



UNIVERSITÀ
DEGLI STUDI DI TRIESTE



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

Introduzione all'inferenza statistica

Francesco Pauli (DEAMS, Università di Trieste)

Sommario

- 1 Concetti alla base, popolazione, parametro, campione, stima
- 2 La distribuzione campionaria
- 3 Statistica in guerra
- 4 Riepilogo

Cos'è l'inferenza statistica?

In breve, c'è **qualcosa** che vogliamo conoscere ma **non possiamo** semplicemente **osservarla** direttamente.

Possiamo però **osservare un'altra cosa**, che è **diversa** da quello che vogliamo sapere ma è **più probabile che sia simile che non che sia distante**.

Sommario

- 1 Concetti alla base, popolazione, parametro, campione, stima
- 2 La distribuzione campionaria
- 3 Statistica in guerra
- 4 Riepilogo

Popolazione e parametro

- Consideriamo una **popolazione**,
 - ad es.: corpo elettorale della CdD italiana;



Popolazione e parametro

- Consideriamo una **popolazione**,
 - ad es.: corpo elettorale della CdD italiana;
- Di questa popolazione interessa una certa caratteristica, questa è detta **parametro**,
 - ad es.: percentuale di elettori del PD.



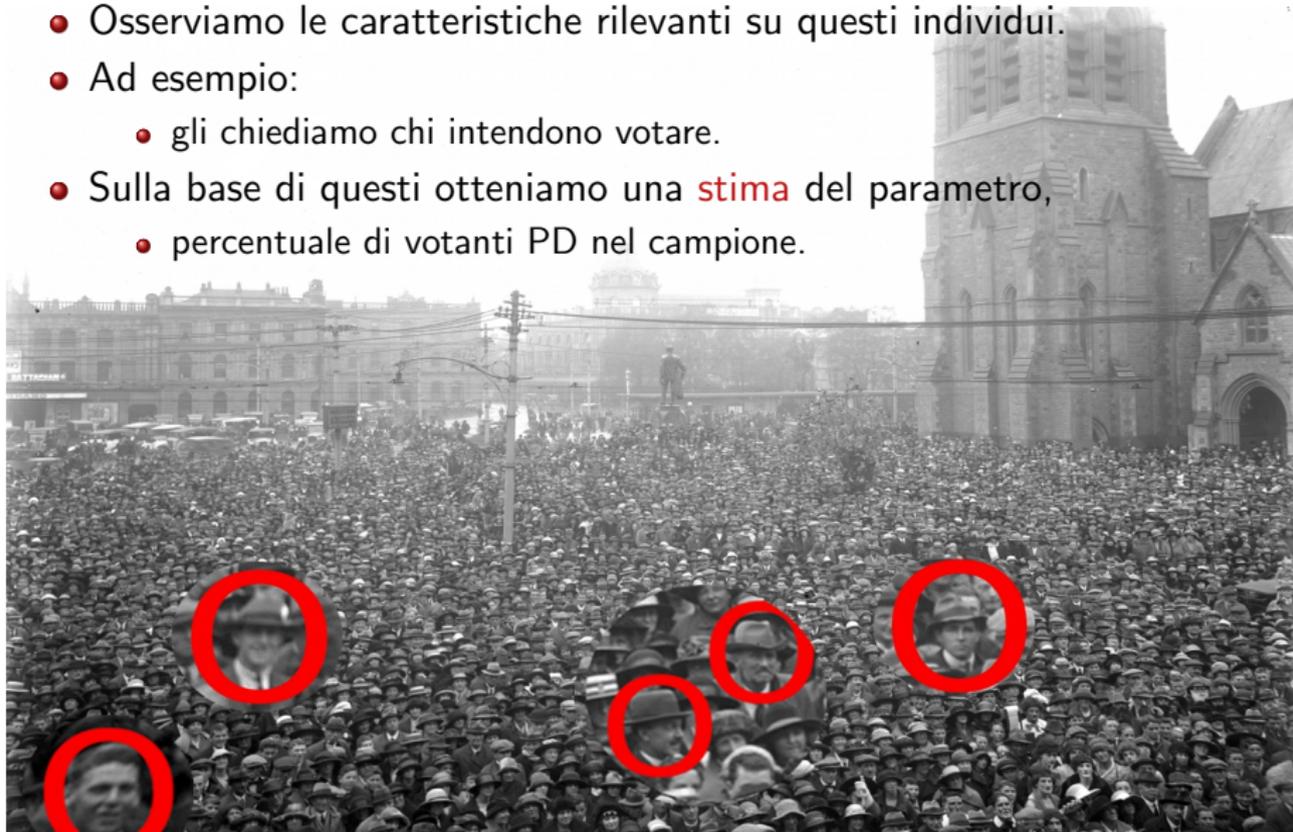
Campione

- Non possiamo/vogliamo osservare tutta la popolazione (troppo grande, non c'è tempo),
- Scegliamo **a caso** alcuni individui: il **campione**.

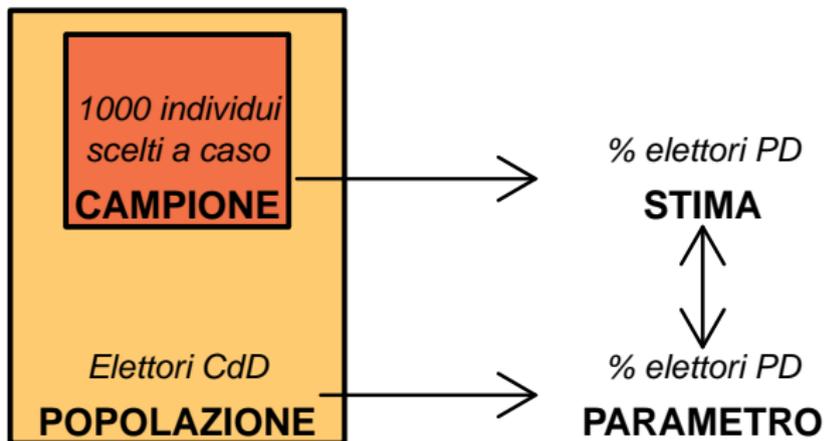


Campione e stima

- Osserviamo le caratteristiche rilevanti su questi individui.
- Ad esempio:
 - gli chiediamo chi intendono votare.
- Sulla base di questi otteniamo una **stima** del parametro,
 - percentuale di votanti PD nel campione.

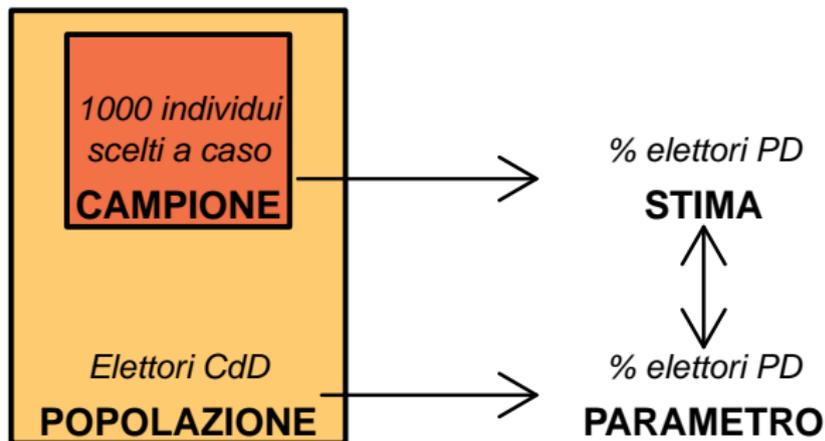


Schematicamente: popolazione-parametro, campione-stima



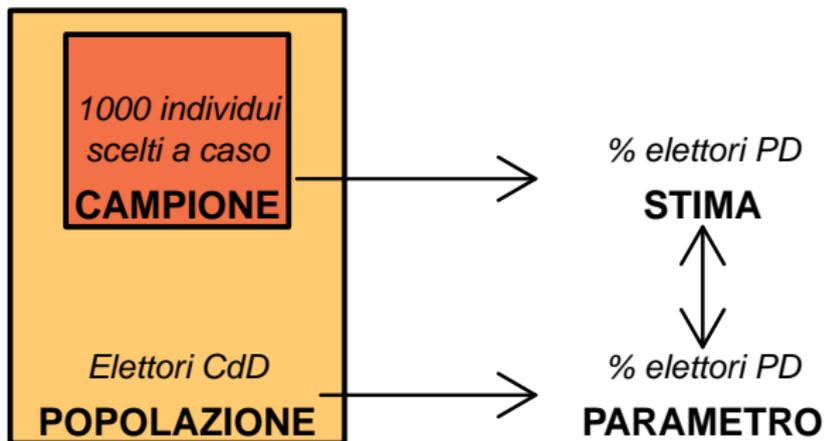
- Non è ovvio, ancora, cosa sia una stima e come sia collegata al parametro.

Schematicamente: popolazione-parametro, campione-stima



- Non è ovvio, ancora, cosa sia una stima e come sia collegata al parametro.
- Per intanto una stima è una caratteristica quantitativa del campione che si ritiene 'simile' al parametro.

Schematicamente: popolazione-parametro, campione-stima



- Vediamo intanto alcuni esempi, poi discuteremo perché questa dovrebbe essere 'simile' al parametro.

Cos'è l'inferenza statistica?

Vogliamo conoscere una caratteristica di una **popolazione** – il **parametro** – ma non possiamo osservarla direttamente.

Osserviamo un sottoinsieme della popolazione – il **campione** – dal quale calcoliamo la **stima**, che è diversa dal parametro ma è più probabile che sia vicino ad esso che lontano.

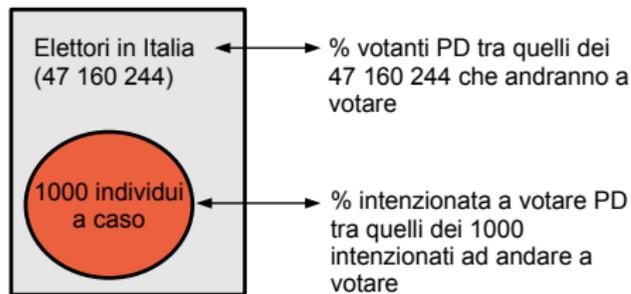
Esempio: sondaggio elettorale

Quanti voteranno PD alla Camera dei Deputati nelle prossime elezioni?



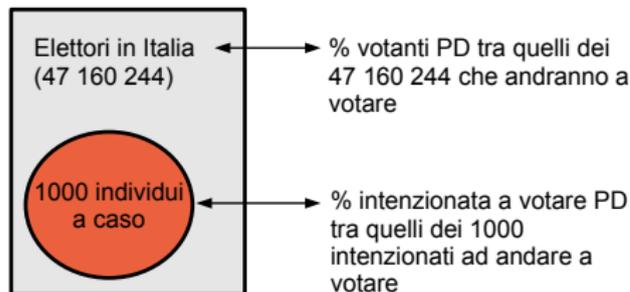
Esempio: sondaggio elettorale

Quanti voteranno PD alla Camera dei Deputati nelle prossime elezioni?



Esempio: sondaggio elettorale

Quanti voteranno PD alla Camera dei Deputati nelle prossime elezioni?



In un sondaggio del 28-29 gennaio (Istituto Piepoli per Sky TG24) la proporzione è risultata pari a 31,5% .

Esempio: stranieri e lavoro

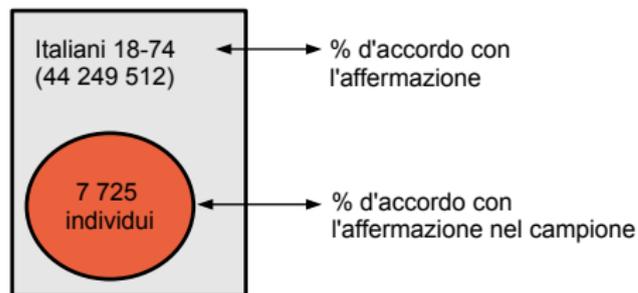
Quante sono le persone dai 18 ai 74 anni d'accordo con l'affermazione 'in condizione di scarsità di lavoro i datori di lavoro dovrebbero dare la precedenza agli italiani'?



Immigrati a Ellis Island, 1902

Esempio: stranieri e lavoro

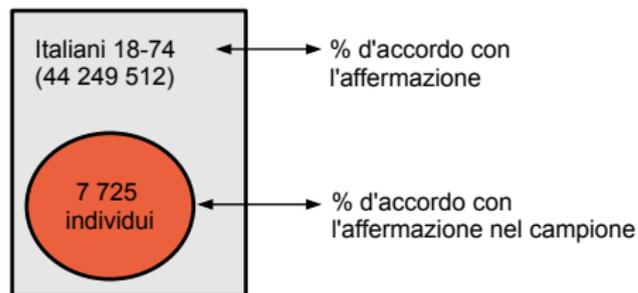
Quante sono le persone dai 18 ai 74 anni d'accordo con l'affermazione 'in condizione di scarsità di lavoro i datori di lavoro dovrebbero dare la precedenza agli italiani'?



Immigrati a Ellis Island, 1902

Esempio: stranieri e lavoro

Quante sono le persone dai 18 ai 74 anni d'accordo con l'affermazione 'in condizione di scarsità di lavoro i datori di lavoro dovrebbero dare la precedenza agli italiani'?



Immigrati a Ellis Island, 1902

Secondo l'ISTAT la percentuale è intorno al **48.7%** (I migranti visti dai cittadini, 2011).

Come dev'essere il campione

Quando diciamo che il campione sono n individui selezionati nella popolazione, questo non vuol dire che qualunque gruppo di n individui vada bene.

Campione “rappresentativo”

Un campione “rappresentativo” è un sottoinsieme della popolazione che ne riflette le caratteristiche.
(Una versione in miniatura della popolazione.)

È il fatto che il campione è rappresentativo che consente di generalizzare i risultati che si ottengono sulla base di calcoli fatti sul campione, alla popolazione.

Come NON dev'essere il campione

NON si ottiene un campione rappresentativo

- prendendo le persone presenti in quest'aula,
- prendendo gli amici/parenti/conoscenti,
- ponendo una domanda in una trasmissione televisiva e invitando il pubblico a rispondere via telefono o sms o internet.

questi gruppi di persone hanno caratteristiche peculiari, non possiamo escludere che queste siano legate alle caratteristiche che stiamo indagando, quindi introdurremmo delle distorsioni.

Per grande che sia, un campione non rappresentativo non consente generalizzazioni.

Come ottengo un campione rappresentativo?

L'idea è di selezionare le unità da includere nella popolazione in modo casuale, poi ci sono diversi metodi

- Il modo più semplice è scegliere n individui in modo che **ciascun individuo della popolazione abbia la stessa probabilità di essere estratto**.
- Altre opzioni sono spesso usate allo scopo di
 - migliorare la rappresentatività,
 - semplificare la procedura (risparmiare quattrini);

tra queste

- campione stratificato,
- campione a grappoli,
- campione a più stadi;

tutti possono essere estratti, le probabilità possono variare.

Campione casuale \neq stime non distorte

Anche se il campione è scelto in modo casuale, non è detto che si siano eliminati tutti i problemi.

- Sottocopertura: può essere che il campione venga dalla popolazione sbagliata, un'indagine telefonica 'manca' tutti quelli che non hanno telefono fisso.

Anche se il campione è costruito perfettamente, le stime possono essere distorte da altri fenomeni:

- non risposte: i democratici tendono a rispondere più frequentemente dei conservatori;
- false risposte: si pensi a un'indagine sul consumo di stupefacenti.

Campione nei sondaggi elettorali

I sondaggi elettorali sono fatti con una logica diversa, e i risultati che vengono forniti sono basati anche su considerazioni politologiche, in particolare si discosta da quello che abbiamo definito campione qui perché

- non copre tutta la popolazione ma solo le persone reperibili per telefono (spesso solo telefono fisso);
- la non risposta potrebbe essere collegata con il voto (si afferma spesso che chi vota a destra risponde meno frequentemente);
- momento del sondaggio temporalmente distante dall'elezione, le persone possono cambiare idea.
- Spesso, s'impiegano campioni non probabilistici (per quote).

Campione per quote

- La popolazione è divisa in strati (e.g. per età e sesso).
- All'interno degli strati il campione è scelto per convenienza.
- Il campione somiglierà alla popolazione quanto alla frequenza delle caratteristiche oggetto di stratificazione.
- Può esserci distorsione da selezione.
- La generalizzabilità non è garantita.

Esempio: consumi

Non sempre la popolazione è composta da individui.

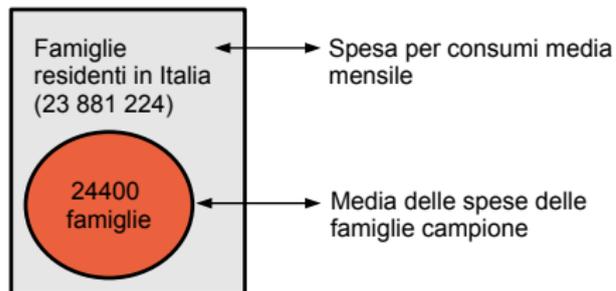
Quanto spende in media una famiglia italiana ogni mese?



Esempio: consumi

Non sempre la popolazione è composta da individui.

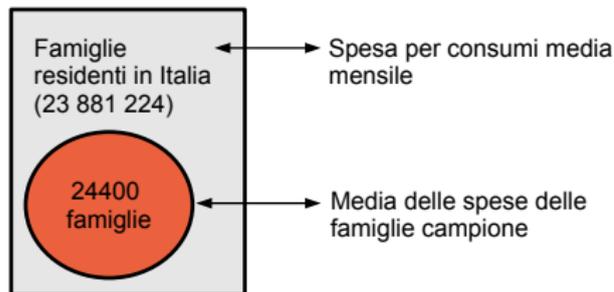
Quanto spende in media una famiglia italiana ogni mese?



Esempio: consumi

Non sempre la popolazione è composta da individui.

Quanto spende in media una famiglia italiana ogni mese?



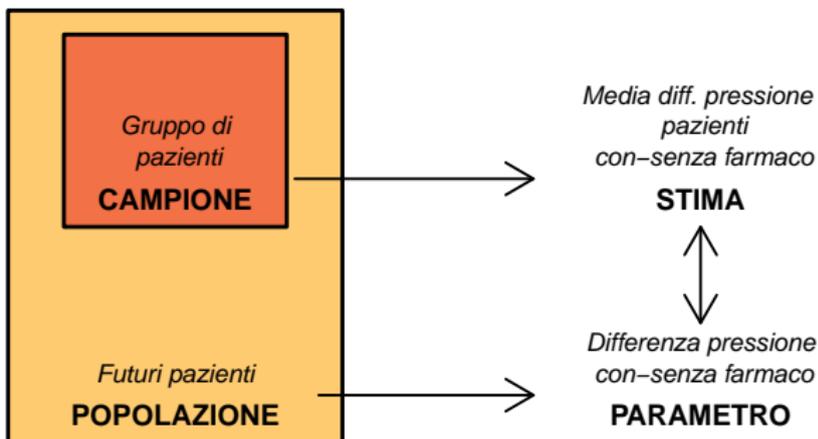
Secondo l'ISTAT la spesa media è **2480 €** (Indagine corrente sui consumi delle famiglie) e l'intervallo di confidenza è **2450-2510 €**.

Generalizzazione dello schema

- I concetti sono stati introdotti con riferimento a una situazione tipica che costituisce un prototipo del problema statistico:
 - c'è una popolazione ben definita (e finita);
 - il campione è un sottoinsieme scelto a caso nella popolazione.
- Questo non è necessariamente vero, come negli esempi seguenti.
- Con qualche *caveat* però la logica dell'inferenza rimane la medesima.

Esempio: efficacia di un farmaco

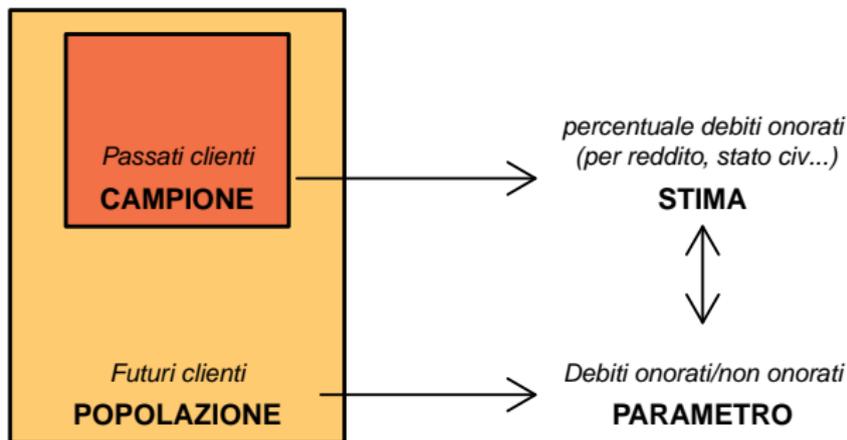
Una casa farmaceutica vuole verificare l'efficacia di un farmaco nell'abbassare la pressione.



- La popolazione è ideale.
- Potremmo voler tenere conto di fattori quali l'età o il genere che possono influenzare l'efficacia.

Esempio: concessione di credito

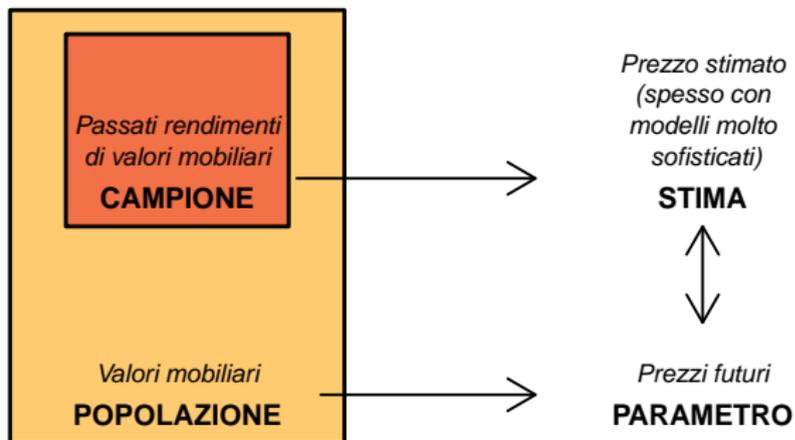
Una banca vuole valutare la probabilità che un potenziale cliente onori un debito sulla base di sue caratteristiche.



- Il campione non è un sottoinsieme della popolazione: assumiamo che i futuri clienti siano analoghi ai passati.
- Correzioni per fattori 'ambientali', ad es. il periodo di crisi.

Esempio: finanza

Consideriamo un fondo di investimento: qual è la probabilità che aumenti di valore il prossimo anno?



- Il futuro ripeterà il passato, o no?

Rappresentatività in generale

- Negli esempi il rapporto tra campione e popolazione non è quello 'ideale':
 - la popolazione può essere un'idealizzazione;
 - il campione è il passato, la popolazione d'interesse è il futuro.
- In molti casi stiamo ipotizzando vi sia rappresentatività, ipotesi che è critica per l'inferenza.
 - Farmaco:
 - popolazione ideale, il campione però ne è parte in un certo senso,
 - è ragionevole assumere la rappresentatività.
 - Concessione di crediti:
 - sappiamo che il campione è diverso dalla popolazione,
 - potremmo valutare degli aggiustamenti.
 - Finanza:
 - la rappresentatività è fortemente in dubbio,
 - aggiustamenti?

Stima e parametro

Abbiamo individuato i personaggi e la relazione tra alcuni di essi.



- Il punto è: perché sono legati stima e parametro?
- Ci aspettiamo che la stima sia vicina al valore del parametro, ma perché?
- Cerchiamo di capire meglio cos'è la stima.

Campioni e stime

- Consideriamo di nuovo l'esempio relativo alle elezioni, alla ricerca della percentuale di elettori del PD.
- Il problema se lo pongono in molti, e molti fanno dei sondaggi, ad esempio il 14 gennaio 2013 ne sono stati fatti 3.

committente	sondaggista	num.camp	PD
RAI Ballar	Ipsos	800	33.40
Sky	Tecn	600	31.80
Reuters Affaritaliani.it	Piepoli	1003	32.50

Campioni e stime

- Consideriamo di nuovo l'esempio relativo alle elezioni, alla ricerca della percentuale di elettori del PD.
- Il problema se lo pongono in molti, e molti fanno dei sondaggi, ad esempio il 14 gennaio 2013 ne sono stati fatti 3.

committente	sondaggista	num.camp	PD
RAI Ballar	Ipsos	800	33.40
Sky	Tecn	600	31.80
Reuters Affaritaliani.it	Piepoli	1003	32.50

- Nei tre sondaggi s'ottengono stime diverse, eppure sono contemporanei e popolazione e parametro sono gli stessi.

Campioni e stime

- Consideriamo di nuovo l'esempio relativo alle elezioni, alla ricerca della percentuale di elettori del PD.
- Il problema se lo pongono in molti, e molti fanno dei sondaggi, ad esempio il 14 gennaio 2013 ne sono stati fatti 3.

committente	sondaggista	num.camp	PD
RAI Ballar	Ipsos	800	33.40
Sky	Tecn	600	31.80
Reuters Affaritaliani.it	Piepoli	1003	32.50

- Nei tre sondaggi s'ottengono stime diverse, eppure sono contemporanei e popolazione e parametro sono gli stessi.
- Qualcuno si sbaglia?

Campioni e stime

- Consideriamo di nuovo l'esempio relativo alle elezioni, alla ricerca della percentuale di elettori del PD.
- Il problema se lo pongono in molti, e molti fanno dei sondaggi, ad esempio il 14 gennaio 2013 ne sono stati fatti 3.

committente	sondaggista	num.camp	PD
RAI Ballar	Ipsos	800	33.40
Sky	Tecn	600	31.80
Reuters Affaritaliani.it	Piepoli	1003	32.50

- Nei tre sondaggi s'ottengono stime diverse, eppure sono contemporanei e popolazione e parametro sono gli stessi.
- Qualcuno si sbaglia?
- In un certo senso tutti sbagliano.

Campioni e stime

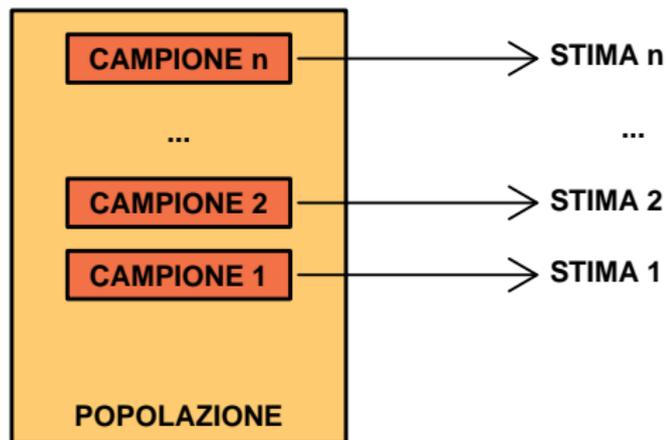
- Consideriamo di nuovo l'esempio relativo alle elezioni, alla ricerca della percentuale di elettori del PD.
- Il problema se lo pongono in molti, e molti fanno dei sondaggi, ad esempio il 14 gennaio 2013 ne sono stati fatti 3.

committente	sondaggista	num.camp	PD
RAI Ballar	Ipsos	800	33.40
Sky	Tecn	600	31.80
Reuters Affaritaliani.it	Piepoli	1003	32.50

- Nei tre sondaggi s'ottengono stime diverse, eppure sono contemporanei e popolazione e parametro sono gli stessi.
- Qualcuno si sbaglia?
- In un certo senso tutti sbagliano.
- Il fatto è che sono intervistate persone diverse, e quindi il risultato è casuale.

Stimatore, stima, distribuzione campionaria

- Una stima è il risultato di un esperimento casuale:
 - estraggo **a caso** il campione,
 - applico lo stimatore (ad es. la proporzione campionaria),
 - ottengo la stima (valore assunto dallo stimatore),
- **se estraessi un altro campione otterrei una diversa stima!**



- Quindi otteniamo una risposta a caso?

Stimatore, stima, distribuzione campionaria

- Quindi otteniamo una risposta a caso?
- La risposta è sì, la risposta che ottengo è casuale.
- Questo però non significa che non sia informativa.
- La stima è sì casuale, ma è probabile che sia vicina al parametro.
- (Sempre se sto facendo tutto bene.)
- Per capire cosa significa questo, immagineremo di estrarre tanti campioni, cioè di ripetere più volte la procedura.

Sommario

- 1 Concetti alla base, popolazione, parametro, campione, stima
- 2 La distribuzione campionaria**
 - Incertezza della stima
- 3 Statistica in guerra
- 4 Riepilogo

Stipendio dei laureati

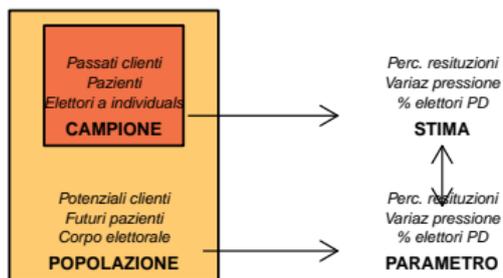
- Una (ex-)facoltà universitaria vuole valutare lo stipendio medio dei suoi laureati a 10 anni dalla laurea.
- La lista dei laureati di dieci anni prima include 1000 individui (**popolazione**).
- È (naturalmente) visto come troppo dispendioso rintracciarli tutti e chiedere loro il reddito.
- Teoricamente è comunque fattibile, il valore medio sui 1000 è il nostro **parametro**.
- Si selezionano allora a caso 20 laureati tra i 1000 (**campione**) e gli si chiede il reddito (per telefono o simili).
- La media dei 20 redditi è la nostra **stima**.

Principio del campionamento ripetuto

Per capire come (e perché) funzioni occorre usare un po' di immaginazione:

- facciamo finta di ripetere il campionamento (cioè di ottenere diversi campioni) più e più volte,
- per ogni campione otteniamo una diversa stima (media campionaria).
- Qual è il legame tra queste stime e il parametro?

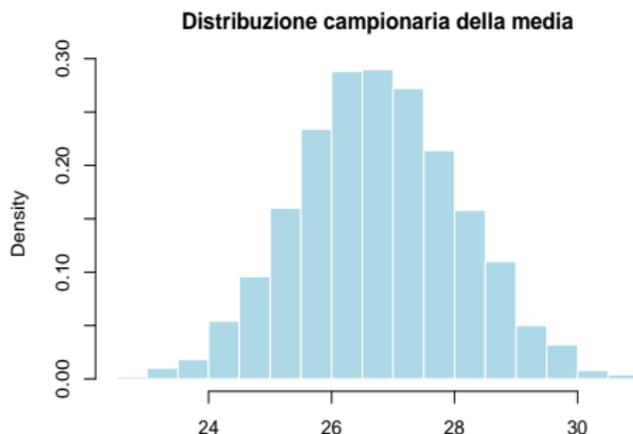
Simuleremo quello che succederebbe con una popolazione nota.



Distribuzione campionaria

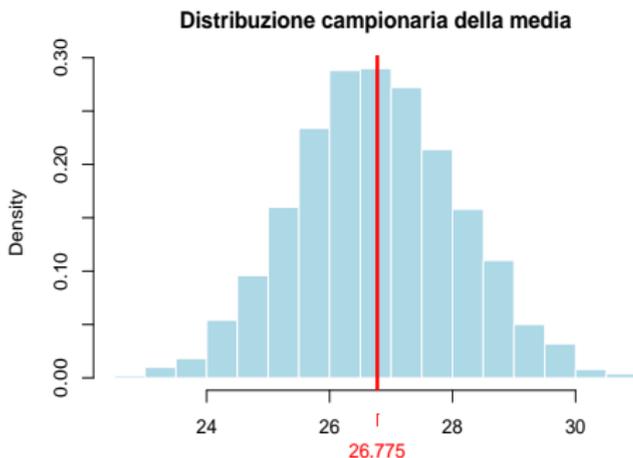
Distribuzione campionaria

Le 1000 medie derivate dai nostri 1000 campioni si distribuiscono così



Distribuzione campionaria

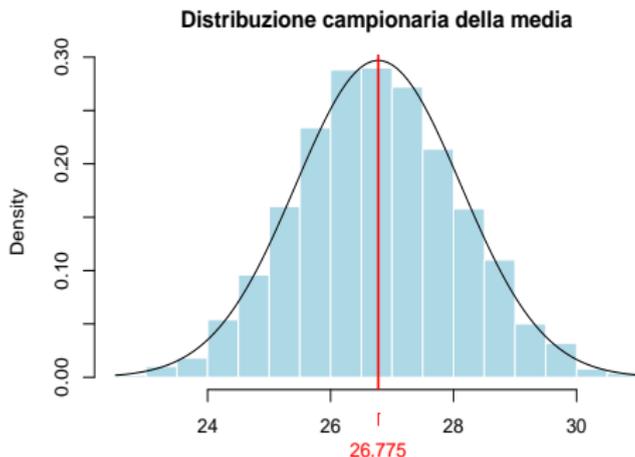
Le 1000 medie derivate dai nostri 1000 campioni si distribuiscono così



- Intorno al vero valore del parametro, nel senso che sono più probabili valori vicini alla media della popolazione, 26.78, che valori lontani.

Distribuzione campionaria

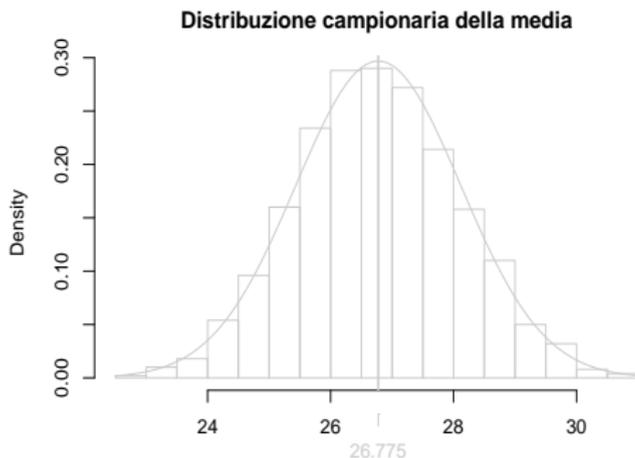
Le 1000 medie derivate dai nostri 1000 campioni si distribuiscono così



- Intorno al vero valore del parametro, nel senso che sono più probabili valori vicini alla media della popolazione, 26.78, che valori lontani.
- Semplifichiamo la rappresentazione, immaginiamo di avere un'infinità di campioni e passiamo alla funzione di densità.

Distribuzione campionaria

Le 1000 medie derivate dai nostri 1000 campioni si distribuiscono così



Nella realtà però io

- non conosco la media della popolazione,
- non osservo tanti campioni ma uno.

Distribuzione approssimata della media

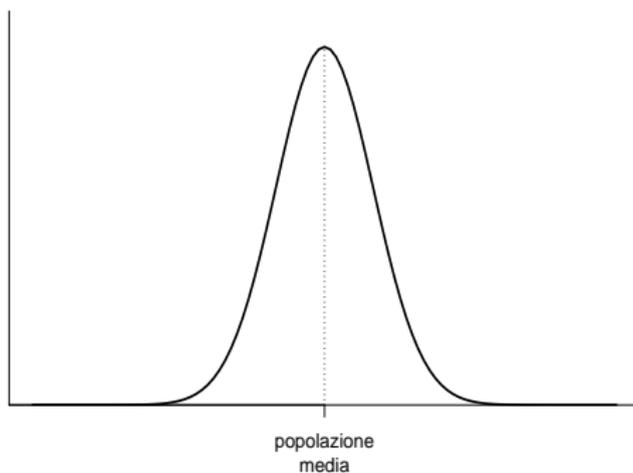
Un risultato fondamentale è il seguente teorema che ci dice che, approssimativamente, la media campionaria è sempre normale, con media pari alla media della popolazione e varianza che dipende dalla varianza della popolazione e dalla dimensione del campione.

Teorema del limite centrale

Siano X_1, \dots, X_n con media μ e varianza σ^2 indipendenti, allora, approssimativamente

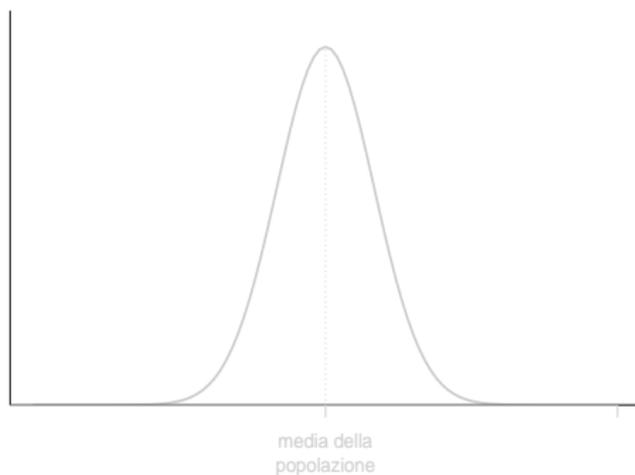
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuzione campionaria



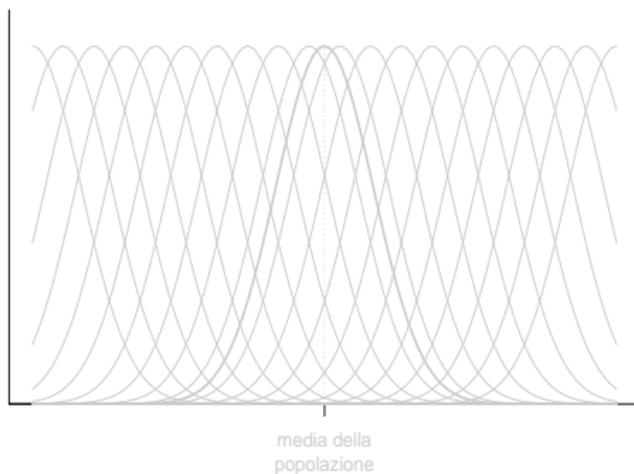
- Sappiamo com'è fatta la distribuzione.

Distribuzione campionaria



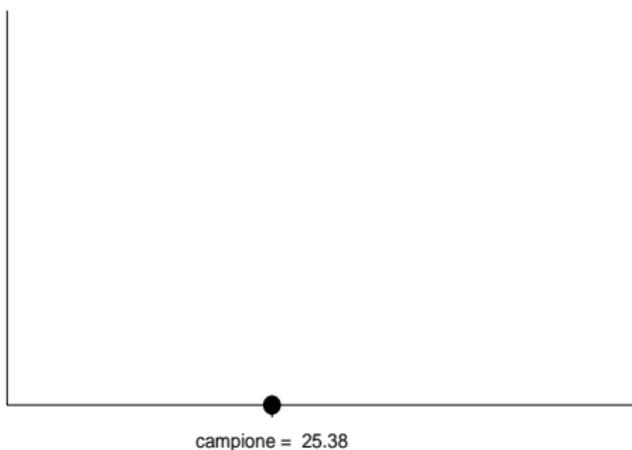
- Sappiamo com'è fatta la distribuzione.
- Non sappiamo dov'è.

Distribuzione campionaria



- Sappiamo com'è fatta la distribuzione.
- Non sappiamo dov'è.
- Sappiamo dove potrebbe essere.

Distribuzione campionaria

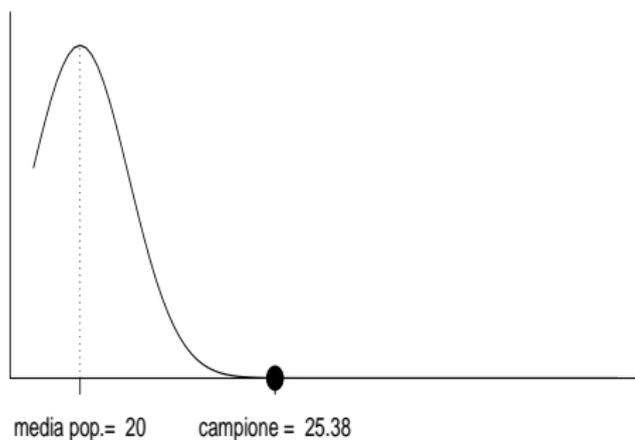


- abbiamo però **un** campione e la relativa media.
- Confrontiamo il campione con le possibili distribuzioni della media campionaria.

Distribuzione campionaria

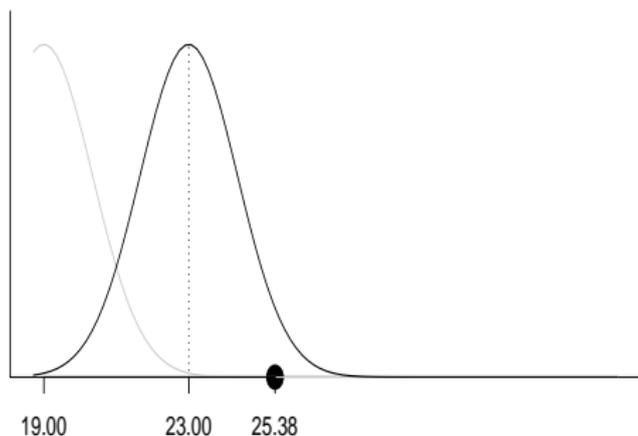
- Il valore più verosimile per la media della popolazione è quindi uguale alla media campionaria, da cui uso quella come stima.
- Questo tipo di ragionamento permette anche di definire intervalli di valori calcolati sulla base del campione che con una certa probabilità contengono il parametro.

Distribuzione campionaria



- Ad esempio, se la media della popolazione fosse 19, sarebbe improbabile osservare una media del campione pari a 25.38.

Distribuzione campionaria



- Ad esempio, se la media della popolazione fosse 19, sarebbe improbabile osservare una media del campione pari a 25.38.
- Mentre se la media della popolazione fosse 23, osservare una media campionaria pari a 25.38 sarebbe molto più verosimile.

Distribuzione campionaria

- Il valore più verosimile è uguale alla media campionaria.

Stima, incertezza, distribuzione campionaria

- Una stima è una quantità che possiamo calcolare a partire dal campione.
- È legata al parametro, però è \neq dal parametro quasi certamente.
- Le procedure statistiche hanno sempre un margine di errore, la stima (puntuale) non fa eccezione.
- Abbiamo però più della stima! Abbiamo la distribuzione campionaria:
- quest'ultima può essere usata per stabilire quanto lontano potremmo essere dal vero valore del parametro.
- (Uno stimatore è tanto più buono quanto più la sua distribuzione è concentrata intorno al vero valore del parametro.)
- Come misurare la concentrazione o come misurare il margine di errore è tema che non affrontiamo nel dettaglio, la varianza della distribuzione campionaria può essere una soluzione, spesso valida.

Numerosità campionaria e precisione dello stimatore

- Nella cittadina di Dunwich, che conta 100 000 abitanti, si terranno a breve le elezioni comunali.
- Ci sono due soli candidati, Arthur Dupin (attuale sindaco) e Auguste Pym.
- L'attuale sindaco Arthur Dupin commissiona un sondaggio: 500 persone vengono intervistate (e tutti rispondono, è una storia di fantasia), il 50% dichiara che voterà per lui.
- Lo sfidante d'altro canto commissiona un suo sondaggio, sulla base di 1000 rispondenti (anche qui rispondono tutti), si stima al 45% la proporzione di votanti per l'attuale sindaco.

che due campioni portino a stime diverse ormai non ci stupisce, ma la domanda è

Quale delle due stime è da ritenersi maggiormente affidabile?

Numerosità campionaria e precisione dello stimatore

- Auguste Pym che, ricordiamolo, è dato favorito come candidato sindaco di Dunwich (100 000 abitanti) secondo un sondaggio che ha coinvolto 1000 persone, si vanta con un amico, William Wilson, anche lui politico, del fatto appunto di essere il favorito nella sua competizione elettorale.
- L'amico William è anch'egli candidato sindaco, ma nella città di Innsmouth (1 000 000 di abitanti) e anche lui è dato al 55% in un sondaggio effettuato intervistando 1000 persone (tutte rispondenti) e dice all'amico 'Ehi, anch'io sono favorito come te a Innsmouth.'
- Auguste replica però: 'Sì, però il mio sondaggio ha intervistato 1000 persone su 100 000 (l'1 per cento), il tuo 1000 su 1 000 000 (l'1 per mille), io sono molto più sicuro di te.'

Auguste ha ragione?

Campionare è come assaggiare una zuppa durante la cottura

- Quando osserviamo un campione osserviamo una parte della popolazione e ne inferiamo le caratteristiche dell'intera popolazione.
- Quando assaggiamo la zuppa per vedere se è giusta di sale ne mangiamo un cucchiaino, verificiamo che quel cucchiaino sia correttamente salato.
- Se la zuppa è stata adeguatamente mescolata, riteniamo che l'intera pentola abbia lo stesso gusto del (piccolo) cucchiaino assaggiato.
- Cambia qualcosa che la pentola sia piccola o grande?

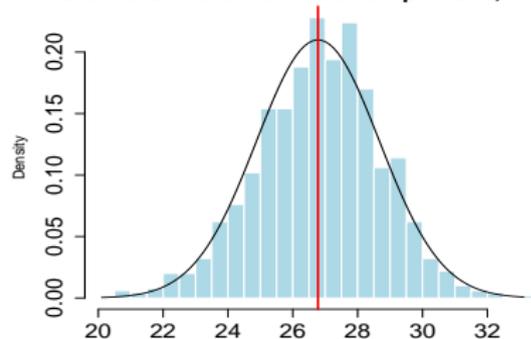


Cosa significa

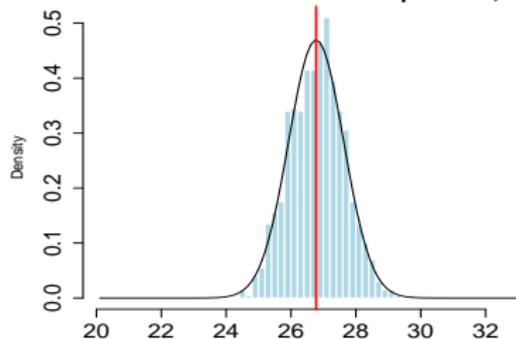
- Che se anche ho una popolazione molto grande, non mi occorre un campione molto grande (un campione di 1000 va egualmente bene per la popolazione della Lombardia, dell'Italia o dell'UE)
- D'altra parte, anche se la popolazione è piccola non è che mi basta un campione più piccolo (un campione di 1000 ha la stessa affidabilità che la popolazione di riferimento sia quella UE, quella italiana o quella di Trieste).
- Attenzione che tutto questo è vero se comunque la popolazione è grande rispetto al campione (un campione di 1000 abitanti di Opicina ha un valore diverso, ovviamente).

Cosa succede se aumenta la numerosità del campione n

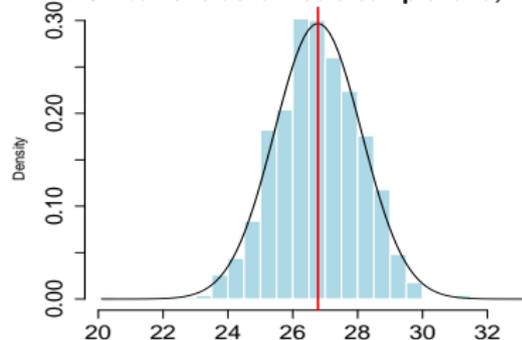
Distribuzione della media campionaria, $n=10$



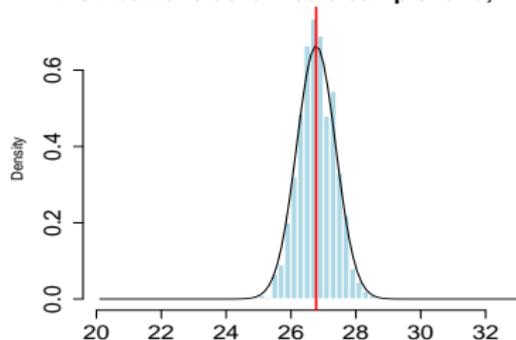
Distribuzione della media campionaria, $n=50$



Distribuzione della media campionaria, $n=20$

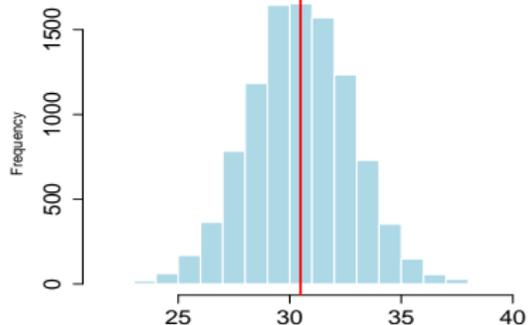


Distribuzione della media campionaria, $n=100$

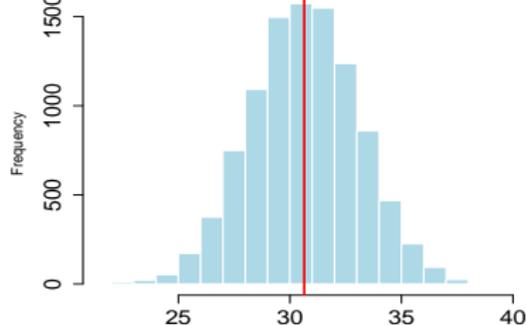


Cosa succede se aumenta la dimensione della popolazione N

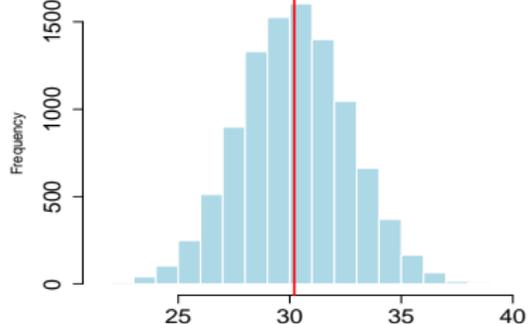
Distribuzione della media campionaria, $N=1000$



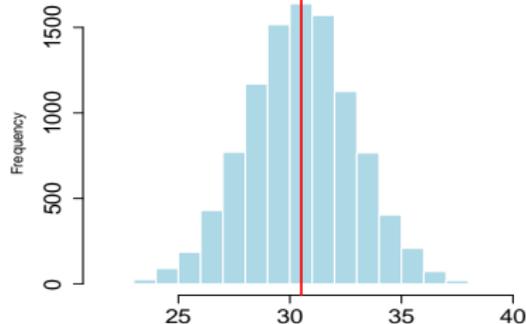
Distribuzione della media campionaria, $N=10000$



Distribuzione della media campionaria, $N=5000$



Distribuzione della media campionaria, $N=50000$



Stimatore e distribuzione campionaria

- Lo stimatore è una quantità che calcolo a partire dal campione.
- Esso è buono se la sua distribuzione campionaria si concentra intorno al vero valore del parametro.
- Quanto più è concentrata tanto meglio.
- La distribuzione campionaria servirà a stabilire se non quanto vale il parametro quanto, probabilmente, siamo vicini.

Sommario

- 1 Concetti alla base, popolazione, parametro, campione, stima
- 2 La distribuzione campionaria
- 3 Statistica in guerra**
- 4 Riepilogo

Le forze del nemico

- Siamo nel 1942, in piena II guerra mondiale.
- Uno degli elementi più importanti della strategia militare nella IIGM è costituito dai carri armati.
- Ovviamente, sapere di quanti carri armati disponesse il nemico, e quanti ne producesse, era considerato di primaria importanza.



Quanti carri armati?

- Mettiamoci nei panni degli alleati.
- Vorremmo stabilire quanti carri armati hanno a disposizione i tedeschi e quanti ne producono.
- Come?
 - ricorrendo alle spie (ad esempio, qualcuno con accesso agli archivi del ministero degli armamenti tedesco);
 - ricorrendo a informazioni prebelliche sulle capacità dell'industria tedesca;
 - C'è però anche un'opzione 'statistica'.



I numeri di serie!

- I carri armati sono numerati.
- Si sa che i carri sono numerati in sequenza.
- Si conoscono i numeri di serie di alcuni carri catturati.

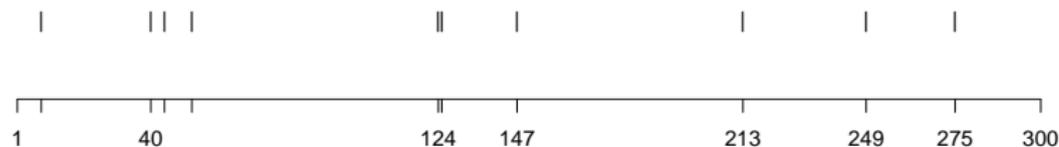


I numeri di serie!

- **Popolazione:** C carri, numerati da 1 a C .
- **Parametro:** C .
- **Campione:** n carri catturati, e i loro numeri, siano
147, 44, 52, 125, 124, 275, 40,
8, 213, 249
- Come stimiamo C ?



I numeri di serie!



Alcune statistiche:

- Media: 127.7;
- Mediana: 124.5;
- Deviazione Standard: 93.5925567;
- Minimo: 8;
- Massimo: 275

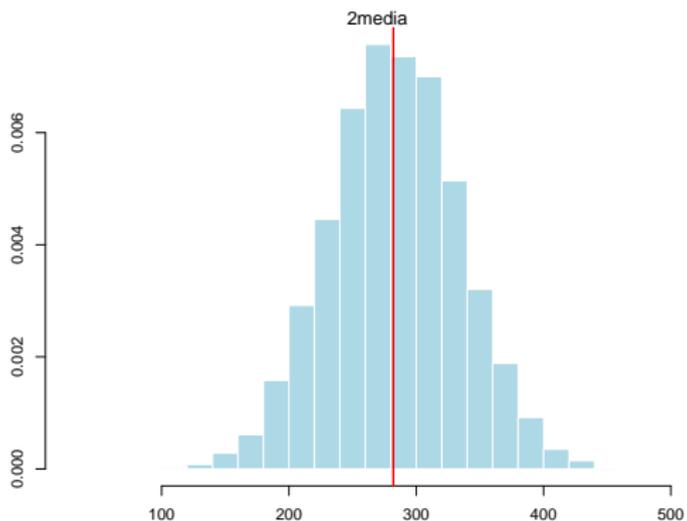
Stimatori proponibili

- $2 \times \text{media} = 255.4$
- $2 \times \text{mediana} = 249$
- $\text{media} + 2 \times \text{Dev. Standard} = 314.8851134$
- $\text{massimo} = 275$
- $\text{massimo} + \text{minimo} = 283$

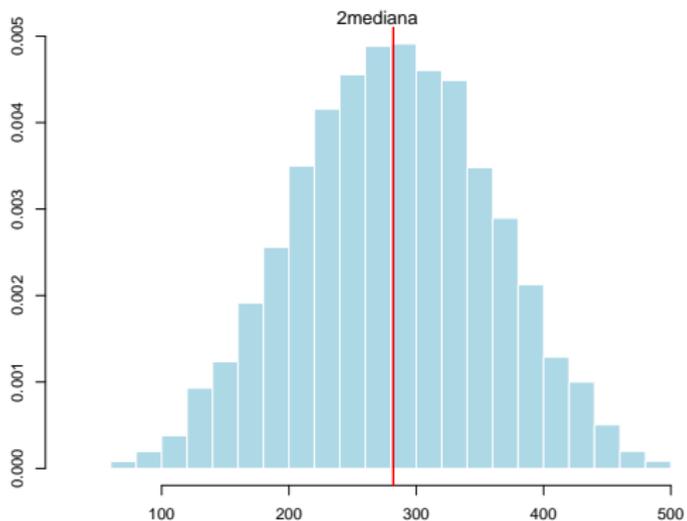
Come facciamo a decidere qual è meglio?

Se ci riferiamo a questo solo campione è impossibile, se ci riferiamo alla distribuzione campionaria invece è possibile.

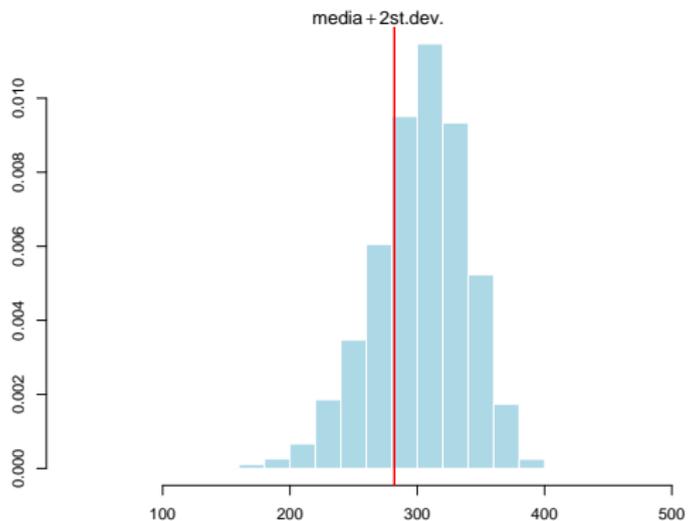
Distribuzione campionaria, $C = 282$; I



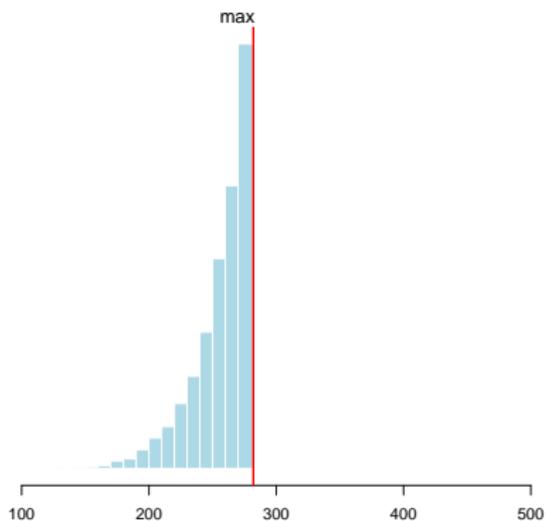
Distribuzione campionaria, $C = 282$; II



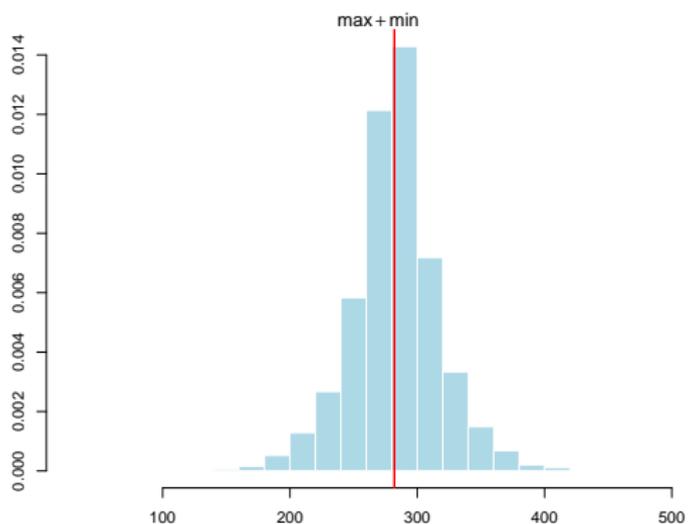
Distribuzione campionaria, $C = 282$; III



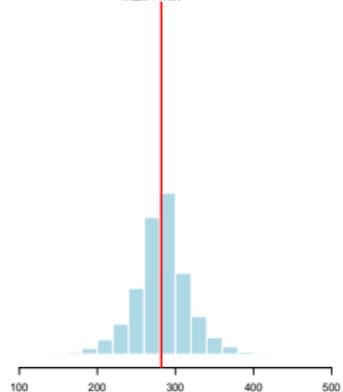
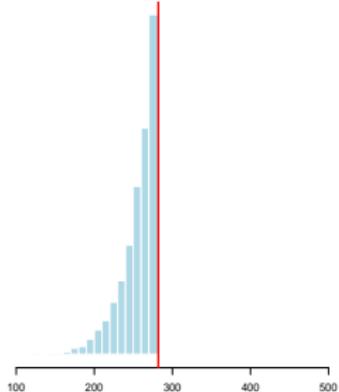
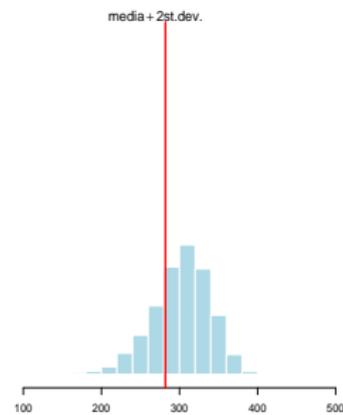
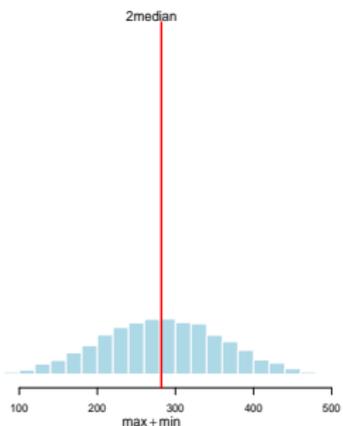
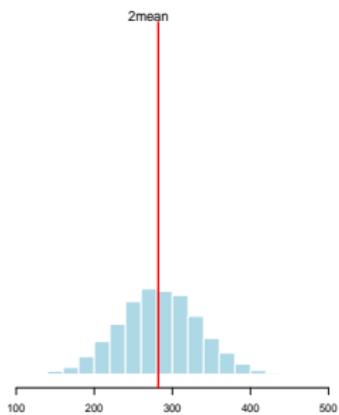
Distribuzione campionaria, $C = 282$; IV



Distribuzione campionaria, $C = 282$; V



Distribuzione campionaria, $C = 282$



La realtà

La situazione reale era più complicata

- i numeri di serie sono più complessi (contengono di solito anno e mese di produzione, produttore, impianto);
- si possono usare i numeri di serie di varie componenti (ad esempio per i carri armati risultava molto utile quello della scatola del cambio).

Quindi in realtà

- Si sono usati metodi più articolati per tenere conto di tutte le informazioni;
- Non si otteneva solo un numero totale, ma maggiore dettaglio, ad esempio la produzione
 - per i vari impianti (utile per scegliere dove bombardare);
 - per diversi periodi (utile per vedere, ad es. se dopo un bombardamento la produzione diminuiva e quindi misurarne il successo).

La metodologia alla base è comunque quella illustrata.

Com'è andata?

Dopo la fine della guerra, si sono potute confrontare le varie previsioni con i dati sulla produzione

Date	Estimated Monthly Production		Monthly Production Speer Ministry
	Serial Number Estimate	Munitions Record 10 Aug. 42	
June, 1940	169	1000	122
June, 1941	244	1550	271
August, 1942	327	1550	342

Sommario

- 1 Concetti alla base, popolazione, parametro, campione, stima
- 2 La distribuzione campionaria
- 3 Statistica in guerra
- 4 Riepilogo**

Riepilogo

- Popolazione: collettivo su cui vogliamo ricavare informazioni;
- Parametro: l'informazione di interesse, è una caratteristica della popolazione;
- Campione: parte della popolazione che osservo
 - dev'essere rappresentativo;
 - selezione casuale;
- Stimatore: corrispondente del parametro
 - calcolato dal campione;
 - ha una distribuzione di probabilità (campionaria)
 - concentrata intorno al valore del parametro;
 - tanto più concentrata quanto più numeroso è il campione.