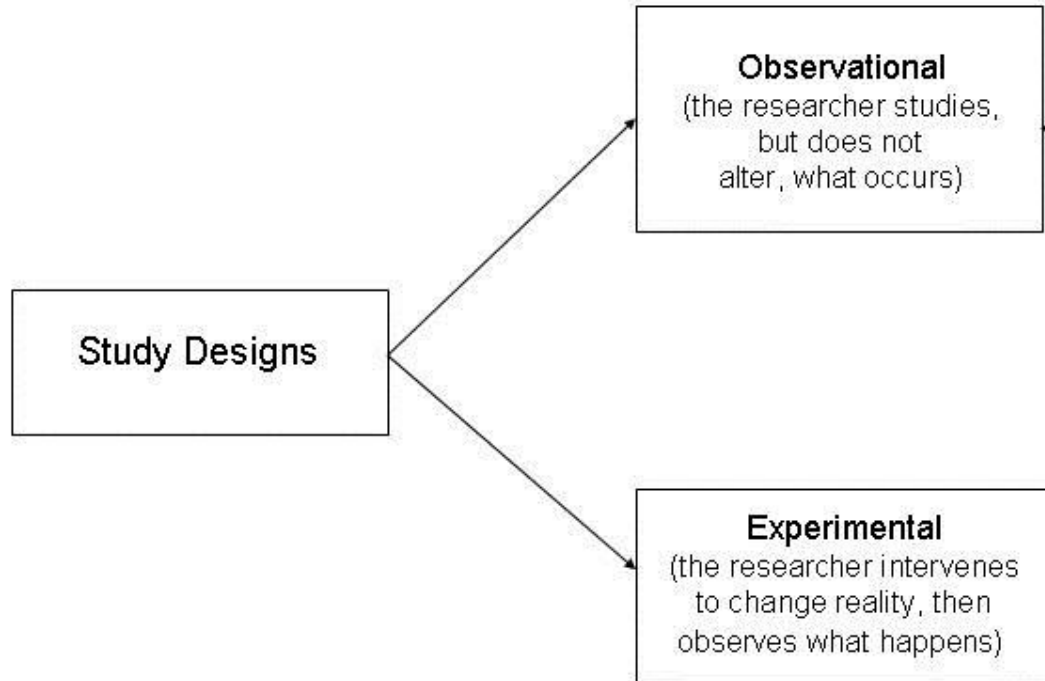
(1) **Esperimenti**

(2) **Indagini** (o studi osservazionali).

Esperimenti:	interferenza pianificata con il corso naturale degli eventi cosicché <u>l'effetto sperimentale</u> possa essere osservato.
Indagine:	il ricercatore è un <u>osservatore</u> più passivo ed interferisce il meno possibile con il fenomeno che deve essere studiato.

- Indagine sui tipi di veicoli a motore che passano per un certo casello stradale durante un certo intervallo di tempo; (*indagine*)
- Sondaggio di opinione; (*indagine*)
- Studio sulla funzione respiratoria (misurata attraverso vari test) dei lavoratori in una certa industria; (*indagine*)
- Osservazione sui tempi di sopravvivenza di topi di diverse razze dopo avere inoculato loro una certa dose di una sostanza tossica; (*esperimento*)
- Indagine clinica per confrontare i vantaggi tra intervento chirurgico o trattamento farmacologico per pazienti in una certa condizione, avendo assegnato in modo *casuale* i soggetti ai due trattamenti. (*esperimento*)

Che «tipo» di studio fare : esperimento o osservazione?



Gli esperimenti

Studi che vengono pianificati in generale con lo scopo di confrontare gli effetti di diversi trattamenti sullo stesso tipo di unità sperimentali.

In particolare in medicina tali esperimenti si definiscono con il termine inglese: '*clinical trials*'. Il trattamento da applicare a ciascuna unità viene deciso dal ricercatore.

- confronto sull'effetto in diversi animali di diverse dosi una sostanza chimica; **unità**=gli animali, il **trattamento**= le diverse dosi della sostanza.
- studio di profilassi per confrontare l'efficacia sui bambini di diversi vaccini contro il morbillo. **unità**=bambino, il **trattamento**=vaccino.
- confronto in un paziente che soffre di attacchi ricorrenti di una malattia cronica di diversi metodi per alleviare il dolore. **unità**=le occasioni successive degli attacchi, il **trattamento**= metodi anti-dolore.
- studio sui vantaggi e svantaggi relativi a diversi programmi di educazione alla salute. **unità**=aree, **trattamento**= programmi di educazione

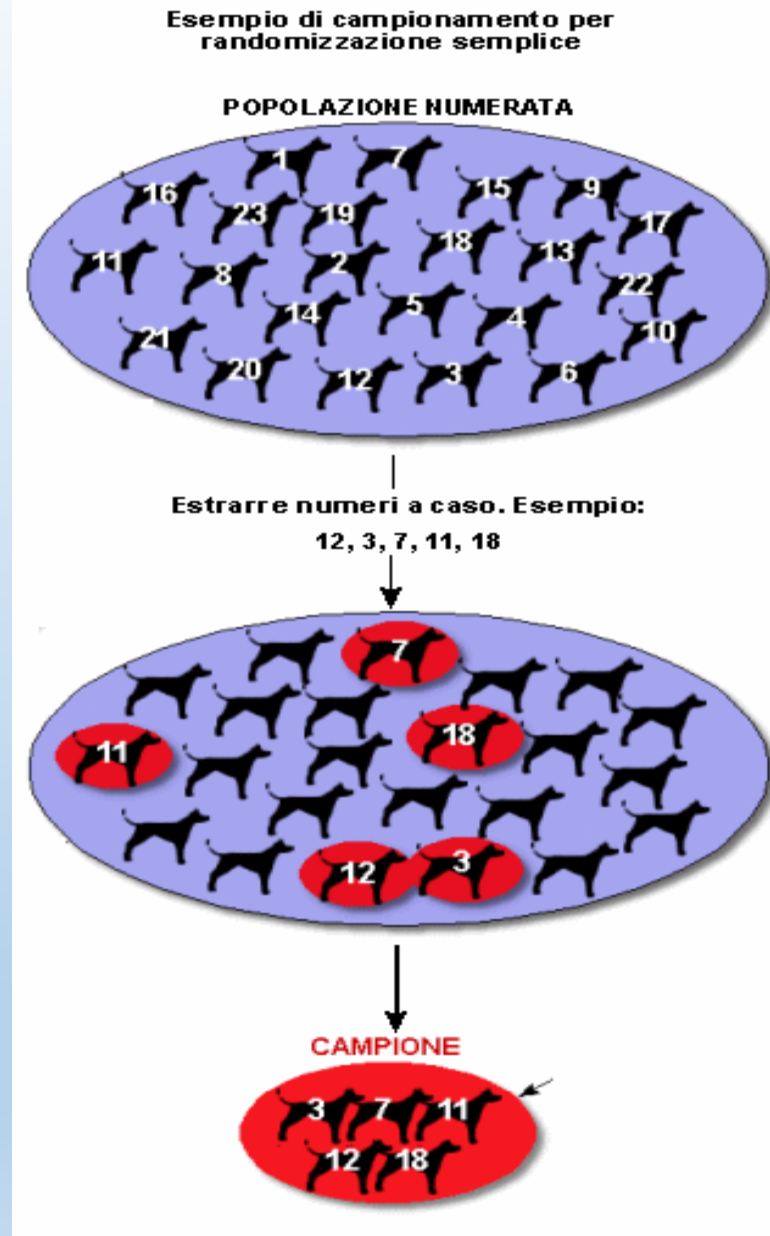
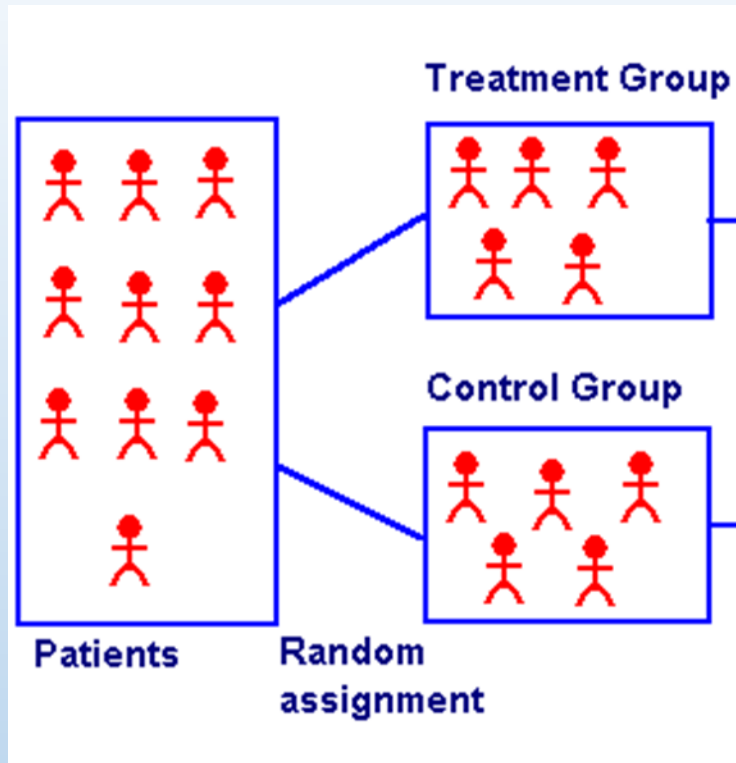
...come **assegnare** i trattamenti alle unità disponibili?

I gruppi di unità sperimentali a cui vengono assegnati i diversi trattamenti devono essere il più *simili* possibile, almeno negli aspetti più rilevanti per la sperimentazione.

Nonostante l'attenzione del ricercatore a che i *fattori* ritenuti importanti siano *bilanciati*, non si può mai essere sicuri che i gruppi non differiscano in modo rilevante per qualche fattore anch'esso importante ma che è stato ignorato durante l'attribuzione ai trattamenti.

Soluzione (semplice) = **randomizzazione**

= fissato il numero n di unità da assegnare ad ogni gruppo, si estraggono dei **campioni casuali** di n unità dal gruppo complessivo di N unità sperimentali.



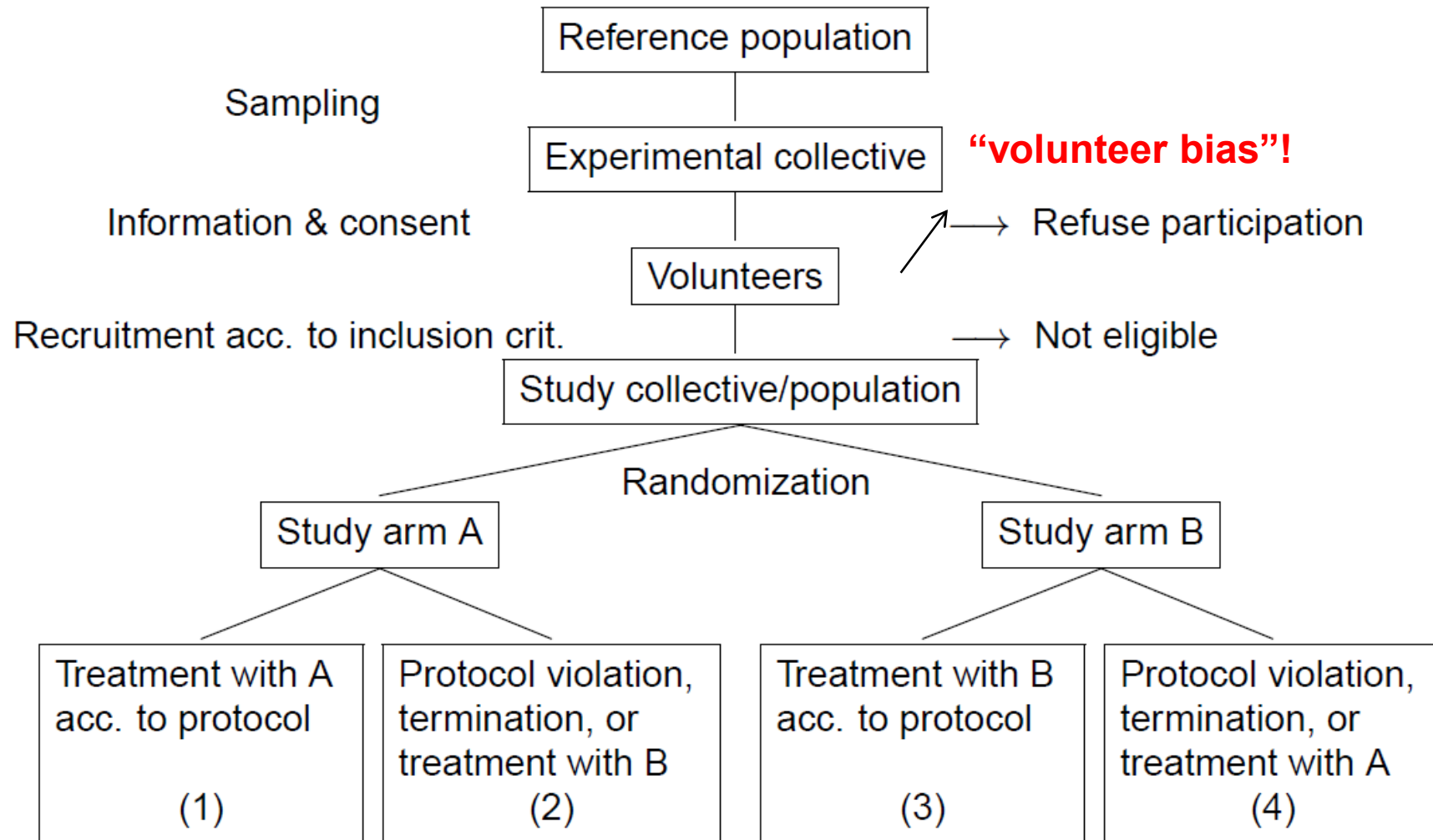
...**Quante** unità assegnare ad ogni trattamento?

...in generale occorre infatti la *replicazione* delle osservazioni, cioè l'uso di più di una unità sperimentale per ogni trattamento.



Problema della dimensione
campionaria
«*minima/ottimale*»....

Studi Clinici Sperimentali (Clinical Trials): SCHEMA



Studi Clinici Sperimentali (Clinical Trials): BLINDING (MASKING)

Per minimizzare la soggettività di chi valuta i pts o analizza i dati.

- **SINGLE BLIND:** i soggetti dello studio non sanno quale trattamento ricevono
- **DOUBLE BLIND:** i soggetti e i ricercatori (medici/biologi) non sanno
- **TRIPLE BLIND:** i soggetti, i ricercatori, gli statistici non sanno



Figure 1: The authors: double blinded versus single blinded

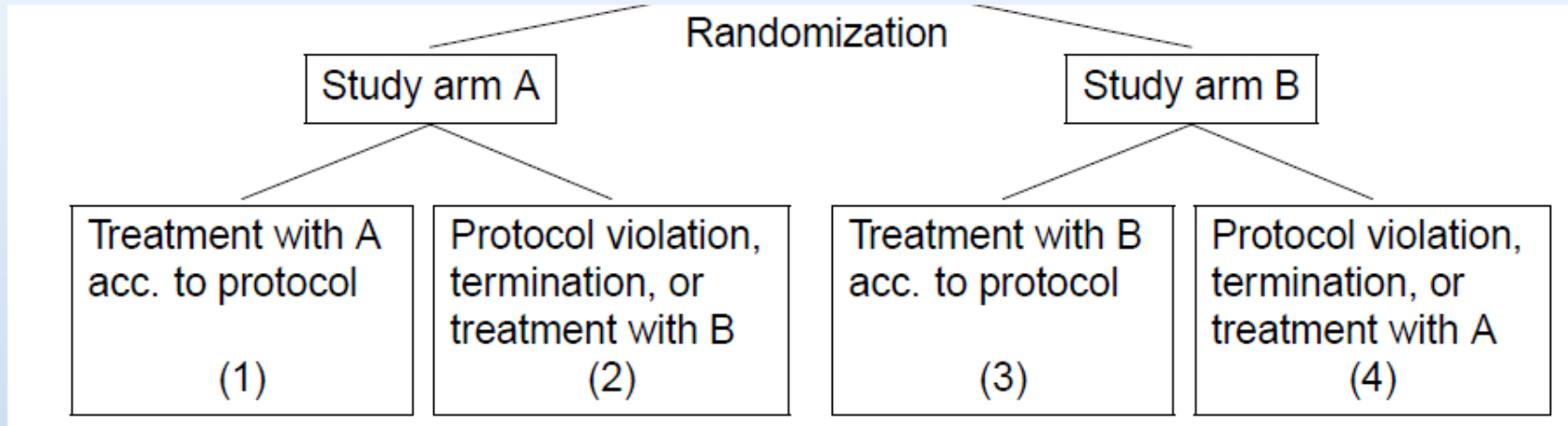
Schulz & Grimes. Lancet 2002

Se non viene effettuata: **OPEN TRIAL**



Deciphering the allocation concealment scheme

Studi Clinici Sperimentali (Clinical Trials): Analisi dei dati



ITT: intention to treat
(1)+(2) vs (3)+(4)

PP: per protocol (1) vs (3)

AT: as treated (1)+(4) vs (2)+(3)
(ma escludendo violazioni del protocollo o chiusure precoci)



GUIDELINES FOR RCTs: The «CONSORT» Statement



Enhancing the QUALity and
Transparency Of health Research



[Website translation help](#)

[Home](#) [About us](#) [Library](#) [Toolkits](#) [Courses & events](#) [News](#) [Blog](#) [Librarian Network](#) [Contact](#)

[Home](#) > [Library](#) > Reporting guideline

Search for reporting guidelines



Browse for reporting guidelines by selecting one or more of these drop-downs:

Study type
Experimental studies ▾ and Clinical area
Please select... ▾ and Section of report
Please select... ▾

Or search with free text

Search Reporting Guidelines

We recommend searching for reporting guidelines in English

[Start again](#) | [Help](#)

Displaying 230 reporting guidelines found.

Refine by: Please select... ▾

Key reporting guidelines, shaded green, are displayed first. [Show the most recently added records first.](#)



[Guidelines for Reporting Outcomes in Trial Reports: The CONSORT-Outcomes 2022 Extension](#)



[Guidelines for Reporting Outcomes in Trial Protocols: The SPIRIT-Outcomes 2022 Extension](#)



[Guidelines for cellular and molecular pathology content in clinical trial protocols: the SPIRIT-Path extension](#)



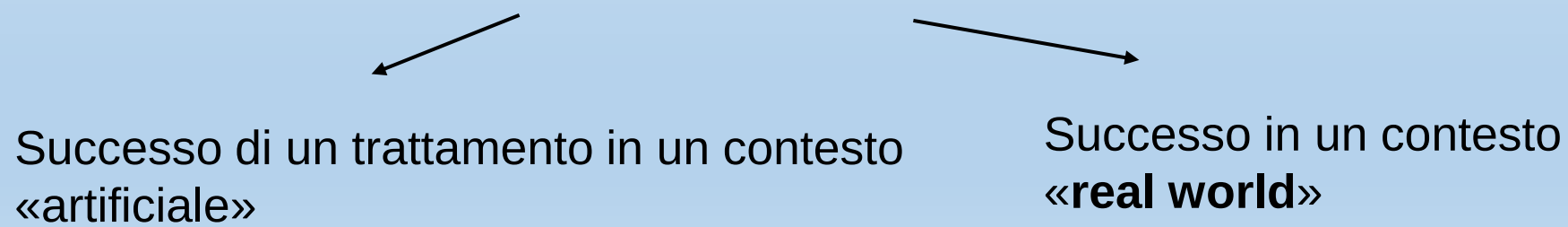
Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Study protocols	SPIRIT	PRISMA-P
Diagnostic/prognostic studies	STARD	TRIPOD
Case reports	CARE	Extensions
Clinical practice guidelines	AGREE	RIGHT
Qualitative research	SRQR	COREQ
Animal pre-clinical studies	ARRIVE	
Quality improvement studies	SQUIRE	Extensions
Economic evaluations	CHEERS	

- **«Volunteer bias»:** partecipano solo soggetti volontari (eligibilità allo studio, *compliance*, vicinanza geografica, stato socio-economico, stato di salute...)
- Criteri inclusione/esclusione molto rigidi
- Follow up «brevi»

Popolazione omogenea selezionata -> validità interna (bassa variabilità)
Ma: limite per la validazione esterna (generalizzabilità) dei risultati.

EFFICACY ≠ EFFECTIVENESS



Successo di un trattamento in un contesto
«artificiale»

Successo in un contesto
«**real world**»

STUDI OSSERVAZIONALI DESCRITTIVI

Gli studi osservazionali descrittivi si focalizzano sulla stima di un parametro di interesse in una popolazione target (es: prevalenza/incidenza...)



Cross-sectional (trasversale) : fotografa una popolazione in un certo tempo

Coorte: longitudinale

I dati:

- Fonte primaria:** raccolti dal ricercatore ad hoc per lo studio
- Fonte secondaria:** già raccolti in precedenza per altri scopi ma utilizzabili ai fini dello studio (es: dati amministrativi).

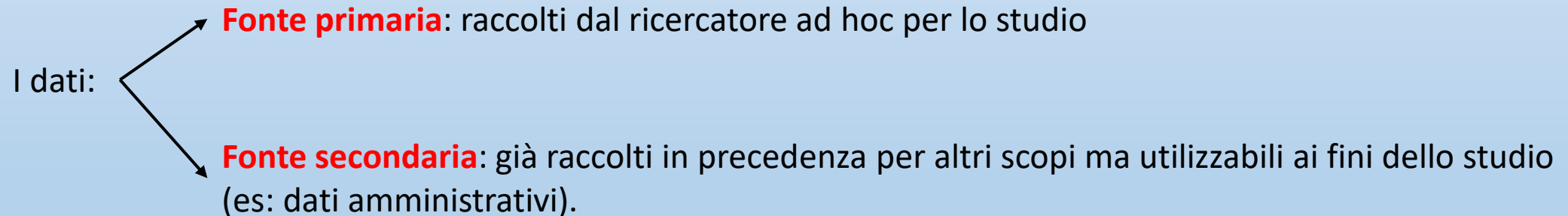
STUDI OSSERVAZIONALI ANALITICI

Gli studi osservazionali analitici si focalizzano sull'**effetto** di una «esposizione» o di un trattamento/intervento su dei soggetti.

Non c'è randomizzazione; non c'è manipolazione da parte del ricercatore.

Osservazione diretta del «**real world**»....

Confondimento: esiste una spiegazione alternativa ai risultati osservati?



Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;312:1215–1218.

Haynes B. *Can it work? Does it work? Is it worth it?* BMJ 1999;319:652.

Primary Data

Advantages

Investigator controls all aspects of the study
including design, sampling techniques, data
collection, and follow-up methods

All variables of interest can be measured

Disadvantages

Time consuming

Expensive

Secondary Data

Advantages

Relatively fast and inexpensive

Sample sizes tend to be large

Sample may cover a large geographic area and
thus provide ability to assess national trends

Unobtrusive to subjects

Disadvantages

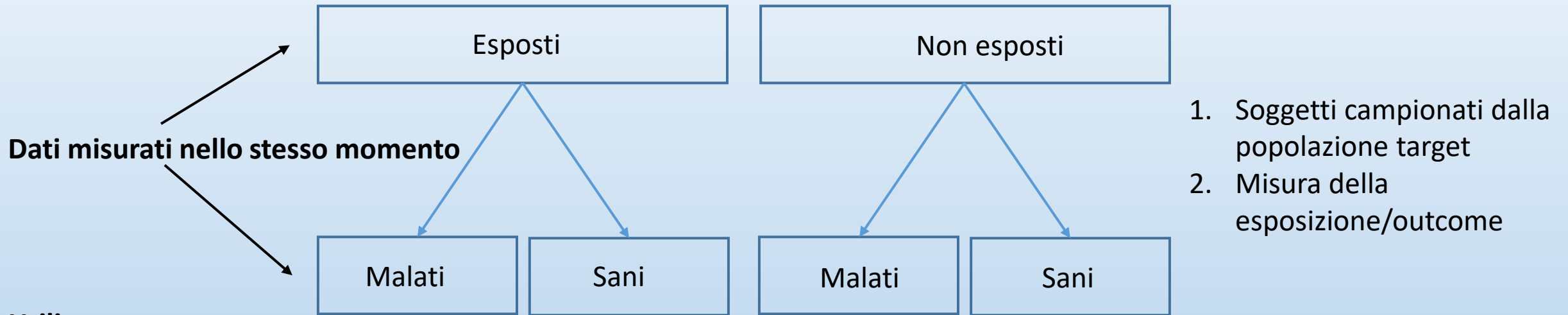
Data may not include all variables of interest
Maybe difficult to understand how and
why data elements were collected

Gli studi osservazionali analitici possono essere suddivisi in diverse **tipi** sulla base di :

- **Come** vengono **selezionati** i soggetti;
- **«Timeline»** della raccolta dei dati
- **CROSS-SECTIONAL/TRASVERSALE**
- **COORTE**
CASO-COORTE
- **CASO-CONTROLLO**
CASE-CROSSOVER

Studi Cross-sectional/trasversali

Studio osservazionale in cui l'esposizione e l'outcome sono determinati **simultaneamente** per ogni soggetto.



Utili per:

- screening delle ipotesi sul quesito in studio, rapidi da realizzare;
- valutazioni di prevalenza;
- valutazioni di fattori associati all'outcome di interesse.

Limiti:

- Non c'è evidenza temporale di causa-effetto tra l'esposizione e l'outcome;
- Outcomes **prevalenti** ma non incidenti, esclusi i soggetti deceduti prima dello studio (bias «long survivors»);
- A causa della presenza di possibili confondenti bisogna valutare spiegazioni alternative alle associazioni osservate

Studi di coorte

1. Soggetti campionati dalla popolazione di esposti e di non esposti
2. Misura dell'outcome

Raccolta dati prospettica

2024

2026

Esposti

Malati

Sani

Non esposti

Malati

Sani

Raccolta dati retrospettiva

2012

2018

Utili quando:

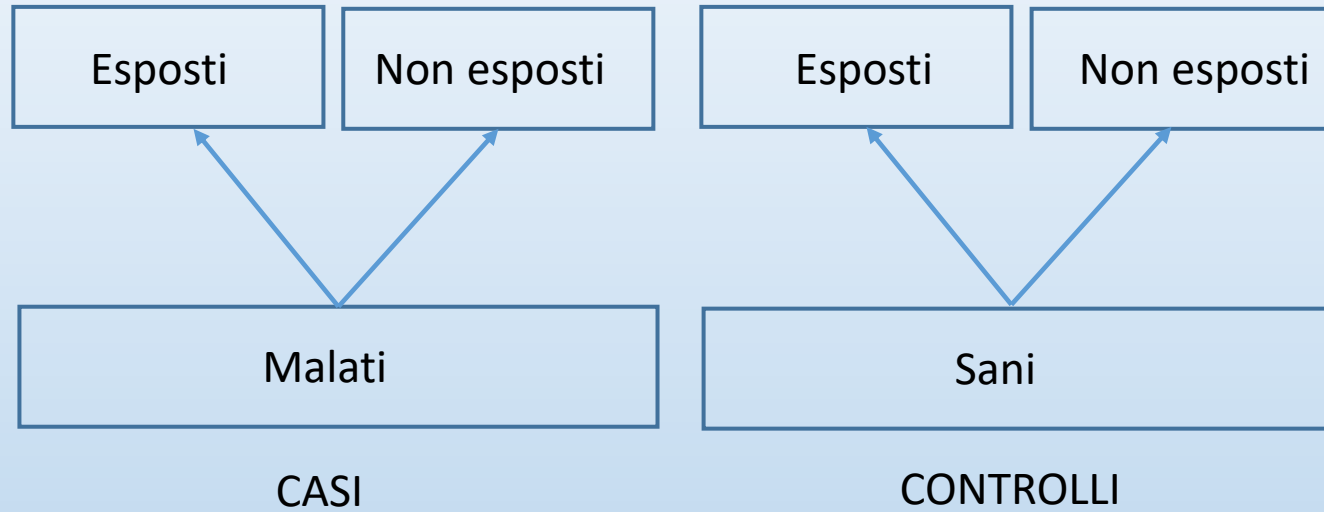
- C'è evidenza di una associazione temporale tra esposizione ed outcome;
- L'intervallo temporale tra esposizione e outcome non è troppo lungo;
- L'outcome non è troppo raro.

Limiti:

- Perdita al follow up;
- A causa della presenza di possibili confondenti bisogna valutare spiegazioni alternative
- Bias nella determinazione dell'outcome: non sono blinded!

Studi caso-controllo

1. Soggetti campionati dalla popolazione di casi e di controlli
2. Misura della esposizione



Utili quando:

- L'outcome è raro;
- C'è una ragionevole evidenza di una esposizione nel passato;

Limiti:

- «Recall/Interviewer bias»: i casi possono essere più precisi nel riferire sulla (passata) esposizione;
- Difficile la selezione di controlli adeguati;
- Se si utilizzano procedure di **«matching»** da valutare la generalizzabilità dei risultati;
- Confondenti : bisogna valutare spiegazioni alternative alle associazioni osservate

Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist



STROBE Statement

Strengthening the reporting of observational studies in epidemiology

u^b

UNIVERSITÄT
DUISBURG
ESSEN

Home

Aims

News

Available checklists

Publications

Translations

Commentaries

Discussion forum

STROBE group

Endorsement

Contact

Links

Member login / logout

What is STROBE?

STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of **STrengthening the Reporting of OBservational studies in Epidemiology**.

The STROBE Statement is being endorsed by a growing number of biomedical journals. Click [here](#) for full list.

For STROBE-related entries in PubMed click [here](#).

What's new in the STROBE Initiative?

01.09.2014

Observational Studies: Getting clear about transparency

New guidelines for observational studies in PLOS Medicine [\[more\]](#)

[\[more\]](#)

01.07.2014

New article of interest

A Review of Published Analyses of Case-Cohort Studies and Recommendations for Future Reporting [\[more\]](#)

Checklist in **22** punti:

- da consultare **prima** di pianificare lo studio
- come linee guida per la scrittura dei risultati

- Concetti generali della Statistica Descrittiva
- Richiami di rappresentazione grafiche dei dati

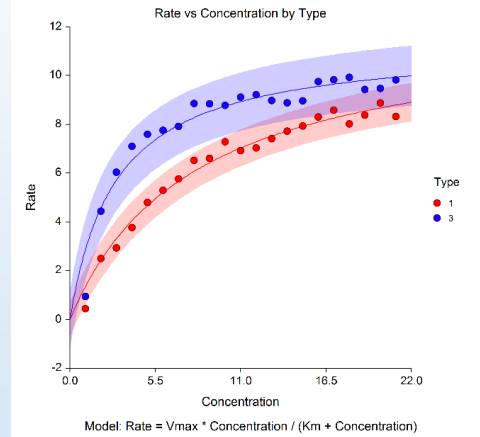
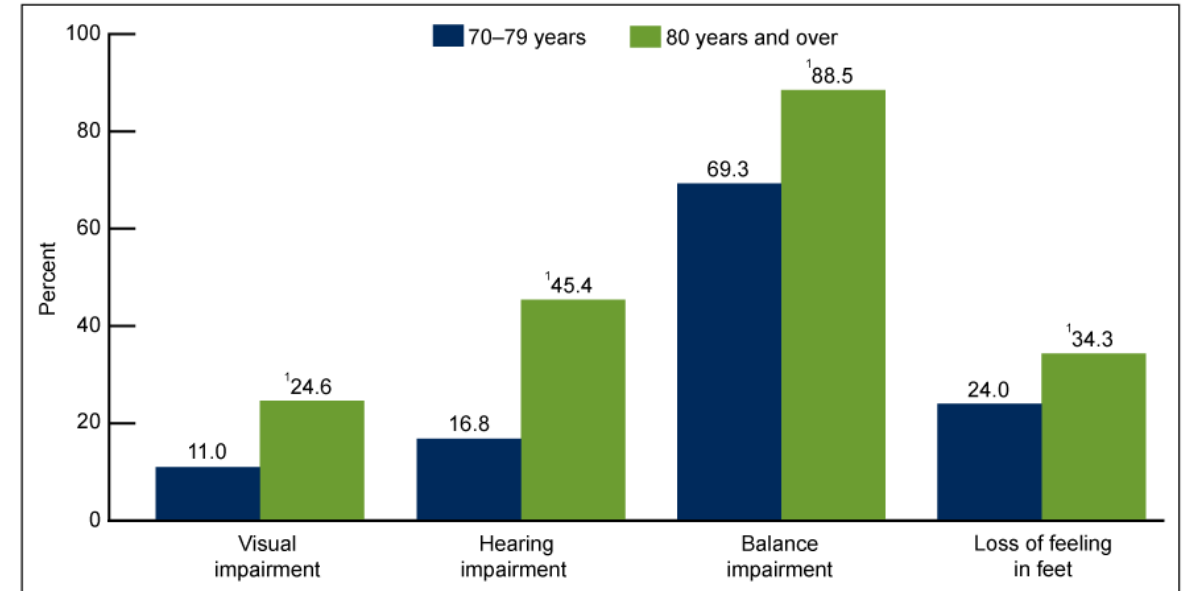


Figure 3. The prevalence of sensory impairments among persons aged 70–79 years compared with persons aged 80 years and over: United States, 1999–2006



¹Significantly different from the 70–79 age group.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey.

Terminologia

L'oggetto dell'osservazione di ogni *fenomeno individuale* che costituisce il *fenomeno collettivo* in studio è detto 'unità statistica'

Ex: un individuo, una A.S.L., un ospedale...

Ciascuna *unità statistica* presenta delle caratteristiche definite 'caratteri/variabili'

Ex: il colore degli occhi in un individuo, il numero di medici che lavorano in una A.S.L., il numero di posti letto in un ospedale....

Ciascun *carattere* è presente in ogni *unità* con una determinata 'modalità'

Ex: colore blu degli occhi, 10 medici in una certa ASL, 300 posti letto in un dato ospedale,...

La classificazione dei caratteri

I caratteri possono essere classificati in:

-Caratteri **qualitativi** (colore degli occhi, tipo di diagnosi...), a loro volta distinti in:

- ***ordinabili***: è possibile *ordinare* le modalità del carattere in senso crescente o decrescente (es: titolo di studio, livello di gravità della diagnosi...);
- ***sconnessi***: non c'è alcun ordinamento intrinseco tra le modalità (es: colore degli occhi, sesso...);

-Caratteri **quantitativi** (es: età, peso, numero di medici,...) distinguibili in:

- ***discreti***: le modalità del carattere sono numeri interi (es: numero di medici...)
- ***continui***: le modalità del carattere sono misurate su una scala continua (es: peso, altezza...).

Alla base di tale classificazione dei caratteri vi è la **'scala di misura'** con cui sono espresse le modalità: se attraverso dei **numeri** o delle **'etichette'**.

Nominale
(qualitativi
sconnessi)

Ordinale
(qualitativi
ordinabili)



«factor data»



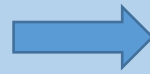
SCALA DI MISURA

«date/time data»



Discreta
(quantitativo)

Continua
(quantitativo)



«numeric data»



Nominale:

Variabili categoriali/qualitative non ordinate. Queste possono essere **binarie** (due sole categorie, come il genere: maschio o femmina) o **multinomiali** (più di due categorie, come lo stato civile: sposato, divorziato, mai sposato, vedovo, separato). La cosa fondamentale qui è che **non esiste un ordine logico** per le categorie.

Ordinale:

Categorie ordinate. Sempre variabili categoriali/qualitative, ma con un ordine. Gli elementi delle scale «Likert*» con risposte come: "*Mai, a volte, spesso, sempre*" sono ordinali.

Le variabili numeriche/quantitative possono essere ulteriormente suddivise in due tipologie: **discrete** e **continue**. Le variabili discrete, come i **conteggi**, possono assumere solo **numeri interi**: numero di bambini in una famiglia, numero di giorni persi dal lavoro. Le variabili numeriche **continue** possono assumere qualsiasi numero, anche oltre la virgola decimale.

Non è sempre ovvio che questi livelli di misurazione non riguardino solo il significato intrinseco della variabile stessa.

Altrettanto importanti sono il significato della variabile **all'interno del contesto della ricerca** e **come è stata misurata**.

***Scale likert**: costituite da un certo numero di affermazioni - definiti **item** - che esprimono un atteggiamento positivo e negativo rispetto ad uno specifico tema. La somma di tali giudizi tenderà a delineare l'atteggiamento del soggetto nei confronti del tema. Per ogni item si presenta una scala di accordo/disaccordo, generalmente a 5 o 7 modalità.

L'età è una **variabile numerica continua**. L'età di una persona è *continua* se la si misura con *sufficiente precisione*. È significativo dire che qualcuno (o qualcosa) ha 7.28 anni (?).

Età come ordinale: categorie di età.

Età come discreta: età misurata in **numero di giorni** in cui i semi germinati di una specie iniziano a germogliare. La maggior parte lo farà entro pochi giorni e il valore potrebbe variare da 2 a 9 giorni. In questo contesto, l'età è sicuramente un conteggio discreto: il numero di giorni.

Età come multinomiale: a volte le variabili numeriche sono rese categoriche a causa della mancanza di valori. In uno studio, si misuravano le età di alcuni testimoni in un processo. Tecnicamente le età sono continue, ma in questo studio c'erano solo quattro valori: 49, 59, 70 e 80.

Età come binaria: sopra/sotto una certa soglia.

Esempio: studio sull'equilibrio tra lavoro e vita privata di genitori
variabile di interesse: numero di ore lavorate a settimana.
Un predittore chiave: età del figlio più piccolo.

[C'è una differenza tra un bambino di 5 anni, che può essere idoneo solo per la scuola materna part-time e un bambino di 6 anni, che è abbastanza grande per andare a scuola a tempo pieno].

Esempio di **matrice dei dati**:

Abbiamo chiesto ad un campione di studenti se avessero visto l'ultimo Festival di Sanremo e se lo avessero gradito. Inoltre, abbiamo chiesto il **genere** (maschio/femmina) e il **numero di ore giornaliere** dedicate al sonno e ad attività di studio.

variabile
↓

Sappiamo definire la **scala di misura** di queste variabili?

Stu.	sesso	Sanremo	...	sonno	studio
1	maschio	Non l'ho visto	...	8	2
2	femmina	L'ho visto e mi è piaciuto	...	6	30
3	maschio	Non l'ho visto	...	9	5
4	femmina	Non l'ho visto	...	8	25
⋮	⋮	⋮	⋮	⋮	⋮
52	femmina	Non l'ho visto	...	8	20

←
unità statistica

Classificazione dei caratteri e scala di misura

CARATTERE		SCALA
qualitativo	Sconnesso	Nominale
	Ordinabile	Ordinale
quantitativo		Ad intervalli (scala numerica discreta o continua)

Operazioni che è possibile fare sui caratteri in base alla loro scala di misura:

Operazioni sulle modalità del carattere	Carattere		
	qualitativi		Quantitativi
	<i>sconnessi</i>	<i>ordinabili</i>	<i>(discreti/continui)</i>
= ; ≠	si	si	si
> ; <	no	si	si
+ ; -	no	no	si

Distribuzioni

Quando si rilevano le **modalità** con cui uno (o più) **caratteri** si presentano in ciascuna **unità** di una popolazione (o di un campione) si ottiene una **'distribuzione'** della popolazione/campione secondo il carattere considerato.

Ex: distribuzione 'unitaria' del peso in una classe:

Studente	Peso (kg)
Marco	62
Chiara	59
Luca	56
...	...

distribuzione «unitaria» = ad ogni unità del gruppo è associato il suo peso, cioè la corrispondente modalità del carattere considerato.

Possiamo suddividere in **'classi'** la popolazione secondo il carattere considerato, allora le modalità del carattere vengono raggruppate in classi ed otteniamo una **distribuzione di 'frequenze'** -> *frequenza della classe* = numero di unità statistiche che appartengono alla classe.

Distribuzioni di frequenze

Tipo di scuola			
		Frequenza	Percentuale
Validi	Altro	11	34,4
	Istituto Professionale	4	12,5
	Liceo Classico	3	9,4
	Liceo Scientifico	14	43,8
	Totale	32	100,0

Il numero di unità che appartengono ad una classe= **frequenza assoluta** della classe.

E' possibile calcolare anche le distribuzioni di **frequenze relative** e di **frequenze percentuali**

Alcuni simboli:

Abbiamo k classi di un dato carattere (nell'ex: 6 classi) cioè $k=6$; le possiamo indicare tramite un indice i che varia da 1 a 6: $i=1,2,3,4,5,6$. Abbiamo rilevato le modalità del carattere su N unità (nell'ex: $N=30$).

- Frequenza **assoluta** della classe i $\rightarrow n_i$
- Frequenza **relativa** della classe i $\rightarrow f_i = n_i / N$
- Frequenza **percentuale** della classe i $\rightarrow p_i = f_i * 100 = (n_i / N) * 100$

Esiste poi un simbolo matematico indica l'operazione di **somma** su un gruppo di oggetti ed è chiamato '**sommatoria**' :

$$\sum_{i=1}^k n_i = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 = 30$$

$$\sum_{i=1}^k n_i = n_1 + n_2 + n_3 + n_4 + n_5 + n_6$$

$$\sum_{i=1}^k n_i = N ; \quad \sum_{i=1}^k f_i = 1 ; \quad \sum_{i=1}^k p_i = 100$$

(per $k=6$)

Distribuzioni doppie (dati categorici/in classi)

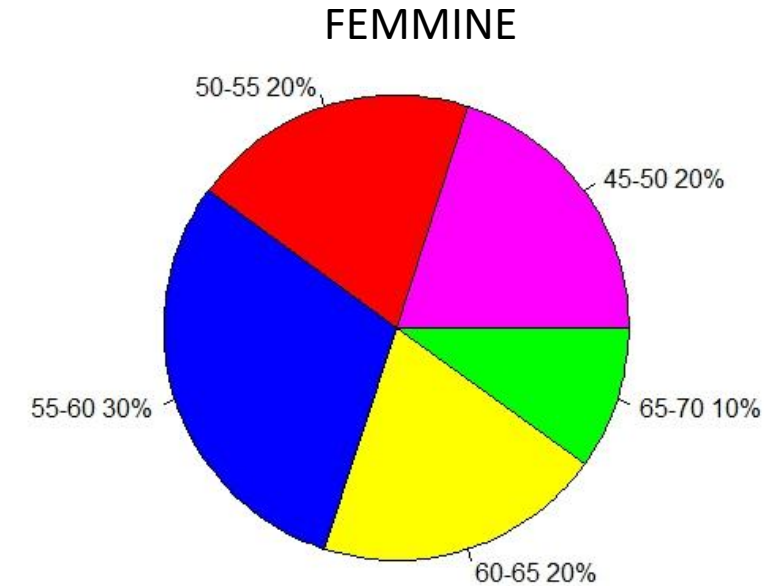
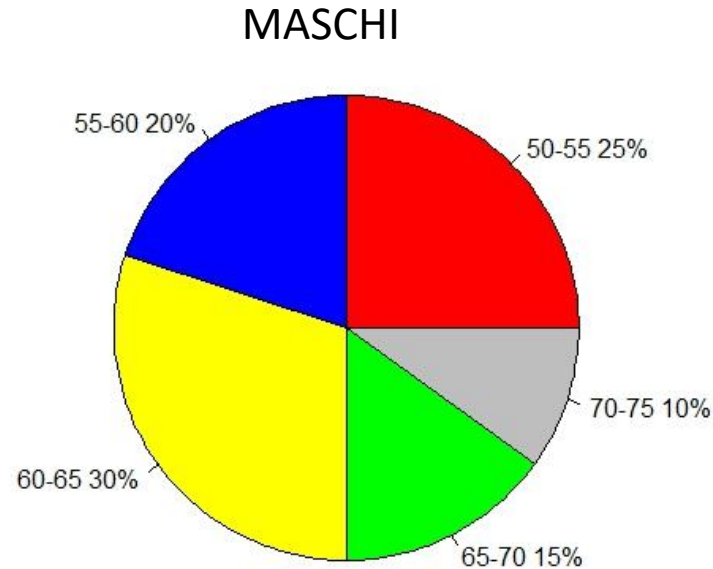
Se si rilevano **due** caratteri in una popolazione si ottiene una **distribuzione doppia**; è uso comune rappresentare le distribuzioni doppie delle modalità rilevate per due caratteri (**categorici o espressi in classi**) tramite la '**tabella di contingenza**' o '**tabella a doppia entrata**':

CLASSI DI PESO	M	F	TOT
45-50	0	2	2
50-55	5	2	7
55-60	4	3	7
60-65	6	2	8
65-70	3	1	4
70-75	2	0	2
TOT	20	10	30

Distribuzioni doppie (dati categorici/in classi)

Per confrontare la distribuzione delle classi di peso rispetto al sesso occorre calcolare le **frequenze relative** all'interno della tabella di contingenza:

CLASSI DI PESO	M	F
45-50	0	20
50-55	25	20
55-60	20	30
60-65	30	20
65-70	15	10
70-75	10	0
TOT	100	100



La tabella di contingenza permette inoltre di studiare **l'associazione** fra i caratteri analizzati
→ **test di ipotesi** che vedremo in seguito → **Statistica Inferenziale**

Tabella di Contingenza (a doppia entrata):

n_{aibj} = numero di soggetti che presentano *contemporaneamente* la modalità i del carattere A e la modalità j del carattere B.

		B				Totale
		B ₁	B ₂	...	B _m	
A	A ₁	na_1b_1	na_1b_2	...	na_1b_m	NA ₁
	A ₂	na_2b_1	na_2b_2	...	na_2b_m	NA ₂
	...	...	...	na_ib_j	...	...
	...	...	...	...	...	...
	A _k	na_kb_1	na_kb_2	...	na_kb_m	NA _k
Totale		NB ₁	NB ₂	...	NB _m	N

←
←
←
Marginali di Riga

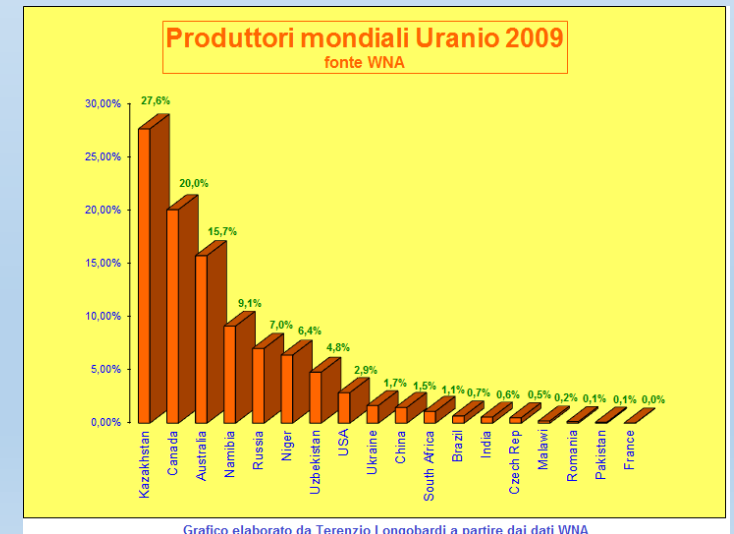
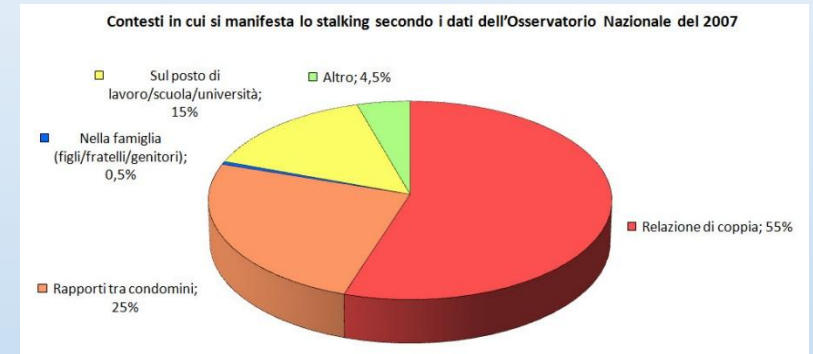
↖ ↗
Marginali di colonna

Rappresentazioni grafiche dei dati

Una parte molto importante della statistica descrittiva è la **rappresentazione grafica** dei dati



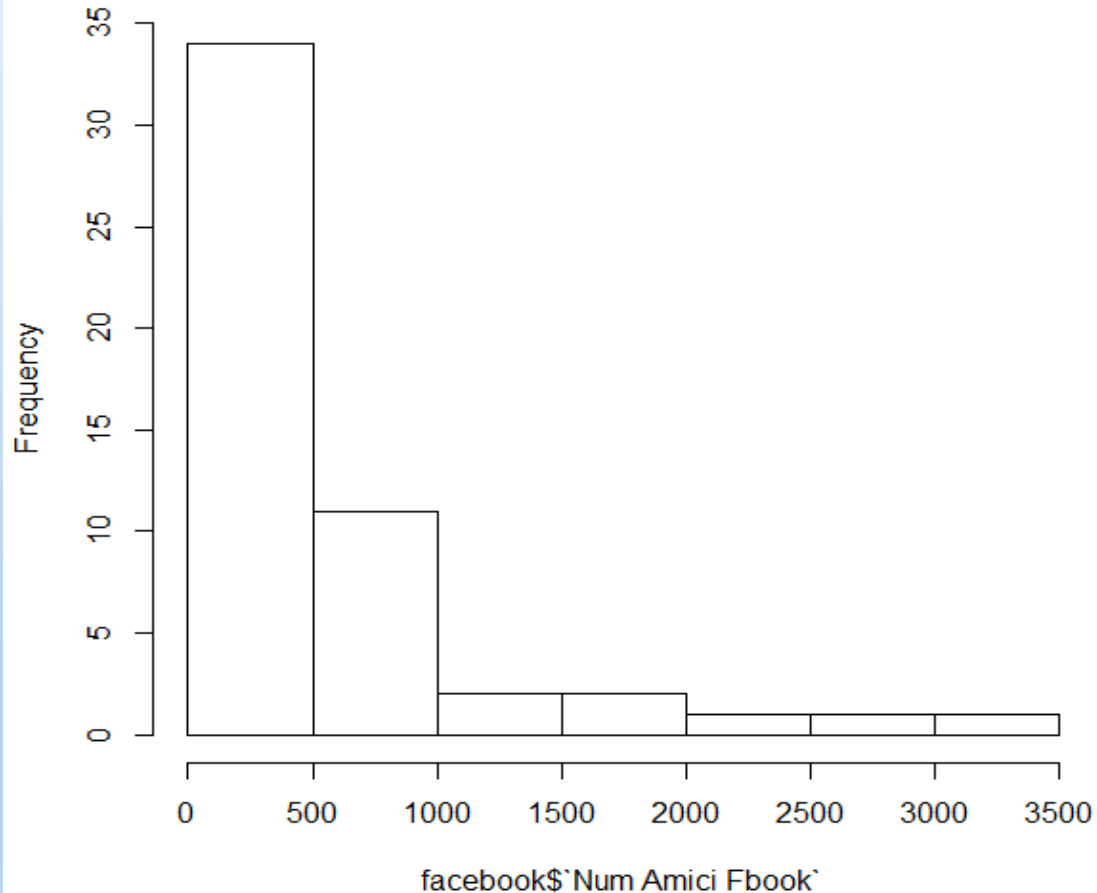
1. Scoprire la **struttura** dei dati;
2. Identificare **i valori più rilevanti** assunti dal fenomeno;
3. Identificare eventuali **valori anomali** (*outliers*)
4. **Comunicare** i risultati dello studio in modo efficace



Caratteri su scala numerica:

Per rappresentare graficamente una distribuzione secondo un carattere di tipo quantitativo possiamo utilizzare l'istogramma:

Istogramma (scelta delle classi: default di R)



Il **numero** dei rettangoli corrisponde al numero di classi in cui è stata suddivisa la variabile.

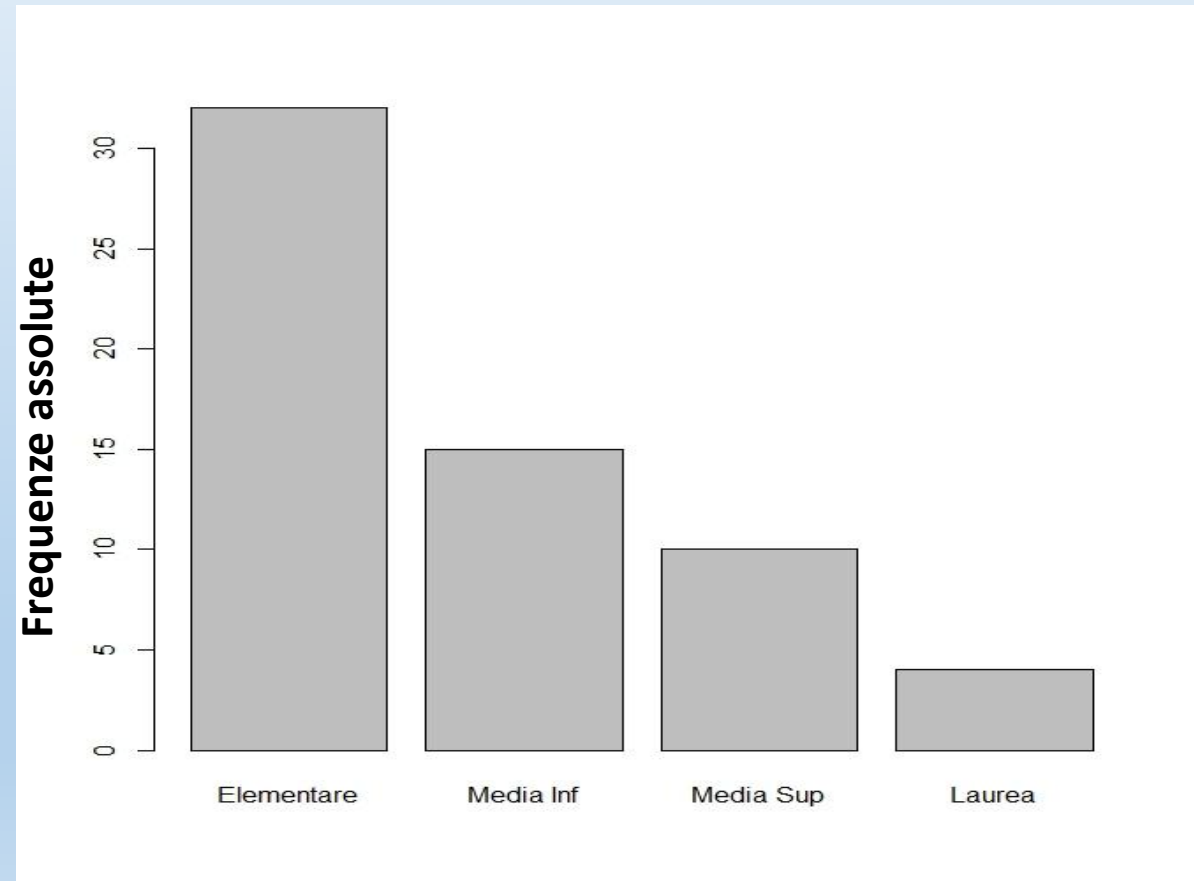
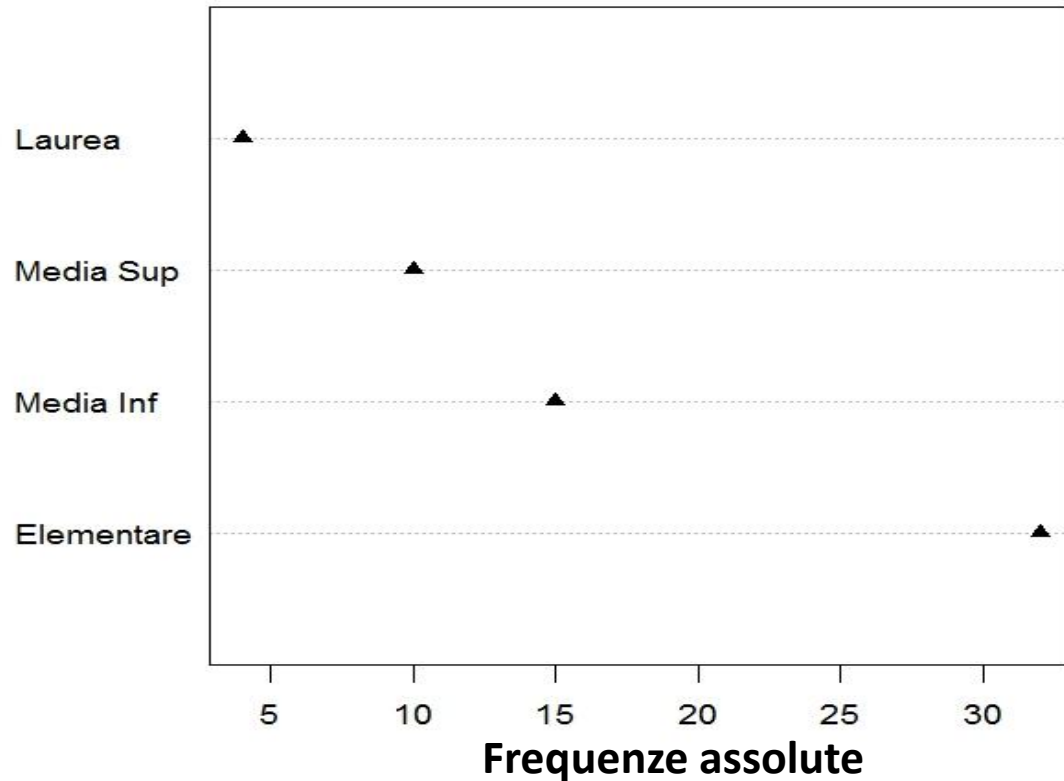
La **larghezza** della base dei rettangoli dipende dall'ampiezza di tali classi e, complessivamente, tutte le basi dei rettangoli affiancate devono coprire l'intera gamma dei valori della variabile.

L' **altezza** di ogni rettangolo invece indica la frequenza dei casi presenti in ogni classe (quante unità statistiche sono presenti all'interno della classe)[classi di uguale ampiezza]

Caratteri su scala nominale/ordinale (I):

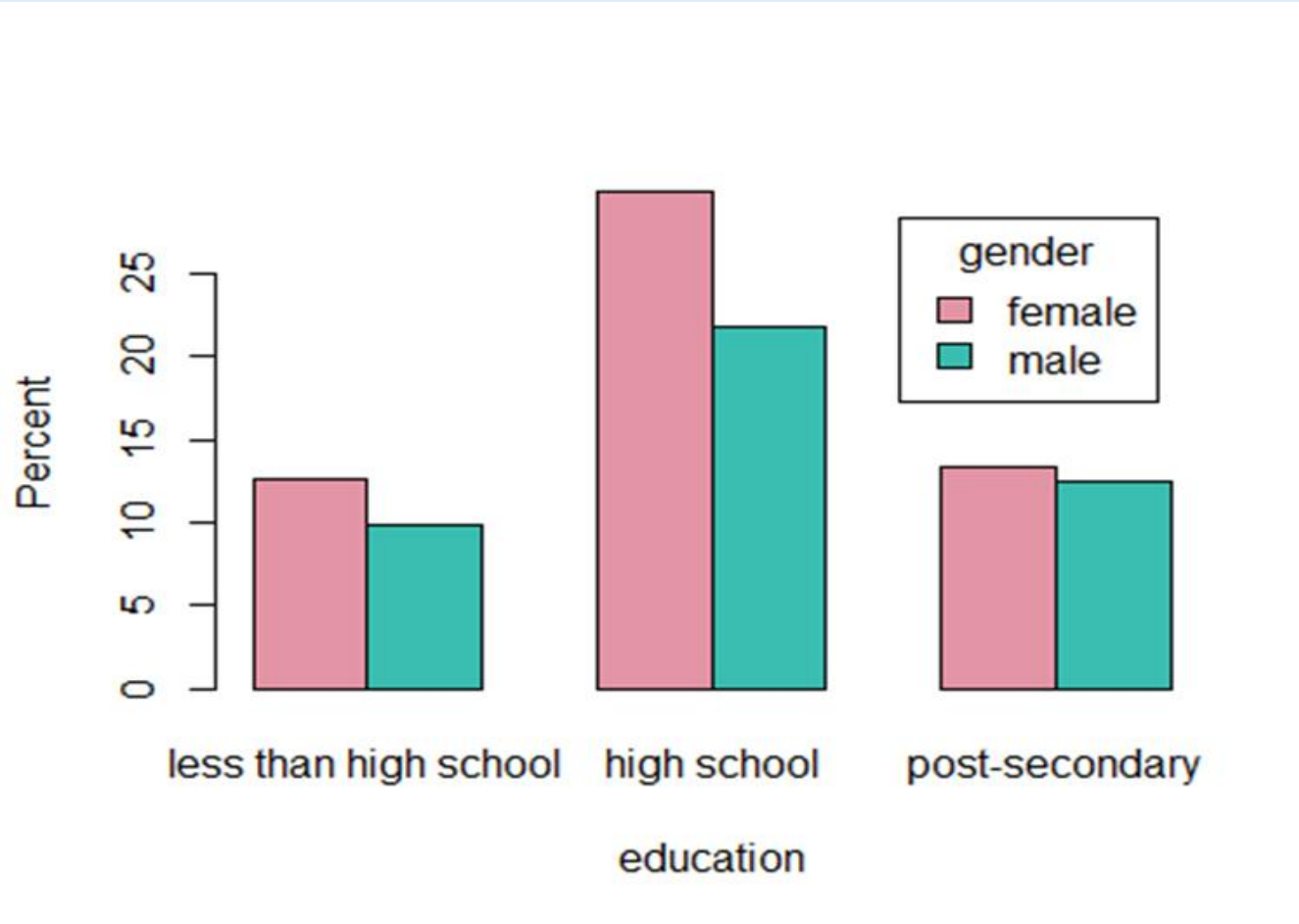
Per rappresentare graficamente una distribuzione secondo un carattere di tipo nominale o ordinale possiamo utilizzare il grafico «dotchart» oppure il «barplot»:

Elementare	Media Inf	Media Sup	Laurea
32	15	10	4



Caratteri su scala nominale/ordinale (II):

Per rappresentare graficamente una **distribuzione doppia** di due caratteri nominali/ordinali possiamo utilizzare il grafico «barplot» [meglio utilizzare le frequenze percentuali]:



Frequency **table**:

education	gender	
	female	male
high school	9987	7264
less than high school	4228	3280
post-secondary	4436	4159

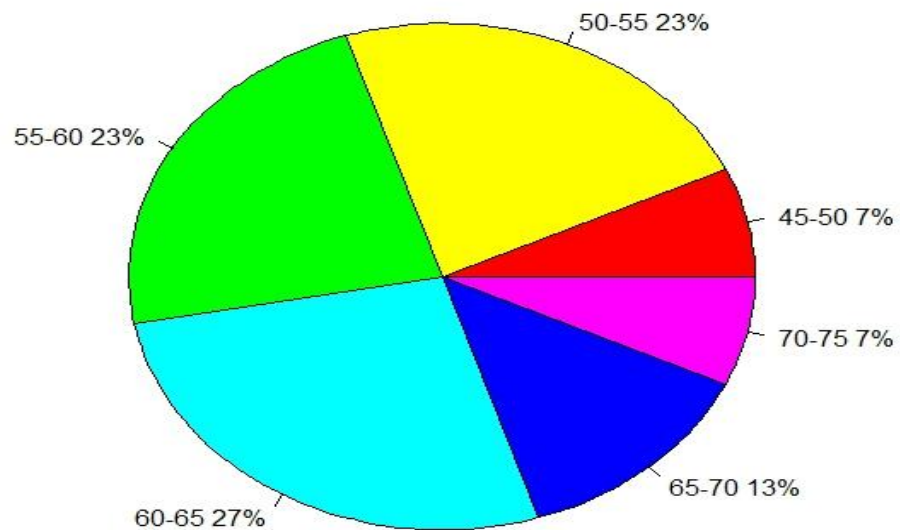
Total percentages:

	female	male	Total
high school	29.9	21.8	51.7
less than high school	12.7	9.8	22.5
post-secondary	13.3	12.5	25.8
Total	55.9	44.1	100.0

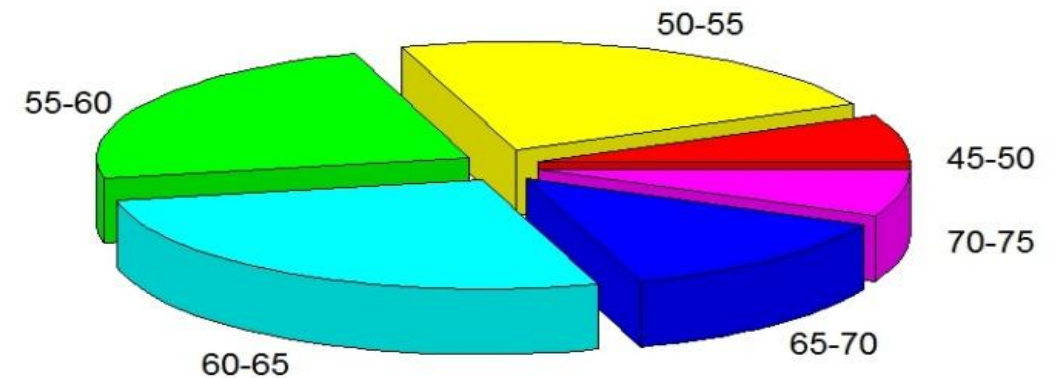
Rappresentazione grafica di distribuzioni in %

Per rappresentare una distribuzione in percentuale secondo un carattere di tipo categorico/in classi possiamo anche utilizzare il grafico a torta:

Pie Chart of Peso



Pie Chart of Peso



Distribuzioni doppie (dati su scala numerica)

Se si rilevano **due** caratteri in una popolazione su scala numerica e si è interessati a studiare l'associazione tra i due caratteri, la rappresentazione grafica usuale è lo **scatter plot (diagramma cartesiano)**:

Distribuzione su un campione di 252 maschi delle dimensioni del polso e del collo: si ipotizza che l'ampiezza del collo sia circa due volte quella del polso.

Vedremo in seguito come sintetizzare da un punto di vista statistico queste relazioni
(**correlazione/regressione lineare...**)

