

STATISTICA SOCIALE

0011 0010 1010 1101 0001 0100 1011

Introduzione alla Statistica per le scienze sociali

Prof. Claudio Capiluppi - Facoltà di Scienze della Formazione - A.A. 2007/08

STATISTICA PER LE SCIENZE SOCIALI

A cosa serve la statistica ?

0011 0010 1010 1101 0001 0100 1011

per fare **ricerca (empirica)**

Punto di partenza di una ricerca empirica:

qualcosa non torna nel nostro sistema di conoscenze,
abbiamo un problema che nasce dalle nostre conoscenze su un
determinato fenomeno di interesse, insufficienti o incoerenti

Obiettivo di una ricerca:

verificare una ipotesi sul fenomeno di interesse
scegliere tra ipotesi alternative

La statistica è uno degli strumenti del metodo scientifico moderno
necessari per progettare e condurre una ricerca che abbia una
validità scientifica

FASI DI UNA RICERCA SOCIALE (SURVEY)

Astrazione	}	Definizione del problema di ricerca: paradigma, teorie, letteratura scientifica, definizioni, costrutti teorici, ... Formulazione ipotesi di ricerca → previsione conseguenze Operazionalizzazione: definizione variabili, unità, tempi Progettazione ricerca: scelta metodo di ricerca, disegno di rilevazione, tecniche e strumenti di osservazione e misura
Rilevazione	}	Formazione del campione: tecnica selezione e dimensione Indagine pilota Indagine principale Indagine suppletiva di controllo
Elaborazione	}	Registrazione dati: codifica risposte (aperte) Analisi qualità dati: controlli validità e coerenza Revisione dati: trattamento dati anomali e mancanti Elaborazione statistica dei dati
Interpretazione	}	Verifica (statistica) delle ipotesi Interpretazione dei risultati in relazione alle teorie Comunicazione, pubblicazione dei risultati

IL PRINCIPIO DI FALSIFICAZIONE

- Verificare una ipotesi non significa dimostrare che l'ipotesi è vera
- Una proposizione di tipo generale (teoria o ipotesi) non può mai essere *verificata empiricamente*: cioè per quante osservazioni si producano, non si potrà mai dimostrare in modo definitivo che essa è vera
- La verifica empirica è un processo che, dal punto di vista logico, non può avere termine:
 - le conseguenze (enunciati particolari) che possono essere derivate logicamente da una ipotesi sono infinite
 - è sempre possibile che una nuova osservazione risulti in contrasto con una delle conseguenze previste dall'ipotesi
- È invece sufficiente che una sola osservazione risulti in contrasto con una delle infinite conseguenze derivabili dalla teoria, per dimostrarne logicamente e definitivamente la falsità: l'ipotesi si dice falsificata
- Quindi quando si parla di verificare una ipotesi cosa si intende?
 - quello che si valuta è se l'ipotesi risulta compatibile con la evidenza empirica disponibile (i dati osservati finora): l'ipotesi si può considerare (temporaneamente) confermata
 - se invece una delle conseguenze previste dall'ipotesi risulta incompatibile con i dati osservati, l'ipotesi è logicamente falsificata

IL PRINCIPIO DI FALSIFICAZIONE

Una proposizione generale può essere dimostrata *logicamente* (dedotta) da altre proposizioni generali *vere*, quando è già implicitamente contenuta in esse

E' il caso delle dimostrazioni matematiche (teoremi), dove le conclusioni (tesi) sono derivate per deduzione a partire da premesse (ipotesi)

Le ipotesi sono a loro volta derivate da altre e, in ultima analisi, dagli assiomi (o postulati), affermazioni vere per assunzione.

Gli assiomi non possono essere dimostrati, sono il punto di partenza di tutto l'edificio matematico, che si sviluppa in modo esclusivamente deduttivo:

- tutte le conclusioni sono già contenute nelle premesse,
- quindi tutto lo sviluppo della disciplina consiste "solo" nell'esplicitarle

Deduzione: procedimento logico che procede dal generale al particolare

IL PRINCIPIO DI FALSIFICAZIONE

- **Evoluzione del concetto di scienza**
- Ideale classico (Aristotele): conoscenza dimostrativa basata sulla *Deduzione*; tutte le conclusioni derivano dai postulati (verità per assunzione)
- Ideale positivismo (Bacone): la conoscenza si fonda sull'osservazione dei fatti e l' *Induzione* è il procedimento che permette di arrivare a formulare (scoprire) le leggi della natura
- Epistemologia moderna (Popper): il criterio di scientificità di una teoria è la sua *Falsificabilità*, cioè che sia possibile in linea di principio dimostrarne l'eventuale falsità
 - non siamo disposti a credere (a ragion veduta, non per atto di fede) a qualcosa che non è possibile verificare se sia falso
 - potremmo dire che, per lo scienziato, ciò che non può essere falso, non può nemmeno essere vero
 - quindi nulla se ne può dire (sempre a ragion veduta), ovvero se ne può dire tutto (tanto non è falsificabile) => non ci interessa

IL PRINCIPIO DI FALSIFICAZIONE

- Asimmetria logica tra verifica e falsificazione: tutte le teorie sono provvisorie
 - oggi ci sembrano vere quelle attuali, e ci sembra impossibile che in passato, magari solo pochi anni fa, si potesse credere che ...
 - ma *tutte* le teorie sono destinate ad essere falsificate !
- La falsificazione delle teorie attualmente dominanti deve essere anzi uno degli obiettivi primari della scienza e degli scienziati: è il vero motore del progresso scientifico
- La scienza non produce verità assolute: tutte le teorie, anche quelle oggi dominanti, sono ipotesi che si sono finora dimostrate valide cioè adeguate a descrivere i fenomeni osservati e a produrre previsioni, sempre confermate
- Le teorie falsificate dovrebbero essere abbandonate:
 - c'è la tendenza, anche da parte degli scienziati, a tentare di difenderle, introducendo ulteriori ipotesi ad-hoc per negare la falsificazione
 - non sempre sono proprio "da buttare": possono conservare una validità in un campo di applicazioni più limitato
- A volte la falsificazione di una teoria può comprometterne altre e mettere in discussione l'intero paradigma, innescando una rivoluzione scientifica

IL METODO SCIENTIFICO

- Come si giunge alla formulazione di una ipotesi empirica ?
 - l'**Induzione**, secondo la visione positivista, è il procedimento "logico" di passaggio dal particolare al generale: da singoli fatti osservati (enunciati particolari) ad una teoria o "legge" (enunciato generale)
 - ma non è un procedimento logico: non garantisce la verità della conclusione (che è + generale) a partire da quella delle premesse
- "Razionalità" della natura: le leggi scientifiche esistono in natura, prima che lo scienziato le "scopra" ?
 - le categorie logiche della ragione (es. causalità) sono dipendenti dalla mente dell'uomo, e non invece corrispondenti a nessi reali e naturali
 - lo scienziato *non scopre leggi esistenti in natura, ma inventa modelli* per descrivere/spiegare i fenomeni: il momento induttivo rappresenta il momento creativo del processo scientifico
 - questi modelli/teorie si rivelano adeguati per un certo campo di applicazioni, per un certo periodo... fino a quando non vengono falsificati da nuove osservazioni, che ne mostrano l'inadeguatezza
 - l'avvenuta falsificazione costringe gli scienziati a formulare una nuova teoria, in grado di rendere conto anche delle nuove osservazioni

IL METODO SCIENTIFICO

- Il mondo reale è veramente come noi ce lo rappresentiamo ?
 - conosciamo il mondo attraverso strumenti di rilevazione: filtri sensoriali (specie-specifici) con precise specifiche e limitazioni
 - i dati rilevati vengono poi sottoposti ad elaborazione: una prima è spesso integrata nell'organo di senso (es. la retina), seguita da quella eseguita a livello del SNC (es. integrazione del campo visivo)
 - il risultato di tutti questi processi è il mondo come ci appare, la nostra rappresentazione della realtà, ma non coincide con la realtà: è una astrazione della realtà
- Tutti i dati su cui ci basiamo sono il prodotto di un processo di rilevazione condotto con uno strumento di osservazione, le cui caratteristiche condizionano, anzi determinano i dati prodotti
- E' importante conoscere e tenere sempre presenti le caratteristiche dello strumento che stiamo utilizzando: sia che sia un sofisticato apparecchio, piuttosto che un questionario, o il nostro occhio ...

Esempi:

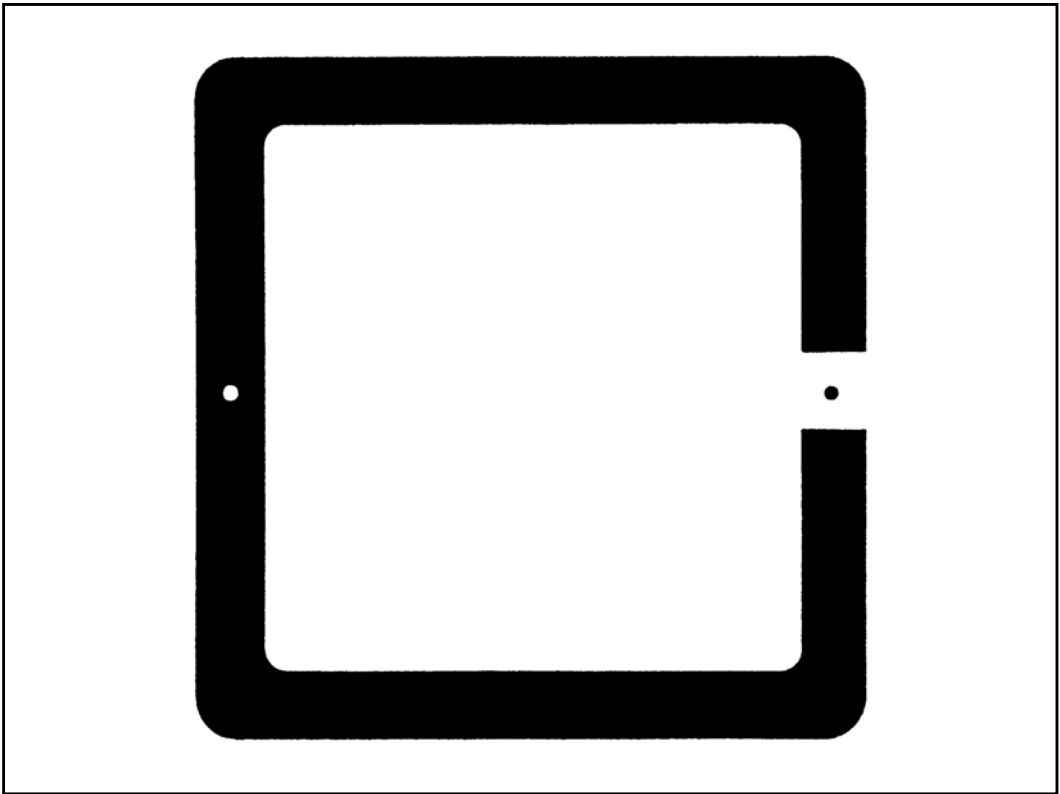
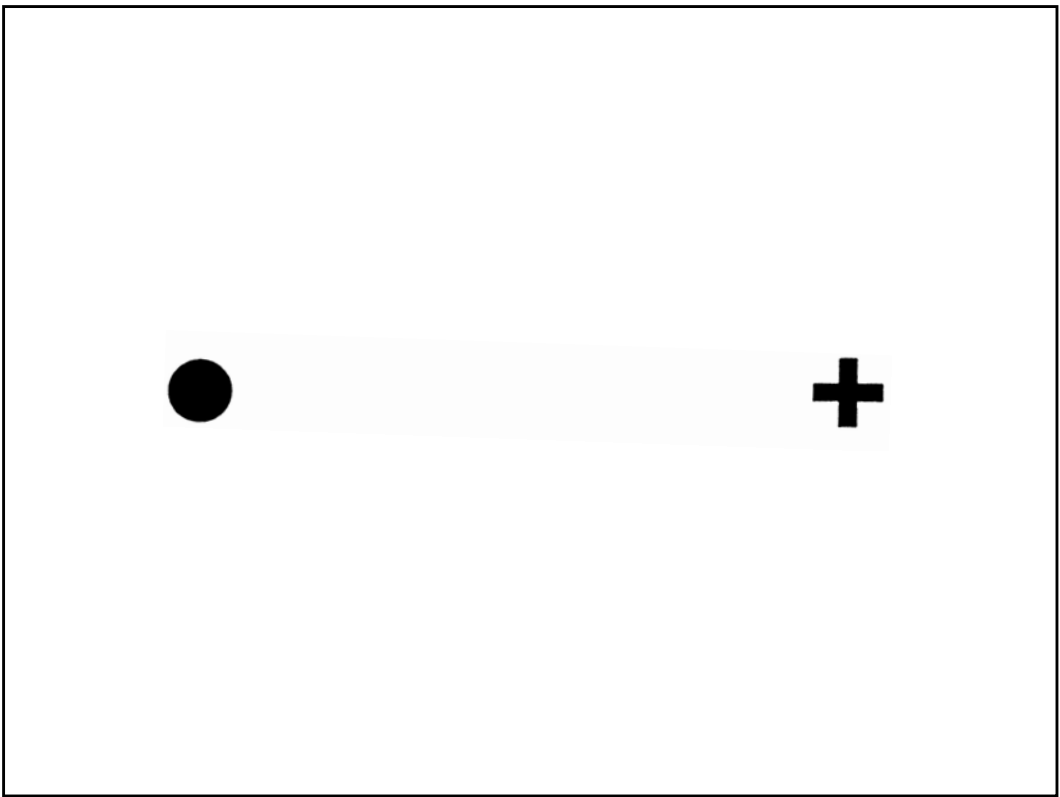
Il nostro concetto di buio fa riferimento allo spettro di radiazioni elettromagnetiche che risultano a noi visibili e che chiamiamo luce. Quello che ai nostri occhi è buio, non lo è per altri occhi...

Gli insetti vedono un mondo completamente diverso dal nostro: ad es. vedono gli ultravioletti, per noi invisibili. I serpenti rilevano invece gli infrarossi, quindi possono individuare un topo nel "buio" più assoluto

Una rana non vede oggetti di grosse dimensioni proprio davanti a lei fino a quando restano fermi... ma individua con precisione un piccolo insetto che si muove velocemente nel suo campo visivo

La retina non è una semplice *bitmap*, una matrice di punti come il sensore di una telecamera, ma un processore neurale che "riconosce" a livello hardware *pattern*, forme, schemi, movimenti

L'osservazione è sempre una selezione di alcuni aspetti della realtà, magari osservati tramite una lente distorta



Nella parte nasale della retina di ciascun occhio c'è una zona priva di recettori (*macula caeca*)



Per far sparire la croce, chiudete l'occhio sinistro e fissate il disco con l'occhio destro



Per far sparire il disco, chiudete l'occhio destro e fissate la croce con l'occhio sinistro

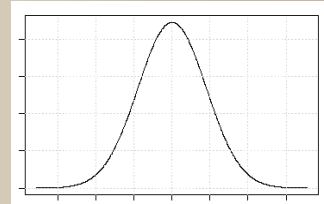
È facile spiegare perché, in visione binoculare, la *macula caeca* eluda la nostra osservazione: la zona non rappresentata nell'occhio sinistro è rappresentata nell'occhio destro, e viceversa. Ma nemmeno in visione monoculare compare un buco; e non perché la *macula caeca* sia piccola, dato che si estende per circa 6° in orizzontale e 8° in verticale, a un'eccentricità di circa 13-19° in direzione nasale (quanto basta per farci cadere dentro un uovo tenuto in mano a braccio stesso). Il buco è invisibile perché ha luogo un'interpolazione visiva: una superficie omogenea continua ininterrotta; una linea non si spezza in due segmenti (figura 3.5). Si realizza cioè un riempimento (*filling-in*) coerente con la metafora del colore che si spande finché non incontra un contorno.

Benché l'interpolazione nella *macula caeca* si presti a interpretazioni contrastanti [Dennett 1991; Kanizsa 1991, 115], possiamo trarre una conclusione generale, valida anche per altri completamenti (par. 3.5). L'assenza di informazione locale va distinta dall'informazione sull'assenza. Per vedere una macchia, un buco o un troncamento, l'occhio deve registrare i contorni che ne segnalano l'esistenza.

Cosa studia la Statistica ?

0011 0010 1010 1101 0001 0100 1011

DISTRIBUZIONE POPOLAZIONE



Soggetti (Unità statistiche)

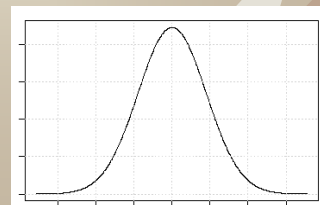
Caratteristiche (Variabili)

Osservazioni/Misurazioni (Modalità)

45

TERMINOLOGIA STATISTICA

- **Popolazione:** insieme delle unità (es. soggetti) di interesse per una ricerca
- **Unità statistiche:** soggetti che costituiscono la popolazione e sui quali si effettuano le osservazioni
- **Variabili:** caratteristiche dei soggetti che vengono osservate e misurate
- **Modalità:** valori o qualità che una variabile può assumere
- **Distribuzione:** descrive il modo in cui un carattere si manifesta in una popolazione, quali modalità sono più frequenti e quali sono invece più rare da riscontrare sulle unità della popolazione



TERMINOLOGIA STATISTICA

- **Popolazione:**
in statistica il concetto di popolazione è da intendersi in senso molto generale, può essere:
 - reale (empirica) / virtuale (teorica)
 - finita / infinita
 - osservata / potenziale
 - di riferimento (pop. obiettivo) / pop. raggiungibile

- **Distribuzione:**
 - Frequenze (osservate) / Probabilità (teoriche)

- **Popolazione ⇔ Distribuzione**
la popolazione di cui si tratta in statistica è sempre una popolazione di *osservazioni (misure)*

TERMINOLOGIA STATISTICA

- **Unità statistiche:**
 - i "soggetti" che costituiscono la "popolazione":
sono le unità portatrici dell'informazione di interesse
 - possono essere individui reali (persone, animali, ...), oggetti fisici
 - oppure unità di riferimento di informazioni relative a eventi o processi (es. giorni, misure ripetute)

- **Unità di rilevazione:**
 - l'unità che si contatta per rilevare le informazioni sui soggetti di interesse
 - può essere diversa dall'unità portatrice dell'informazione di interesse (es. famiglia / componenti)

TERMINOLOGIA STATISTICA

■ Le variabili:

- L'osservazione non è un'operazione neutrale: quello che osserviamo è sempre una astrazione della realtà; si osserva, *si vede*, ciò che interessa, selezionando, separando, alcuni aspetti e ignorandone altri
- La definizione della variabile si colloca sempre all'interno di un paradigma scientifico, culturale e sociale: complesso di teorie e conoscenze accettate da una comunità scientifica-sociale (es. intelligenza, ansia, disoccupazione, ...)
- Non esistono fatti *puri*, svincolati da qualunque teoria, indipendenti da qualunque presupposto teorico: i dati sono un prodotto culturale (ideologico)
- La definizione della variabile di interesse è un costrutto *teorico*: non è direttamente osservabile, misurabile

variabile teorica ->

operazionalizzazione

=> **variabile misurabile**

- L'operazionalizzazione è la traduzione della definizione teorica in termini di operazioni di misurazione empirica:

protocollo di osservazione = strumento e operazioni di misura

TERMINOLOGIA STATISTICA

■ Scala di misura di una variabile:

- **Qualitativa** (o **Categoriale**): descrive la qualità di una caratteristica
 - scala **Nominale**: qualità (nomi, aggettivi) non commensurabili (sconnesse)
 - scala **Ordinale**: qualità ordinabili
- **Quantitativa**: misura numerica di intensità/grandezza di una caratteristica
 - scala **Intervallo**: i valori sono sommabili, ovvero ha senso calcolare la differenza tra due valori, e differenze uguali indicano una uguale differenza di intensità della caratteristica
 - scala **Rapporto**: i valori sono anche rapportabili, cioè il rapporto tra due valori corrisponde ad un reale rapporto di intensità della caratteristica (es. 4 kg pesa il doppio di 2 kg); questo accade quando esiste uno 0 assoluto (completa assenza della caratteristica)
- Una variabile quantitativa può essere inoltre:
 - **Discreta**: può assumere solo valori discreti (numeri interi)
 - **Continua**: può assumere qualunque valore compreso in un intervallo (numeri reali)
- La scala dipende dall'operazionalizzazione, cioè da come misuriamo la variabile

TERMINOLOGIA STATISTICA

- **Variabile Dicotomica** (o Binaria): può assumere solo due modalità
 - può essere quantitativa (es. 0/1) o qualitativa (es. bianco/nero, sì/no, presenza/assenza, trattamento/placebo)
 - possiede proprietà di entrambe le classi e può essere trattata di volta in volta come qualitativa o come quantitativa

- **Livelli di una variabile:**
 - modalità /valori che una variabile può assumere (cioè che vengono presi in considerazione) nell'ambito di una ricerca
 - se la variabile è manipolata (somministrata), possono essere scelti dal ricercatore in base alle ipotesi che si vogliono testare (es. dosaggi di un farmaco)
 - in altri casi invece non possono essere assegnati, ma sono invece rigidamente prestabiliti (es. maschio/femmina)
 - il *numero* di livelli considerati in una ricerca è una scelta che ha un grande impatto sulla sua realizzazione, soprattutto se sperimentale: determina il numero di osservazioni e di soggetti necessari per studiare il fenomeno e verificare l'ipotesi

LO STRUMENTO DI MISURA

- **Strumento (processo) di rilevazione :**
 - *cosa misura realmente ?* problema di adeguatezza dell'operazionalizzazione adottata (es. ansia, intelligenza, ...)
 - più o meno adeguato che sia, ha sempre caratteristiche e limiti intrinseci, dei quali bisogna essere consapevoli e tenere conto
 - l'osservazione stessa può alterare l'oggetto di studio (es. microscopio elettronico, effetto Hawthorne, presenza osservatore, intervista faccia a faccia, ...)
 - dobbiamo ricordarcene soprattutto quando si interpretano i risultati, per non far dire ai dati cose che *non possono dire*

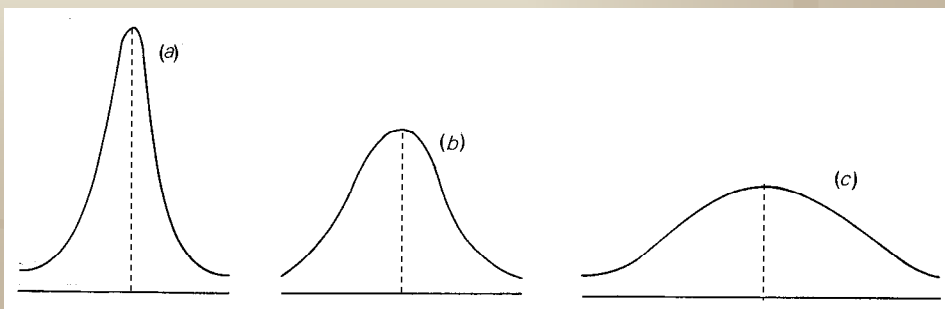
- Le misure prodotte sono sempre *stime* : il *vero* valore è ignoto, e non è possibile conoscerlo con certezza
- Le misure prodotte sono delle "buone" misure ?

LO STRUMENTO DI MISURA

- La valutazione di uno strumento di misura si basa su 4 aspetti principali:
 - **adeguatezza** (validità concettuale o di costrutto): lo strumento misura *veramente* quello che si vuole valutare, ovvero è adeguato rispetto agli obiettivi conoscitivi
 - **oggettività**: diversi osservatori producono la stessa valutazione dello stesso evento, ovvero la misura è indipendente dall'osservatore
 - **fedeltà** (stabilità, affidabilità): capacità di ri-produrre le stesse misure, nelle stesse condizioni; si possono avere invece fluttuazioni dello strumento
 - **sensibilità**: capacità di rilevare differenze (discriminazione), grado di informatività portata dalla misura
 - ...
 - **risoluzione**: più piccola grandezza rilevabile (a inizio scala e su tutto il campo di misura)
 - **prontezza**: rapidità con cui lo strumento è in grado di produrre la misura

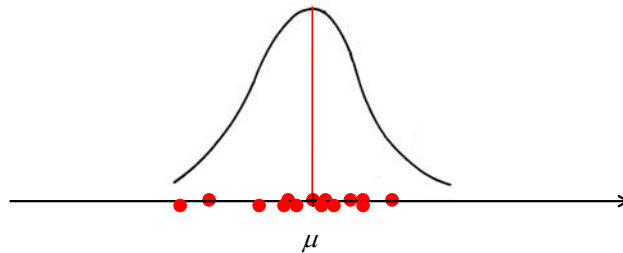
LO STRUMENTO DI MISURA

- Due parametri di valutazione di uno strumento di misura propriamente *statistici* (si riferiscono alla distribuzione delle misure prodotte) sono:
 - **correttezza (non-distorsione)**: la misura prodotta è *centrata* sul valore vero, o invece sbaglia *sistematicamente* in una direzione (sottostima/sovrastima) ?
 - **precisione**: quanto fluttua la misura intorno al valore vero, ovvero quanto le misure prodotte dallo strumento sono disperse intorno al valore centrale ?



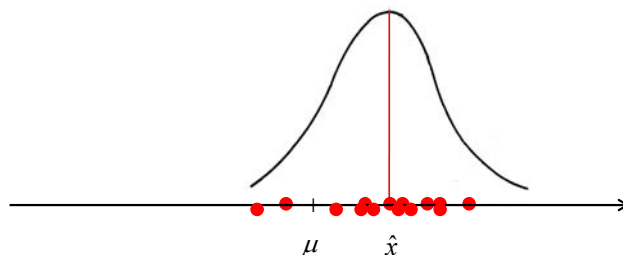
LO STRUMENTO DI MISURA

- Se ripetiamo l'operazione di misura molte volte (al limite infinite volte) otteniamo misure sempre diverse (poco o tanto)
- Come risultato avremo una distribuzione delle misure prodotte dallo strumento.
- Sulla base delle misure osservate possiamo operare una **stima**: che può coincidere oppure no con il vero valore (ignoto)



LO STRUMENTO DI MISURA

- Quando la distribuzione delle misure osservate non è centrata sul vero valore, la stima che operiamo sulla base del nostro strumento non è corretta: si dice che è una stima **distorta**
- Se ad es. le misure prodotte sono sistematicamente spostate a destra del valore vero, il nostro strumento conduce ad una *sovrastima* della variabile misurata:



- La **distorsione** è una delle maggiori problematiche per la validità di una ricerca

TERMINOLOGIA STATISTICA

- **Ruolo delle variabili** all'interno di una (ipotesi di) ricerca
- Quando si ipotizza una relazione di causa-effetto tra due variabili, si distinguono :

$$X \rightarrow Y$$

- variabile **dipendente (Y)** : o risposta, esito, effetto
- variabile **indipendente (X)** : o esplicativa, predittiva, fattore, trattamento
 - **manipolata (controllata)**: la variabile viene assegnata ai soggetti dal ricercatore in modo da poterne controllare l'effetto (es. farmaco, metodo sperimentale)
 - **non manipolata dal ricercatore**: possiamo solo osservare e prendere atto dei valori assunti dalla X; il valore su cui è attestata la variabile è in qualche senso preesistente alla ricerca (es. fumatore/non fumatore, tutte le variabili *soggettive*)
- variabile di **disturbo** : o di confusione, estranea, parassita
 - **interferisce** con i risultati della ricerca facendo variare la Y in modo non previsto (es. altri fattori non considerati che influenzano la Y)
 - per questo si dice che introduce "errori" sulla Y (scostamenti dall'atteso), che possono essere casuali o sistematici

LA CAUSALITA'

- Supponiamo di aver rilevato su un insieme di soggetti due variabili quantitative, ad es. altezza e peso
- Si ha una dipendenza tra due variabili quando si osserva una relazione tra i valori assunti dalle due variabili:

- la relazione si dice di tipo **causale**, quando una variazione del valore assunto da una delle due variabili causa una conseguente modificazione del valore dell'altra variabile

$$X \rightarrow Y$$

- altrimenti non è identificabile una causa e un effetto, ma si osserva solo una *associazione* tra i comportamenti delle due variabili

$$X \leftrightarrow Y$$

- In assenza di qualunque relazione, due variabili si dicono **indipendenti**
- E' importante sottolineare che ipotizzare una relazione di causa-effetto è una operazione concettuale che va al di là dei dati osservati e anche dell'analisi statistica:

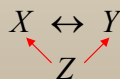
la causalità è una categoria del pensiero umano

LA CAUSALITA'

- La causalità è una categoria del pensiero umano, che ha la disposizione a costruire associazioni tra eventi: è una *funzione* e un *prodotto* della nostra mente (o SNC)
- Si tratta di una qualità *adattiva*, evolutasi in funzione di un migliore adattamento all'ambiente, in quanto consente di *prevedere* situazioni, comportamenti, eventi ... come conseguenze di altri eventi osservati: è una capacità che dà un vantaggio nella competizione per le risorse ambientali (se conduce a previsioni corrette...)
- Come vengono costruite tali relazioni causali nella nostra mente ?
 - alcuni tipi di causalità sono *percepiti*, rilevati a livello di percezione (processi primari), sono funzioni innate, incorporate nella memoria genetica
 - altri sono appresi, sulla base di associazioni tra eventi (condizionamento): le associazioni possono essere costruite in modo arbitrario, ed eventualmente erroneo (es. pensiero primitivo, pensiero del bambino, superstizioni ...)
 - l'uomo ha poi sviluppato il pensiero logico, un ulteriore step evolutivo, in grado di migliorare la correttezza delle previsioni operate: ma anche l'adulto commette ancora spesso degli errori logici (es. post hoc ergo propter hoc)
- Molte credenze e superstizioni sono basate su associazioni arbitrarie:
 - ad es., uno studente si porta all'esame un amuleto portafortuna, l'esame va bene, e quindi conclude che l'amuleto funziona. così decide di portarlo sempre agli esami. se un esame successivo va male però... dovrebbe riconoscere che l'ipotesi è stata falsificata e abbandonarla, invece la credenza rimane !
 - anche il gioco d'azzardo è basato su questi meccanismi...

LA CAUSALITA'

- L'attribuzione di un nesso di causalità è sempre una nostra proiezione: a volte ci sembra naturale (a torto o a ragione), altre volte è molto problematica, al punto che si può confondere la causa con l'effetto !
- Di per sé stessi, i dati permettono di rilevare solo una *associazione*, più o meno forte, tra fenomeni, comportamenti, eventi
- Talvolta due variabili sono strettamente associate, cioè presentano andamenti molto simili, ma nessuno penserebbe di dire che esiste una relazione causale, né in una direzione e nemmeno nell'altra
- Osservare una forte associazione non significa quindi che la prima variabile abbia un effetto causale sulla seconda, cioè che aumentando il suo valore si possa indurre un aumento (o una diminuzione) nel valore assunto dall'altra
- Una forte associazione tra due variabili può essere dovuta, ad es. a un terzo fattore, che agisce direttamente su entrambe :



- Quando ipotizziamo una relazione di causa-effetto, possiamo costruire un modello per prevedere Y in funzione di X: poi dovremo andare a verificare quanto queste previsioni sono valide

LA CAUSALITA'

- Per poter essere di tipo causale, una relazione deve avere :
 - una direzione determinata, dedotta di solito dalla sequenza temporale
 - è necessario avere il controllo delle variabili esplicative (presunti fattori causali), cioè poter assegnare e modificare i valori che assumono
 - occorre garantire l'assenza o comunque controllare anche eventuali fattori di confusione, che possono compromettere la correttezza dell'inferenza
- La possibilità di fare una affermazione di tipo causale dipende in modo fondamentale dal metodo di ricerca utilizzato e dal grado di controllo esercitato sulle variabili esplicative e di disturbo -> **metodo sperimentale**

LA CAUSALITA'

Metodo sperimentale

i livelli delle variabili indipendenti sono somministrate ai soggetti dal ricercatore

le altre variabili che potrebbero influire sulla variabile dipendente sono *tenute sotto controllo*

Controllo:

esperimento in condizioni di isolamento (in laboratorio)

assegnazione randomizzata dei livelli di trattamento ai soggetti

Solo quando si applica il metodo sperimentale si può essere ragionevolmente sicuri che la modificazione della variabile dipendente sia stata effettivamente provocata dalla variazione dei fattori ipotizzati