

INTRODUZIONE ALL'INFERENZA STATISTICA

Prof. N. Ribeco
nunziata.ribecco@uniba.it

PNLS - Statistica

Bari, 15 e 22 novembre 2016



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

DIPARTIMENTO DI
SCIENZE ECONOMICHE
E METODI MATEMATICI

Un mondo pieno di incertezza ...

- ❑ La capacità di prendere decisioni in situazioni di incertezza è indispensabile al cittadino consapevole che vuole capire e controllare i fenomeni (naturali, sociali, economici e politici) che lo circondano.

- ❑ Alcuni esempi:
 - ✓ efficacia di un vaccino
 - ✓ gestione del mutuo per l'acquisto di una casa
 - ✓ andamento dei prezzi
 - ✓ analisi della soddisfazione verso un servizio pubblico
 - ✓ giochi d'azzardo (lotterie, gratta e vinci, ecc..)
 - ✓ scelta del percorso universitario



Da cosa deriva l'incertezza?

- ❑ L'incertezza è generata da meccanismi molto diversi ...
- ❑ ma fondamentalmente dipende dalla **VARIABILITÀ**
 - ✓ se lancio un dado o una moneta, o se estraggo un numero al lotto non so quale sarà il risultato;
 - ✓ se misuro più volte un oggetto le misure non sono tutte uguali (variabilità delle misure, o errore di misura, p.e. due misurazioni della propria statura);
 - ✓ se misuro l'altezza di due studenti ottengo valori diversi (variabilità 'biologica');
 - ✓ non so se domani pioverà, ecc

Prendere decisioni in situazioni di incertezza

- ❑ In tutti i momenti della vita quotidiana dobbiamo prendere delle decisioni.
- ❑ Come decidere, sicuri di non sbagliare?
- ❑ Non è possibile essere **certi** di non sbagliare ma, certamente, è possibile controllare il rischio di un'errata decisione.
- ❑ Ruolo della Statistica

Ruolo della Statistica

- ❑ La Statistica è lo strumento idoneo per prendere decisioni in situazioni di incertezza
- ❑ La Statistica è una scienza quantitativa, ma il modo di pensare 'statistico' è diverso da quello 'matematico' per almeno 2 aspetti:
 - la statistica non può prescindere dal contesto (dati)
 - la statistica permette di prendere decisione nelle situazioni di incertezza non basandosi sulla **deduzione** (come la matematica) ma sull'**induzione**: dal particolare (ciò che si è osservato) al generale.

Procedimento induttivo: Inferenza statistica

- Il procedimento che, partendo dal particolare permette di risalire al generale è il ***procedimento induttivo***
- L'insieme di metodi statistici che, partendo dall'osservazione e dall'analisi del particolare, permettono di estendere la decisione al contesto generale determinano ***l'induzione statistica***



INFERENZA STATISTICA

... Inferenza statistica

- L'inferenza statistica è un procedimento logico ed operativo che si avvale di metodi che portano a dei risultati che vengono ottenuti su una parte dell' intero universo e poi estesi all'universo nel complesso:

Universo  **Popolazione**

Parte  **Campione**

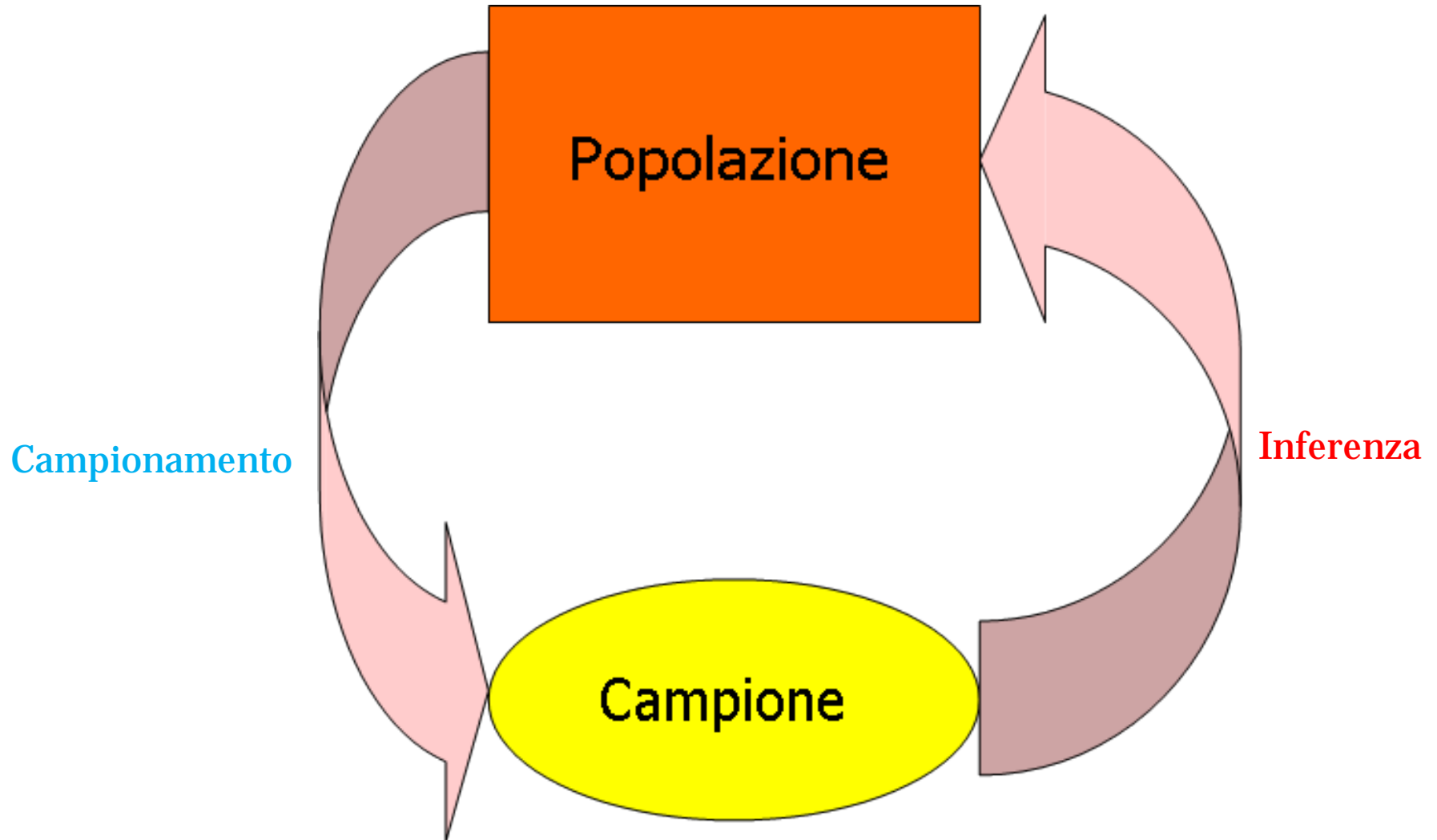
... Inferenza statistica

- ❑ Il percorso logico dell'inferenza si svolge secondo le seguenti fasi:
 - Estrazione di una parte (campione) della popolazione.
 - Calcolo sui dati campionari di quantità (stime) quali ad esempio: media, varianza o altre.
 - Estensione alla popolazione dei risultati forniti dal campione (inferenza).

Popolazione e campione

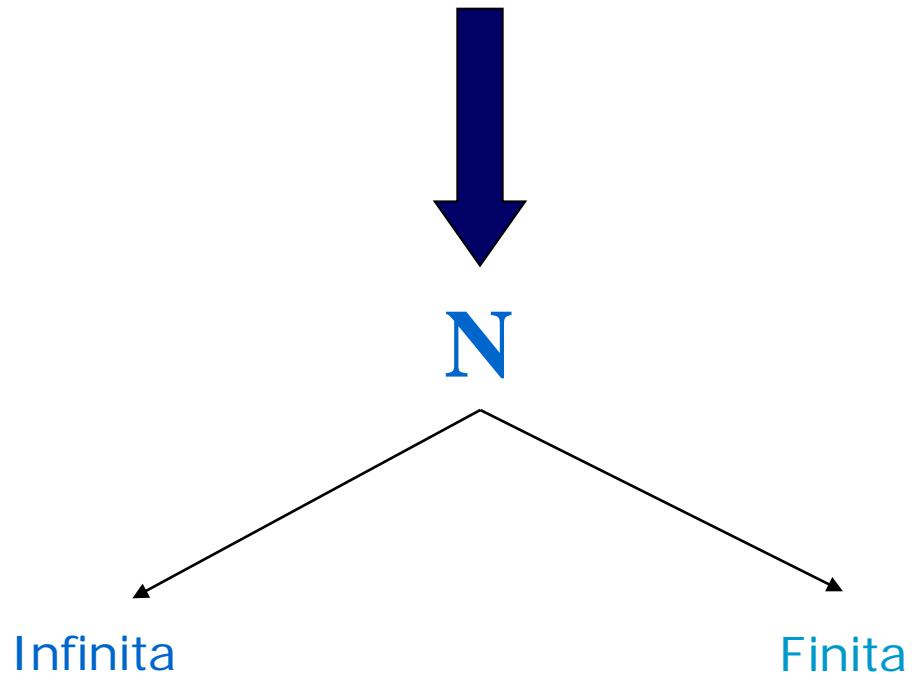
- ❑ **Popolazione**: insieme di tutte le osservazioni possibili, relativamente ad una certa variabile o ad un dato fenomeno. Può essere finita (comunque molto grande) o infinita e non può essere conosciuta in maniera esaustiva. Anche quando è finita, la **popolazione**, per problemi di costo e di tempo, non può essere esplorata nella sua interezza pertanto si considera una sua parte (**campione**)
- ❑ **Campione**: insieme finito di elementi estratti da una popolazione. Deve essere il più possibile rappresentativo della popolazione (molto simile alla popolazione rispetto alle caratteristiche ritenute salienti per la ricerca), al fine di generalizzare i risultati ottenuti nel campione all'intera popolazione di riferimento (**inferenza**).

Introduzione all'inferenza statistica



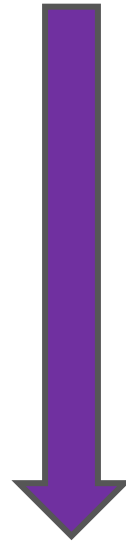
Dimensione popolazione

Dimensione popolazione



Dimensione Campione

Dimensione campione



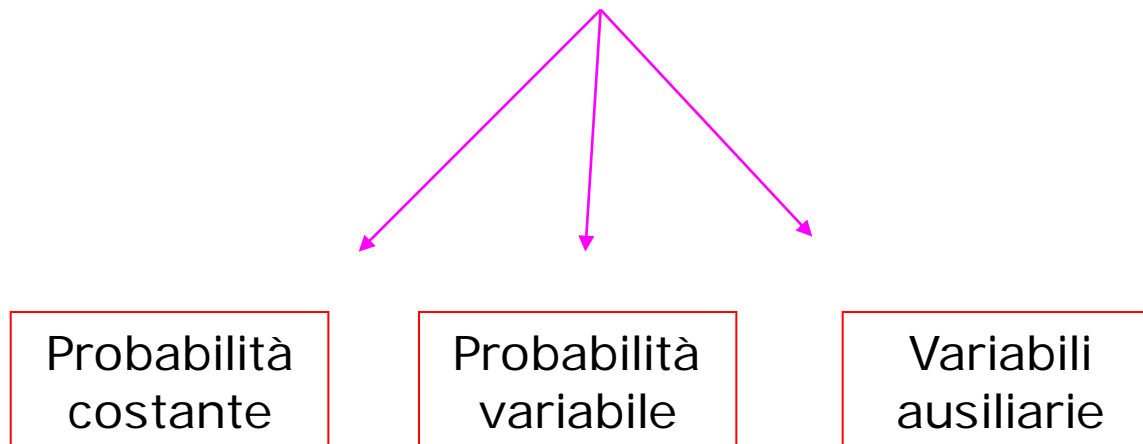
n

Campione

- Campione rappresentativo della popolazione da cui viene estratto \longrightarrow campionamento casuale.



Campionamento probabilistico



Campione

Tecniche probabilistiche:

1. Campionamento casuale semplice;
2. Campionamento sistematico;
3. Campionamento stratificato;
4. Campionamento a grappoli.

Campione

□ **Tecniche non probabilistiche:**

1. per popolazioni rare ed elusive;
2. per quote;
3. a scelta ragionata.

Campionamento casuale semplice

□ Campionamento con ripetizione (o bernoulliano):

Le n unità vengono estratte e rimesse nell'urna. La composizione dell'urna non cambia e la probabilità per ogni unità di essere estratta è costante ($1/N$) e le unità estratte sono fra loro indipendenti.

□ Campionamento senza ripetizione (o esaustivo):

Le n unità vengono estratte e **non reimmesse** nell'urna. La composizione dell'urna cambia ad ogni estrazione e cambia, anche, la probabilità per ogni unità di essere estratta ($1/N$ per la prima estrazione, $1/N-1$ per la seconda, $1/N-2$ per la terza e così via).

Inferenza statistica

- ❑ Le caratteristiche della popolazione a cui siamo interessati prendono il nome di: **parametri**.
- ❑ Sono parametri: la media, la varianza, ecc.
- ❑ Il campionamento che viene utilizzato nelle tecniche inferenziali classiche è il **campionamento con ripetizione**

Inferenza statistica e Calcolo delle probabilità

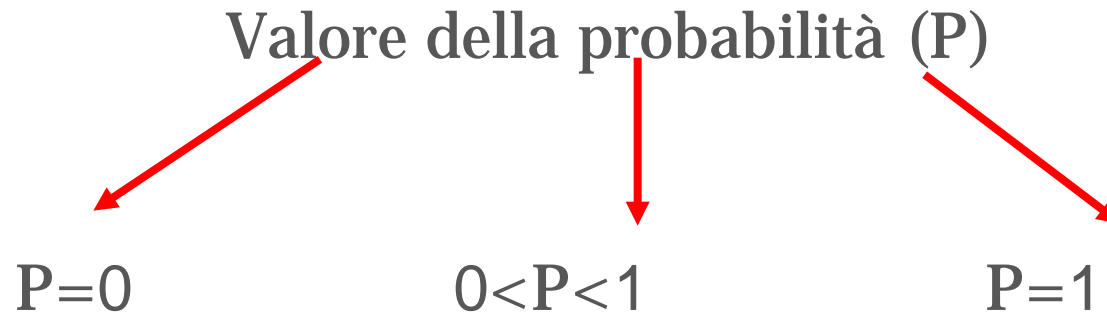
- ❑ Le tecniche di inferenza statistica si avvalgono del calcolo delle probabilità.
- ❑ L'estensione dei risultati campionari viene effettuata controllando (in termini probabilistici) il rischio di aver fatto una valutazione errata

Definizione di probabilità

- ❑ La maggior parte dei fenomeni, ai quali assistiamo quotidianamente, può manifestarsi in vari modi, ma è quasi sempre impossibile stabilire a priori quale di essi si presenterà ogni volta.
- ❑ La **PROBABILITA'** è un numero che si associa ad un evento **E** ed esprime il grado di aspettativa circa il suo verificarsi.
- ❑ Se indichiamo con **P(E)** la probabilità che si verifichi l'evento **E** questo numero può assumere valori fra:

$$0 \leq P(E) \leq 1$$

Valore della probabilità



- ❖ La probabilità di un evento casuale qualsiasi è sempre un numero compreso fra **0** e **1**
- ❖ un **evento casuale certo** ha probabilità **1**
- ❖ un **evento casuale impossibile** ha probabilità **0**



Evento impossibile, certo, possibile!

- **evento impossibile**: estrarre il numero 100 dalla tombola. **PROBABILITA'** sempre uguale a 0 (**$P=0$**)



- **evento certo**: l'uscita di un numero compreso tra 1 e 6 nel lancio di un dado. **PROBABILITA'** sempre uguale a 1 (**$P=1$**)



- **evento possibile**: estrarre una pallina bianca da una scatola contenente sia palline bianche che rosse. **PROBABILITA'** compresa tra 0 e 1 (**$0 < P < 1$**)



Come calcoliamo la probabilità

numero dei **cas**i favorevoli

diviso

numero dei **cas**i possibili



Come calcoliamo la probabilità?



Lanciamo una moneta

- ✓ Qual è la probabilità che esca «croce»?
- ✓ I casi possibili sono 2 «testa» o «croce»
- ✓ Il caso favorevole è 1 cioè «croce»

La probabilità che esca «croce» sarà 1 su 2, ossia $\frac{1}{2}$



Come calcoliamo la probabilità?

- ❑ Lanciamo un dado
- ✓ Qual è la probabilità che esca un «numero pari»?
- ✓ I casi possibili sono 6, tanti quanti sono i numeri riportati sulle facce del dado e che possono risultare dal lancio.
- ✓ I casi favorevoli sono 3: **i numeri 2, 4 e 6**
- ❑ La probabilità che esca un numero pari è **3 su 6**:
 $3/6 = 1/2$.

Legge dei grandi numeri

- Il *numero di volte* che un evento E si è verificato, durante un numero di n prove effettuate, si chiama **frequenza assoluta** dell'evento
- Il *rapporto* fra la frequenza assoluta dell'evento E e il numero n di prove effettuate si chiama **frequenza relativa** dell'evento E ($F(E)$)
- **Per la legge dei grandi numeri:**

$$P(E) \approx F(E) = \frac{\text{frequenza assoluta}}{n}$$

Dalla probabilità alla frequenza relativa

- ❑ Se lancio un dado 20 volte e per 12 volte si presenta una faccia pari:

Facce del dado	Numero di volte
PARI	12
DISPARI	8
TOTALE	20

- ❑ La probabilità della faccia pari sarà:

$$P(E) = \frac{\text{numero di volte}}{\text{numero di lanci}} = \frac{\text{frequenza assoluta}}{\text{numero totale}} = \frac{12}{20} = 0.6 = 60\%.$$



La distribuzione normale o di Gauss

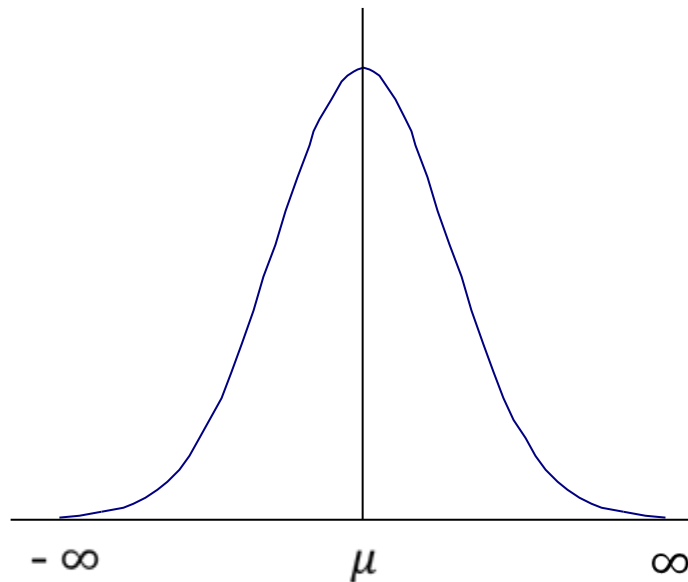
- ❑ Le tecniche inferenziali sono un insieme di metodi sviluppati ponendo come ipotesi fondamentale che la popolazione da cui viene estratto il campione sia normale
- ❑ Moltissime variabili tendono a distribuirsi *normalmente*.
- ❑ La distribuzione normale è una distribuzione di probabilità.
- ❑ Si applica bene alle quantità campionarie.
La formula della distribuzione normale è definita dai parametri media (μ) e scarto quadratico medio (σ).

Introduzione all'inferenza statistica

La distribuzione normale o di Gauss

□ L'espressione di questa distribuzione è:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$



Introduzione all'inferenza statistica

La distribuzione normale o di Gauss: caratteristiche

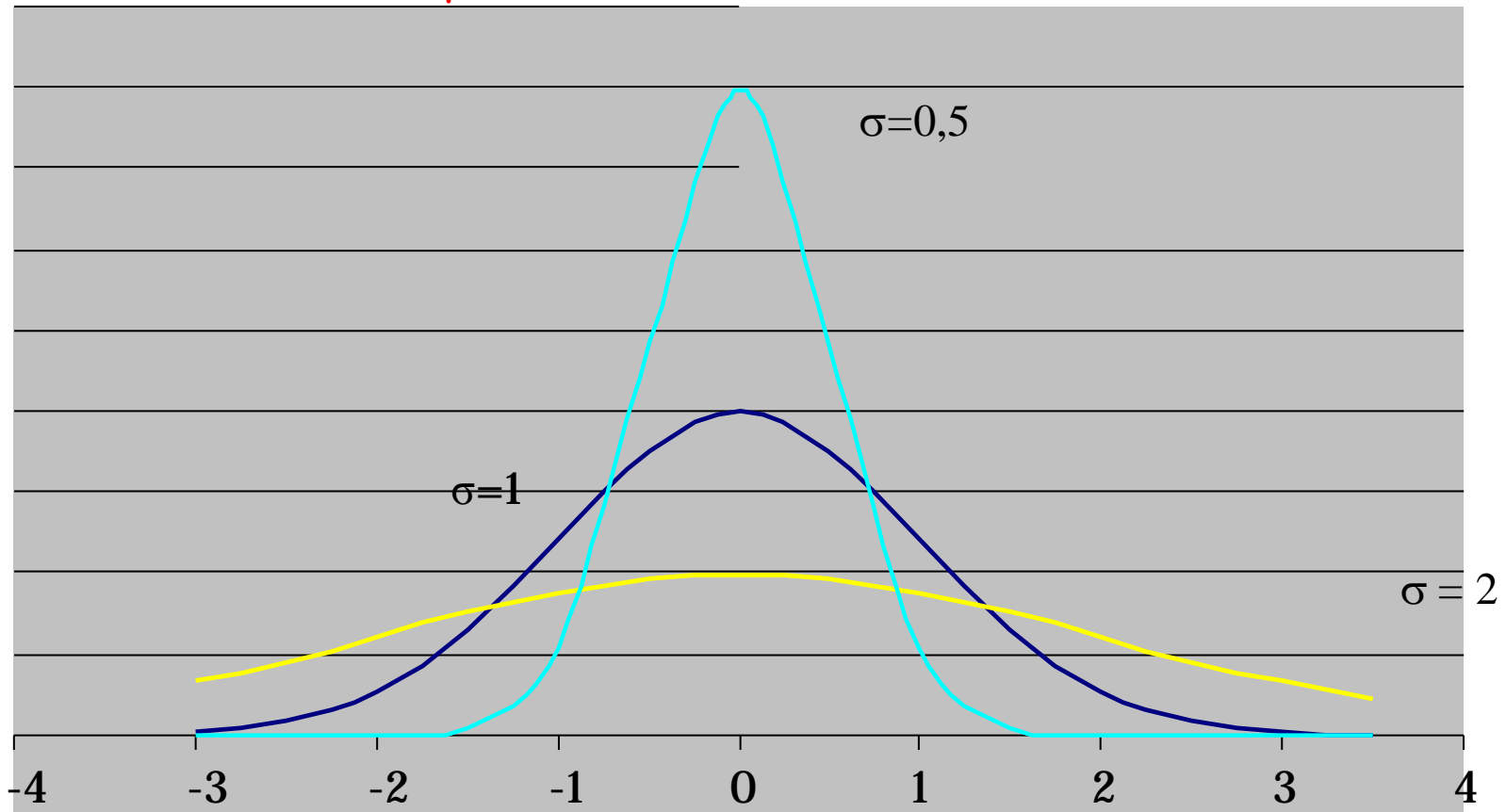
- ❑ Forma campanulare.
- ❑ Asintotica rispetto all'asse delle ascisse.
- ❑ Simmetrica intorno alla media:

moda \equiv **mediana** \equiv **media aritmetica**.

- ❑ Due punti di flesso in corrispondenza di $\pm \sigma$.
- ❑ L'ordinata del punto di massimo è: $\frac{1}{\sigma\sqrt{2\pi}}$.

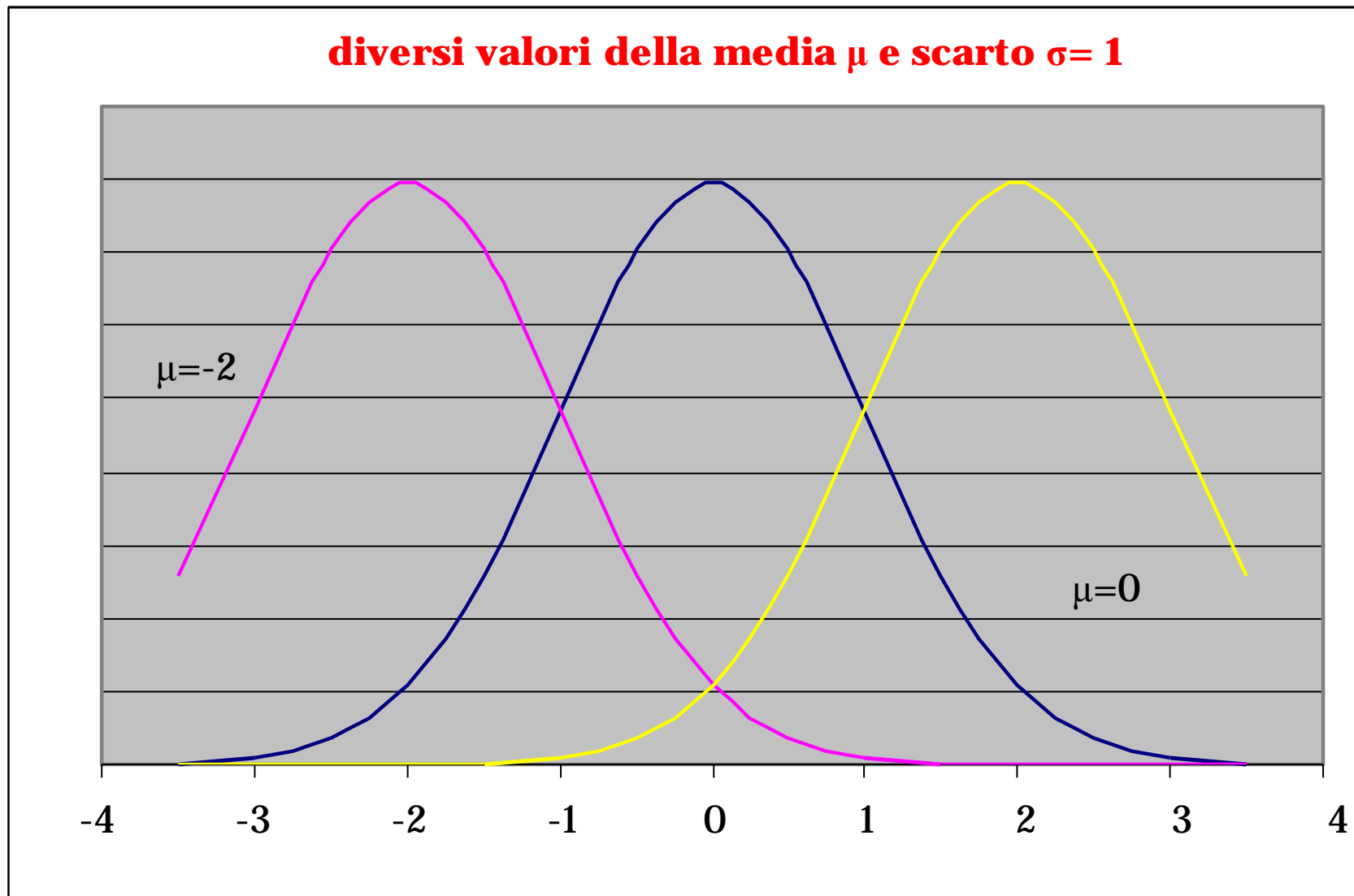
Alcuni esempi di distribuzione normale

media $\mu=0$ e diversi valori dello scarto σ



Alcuni esempi di distribuzione normale

diversi valori della media μ e scarto $\sigma=1$



Normale standardizzata

- Abbiamo visto che possiamo avere infinite variabili normali. Pertanto, al fine di poterle studiare possiamo ricondurle ad una variabile che è sempre distribuita normalmente ma che, grazie ad una trasformazione, ha la caratteristica di avere *media 0 e varianza (o scarto quadratico medio) 1*.
- Tale variabile prende il nome di *Normale standardizzata*.

Introduzione all'inferenza statistica

Normale standardizzata

□ La trasformazione è la seguente:

$$Z = \frac{x - \mu}{\sigma}$$

dove:

x: valore a cui siamo interessati

μ: media nella popolazione

σ: deviazione standard nella popolazione

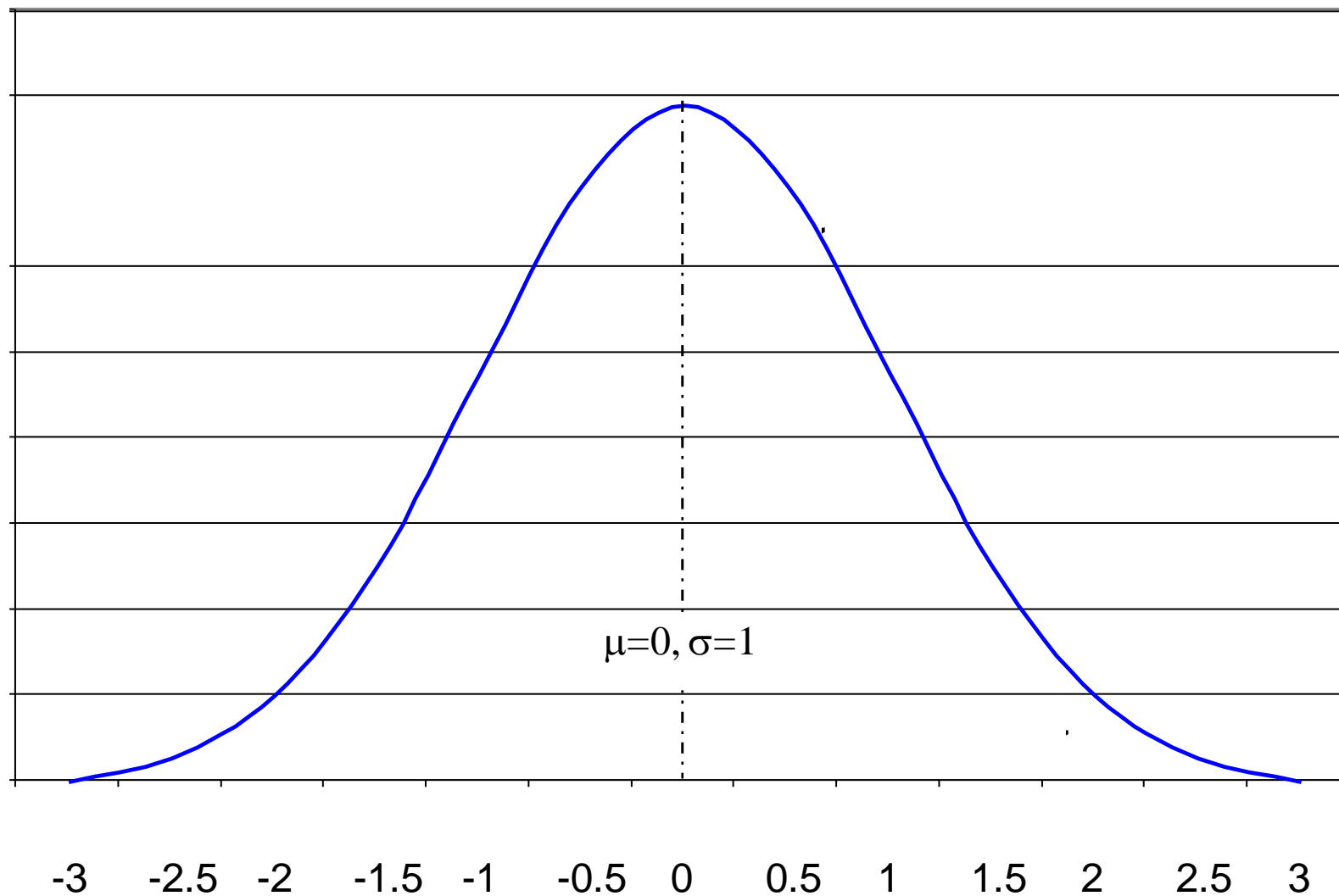
Z: normale standardizzata corrispondente ai valori dati : x, σ, μ .

Normale standardizzata

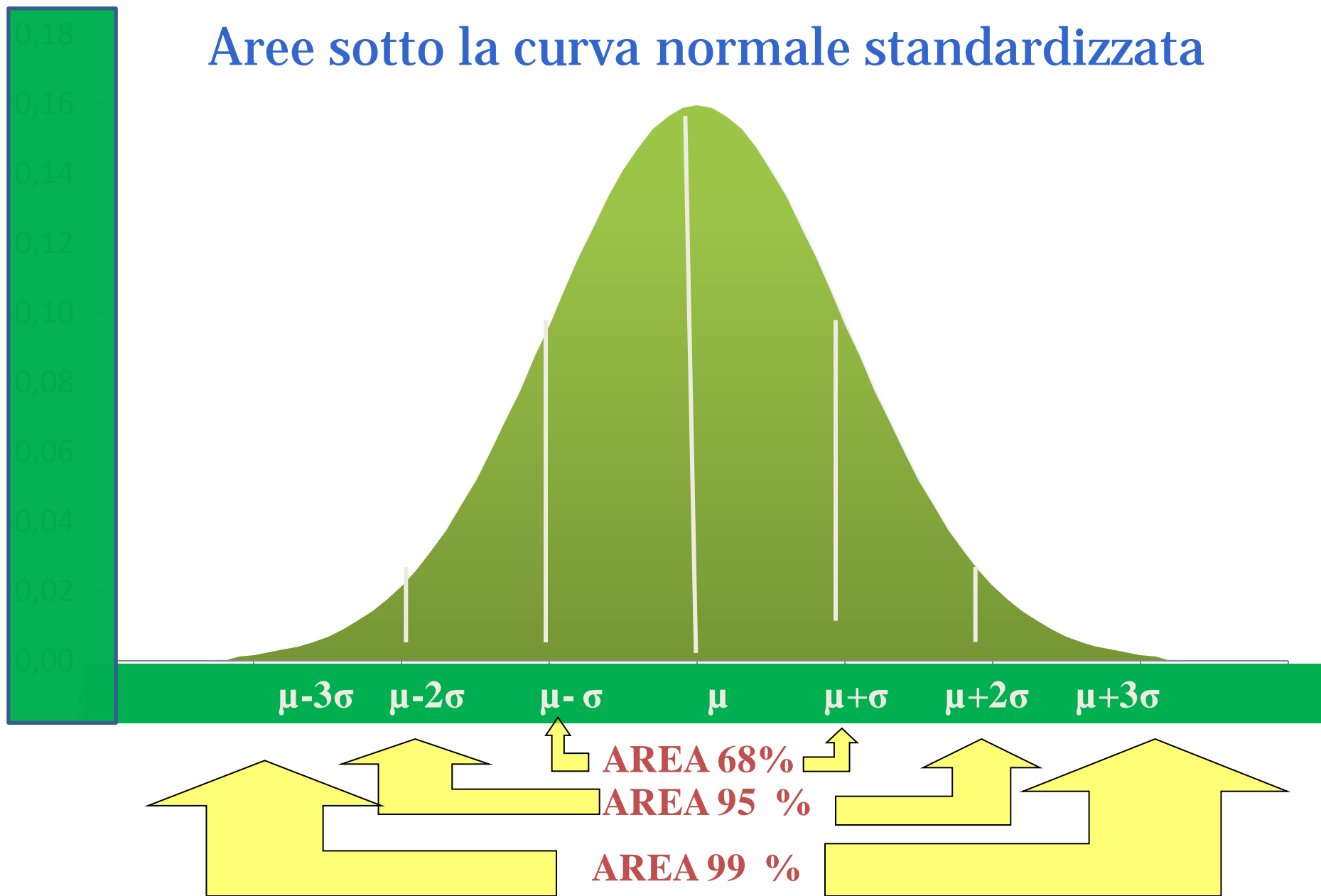
- ❑ Esistono delle tavole che, per determinati valori di z riportano i valori di probabilità.
- ❑ Infatti per una qualunque variabile normale possiamo, utilizzando le tavole della curva normale standardizzata, conoscere la probabilità che la variabile assuma determinati valori
- ❑ **Applicazione: determinazione della probabilità sotto la curva normale standardizzata**

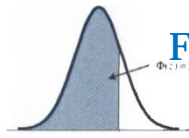
Introduzione all'inferenza statistica

Normale standardizzata



Aree sotto la curva normale standardizzata





Funzione di ripartizione della normale standardizzata.

Seconda cifra decimale di z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

z	1,282	1,645	1,960	2,326	2,576	3,090	3,291	3,891	4,417
$\Phi^{-1}(z)$	0,90	0,95	0,975	0,99	0,995	0,999	0,9995	0,99995	0,999995

Esempio

I pesi di una partita di 1000 cartoni contenenti scatole di biscotti si distribuiscono normalmente con media = 240 kg e $\sigma = 50$ kg.

Calcolare

- a) il numero di cartoni con peso inferiore a 280kg
- b) il numero di cartoni con peso compreso tra 250 e 280 kg;
- c) Il numero di cartoni con peso inferiore a 200kg

Introduzione all'inferenza statistica

Soluzione:

$$X \sim \text{Norm}(240 \text{ kg}, 50 \text{ kg})$$

Calcolare:

- a) il numero di cartoni con $\{X \leq 280 \text{ kg}\}$;
- b) il numero di cartoni con $\{250 \text{ kg} \leq X \leq 280 \text{ kg}\}$;
- c) il numero di cartoni con $\{X \leq 200 \text{ kg}\}$

Introduzione all'inferenza statistica

Soluzione: $X \sim Norm(240 \text{ kg}, 50 \text{ kg})$

a1) Standardizziamo la variabile:

$$z = \frac{280 - 240}{50} = 0,8$$

a2) Cerchiamo sulle tavole della normale standardizzata l'area compresa fra 0 e 0,8: $P(0,8) = 0,2881$

a3) L'area, sotto la curva normale standardizzata, che ci interessa è quella compresa tra $-\infty$ e 0,8. Quindi:

$$0,5 + 0,2881 = 0,7881 \approx 79\%$$

a4) $1000 \times 0,7881 \approx 788$ cartoni hanno peso inferiore a 280 kg.

Introduzione all'inferenza statistica

Soluzione: $X \sim \text{Norm}(240 \text{ kg}, 50 \text{ kg})$

b1) Standardizziamo la variabile per $X = 250$ e per $X = 280$:

$$z_a = \frac{250 - 240}{50} = -0,2 \qquad z_b = \frac{280 - 240}{50} = 0,8$$

b2) Cerchiamo sulle tavole della normale standardizzata l'area compresa fra (0 e -0,2) \equiv (0 e 0,2): $P(0,2) = 0,0793$ e fra 0 e 0,8: $P(0,8) = 0,2881$

b3) L'area, sotto la curva normale standardizzata, che ci interessa è quella compresa tra -0,2 e 0,8. Quindi:

$$P(-0,2) + P(0,8) = 0,0793 + 0,2881 = 0,3674 \approx 37\%$$

b4) $1000 \times 0,3674 \approx 367$ cartoni hanno peso compreso fra 250 kg e 280 kg.

Introduzione all'inferenza statistica

Soluzione:

$$X \sim \text{Norm}(240 \text{ kg}, 50 \text{ kg})$$

c1) Standardizziamo la variabile per $X = 200$:

$$z_a = \frac{200 - 240}{50} = -0,8$$

c2) Cerchiamo sulle tavole della normale standardizzata l'area compresa fra $(0 \text{ e } -0,8) \equiv (0 \text{ e } 0,8)$: $P(0,8) = 0,2881$

c3) L'area, sotto la curva normale standardizzata, che ci interessa è quella compresa tra $(-\infty \text{ e } -0,8) \equiv (0,8 \text{ e } \infty)$.

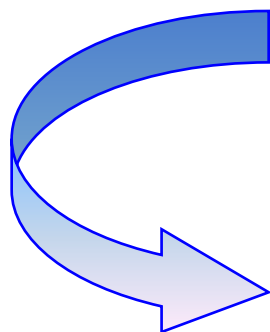
Quindi:

$$0,5 - P(0,8) = 0,5 - 0,2881 = 0,2119 \approx 21\%$$

c4) $1000 \times 0,2881 \approx 288$ cartoni hanno peso inferiore a 200 kg.

Introduzione all'inferenza statistica

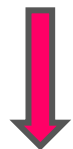
Inferenza statistica



Modelli probabilistici



Valori Caratteristici



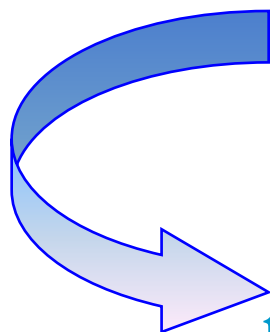
Parametri



Inferenza parametrica

Introduzione all'inferenza statistica

Inferenza statistica



non vincolata da modelli



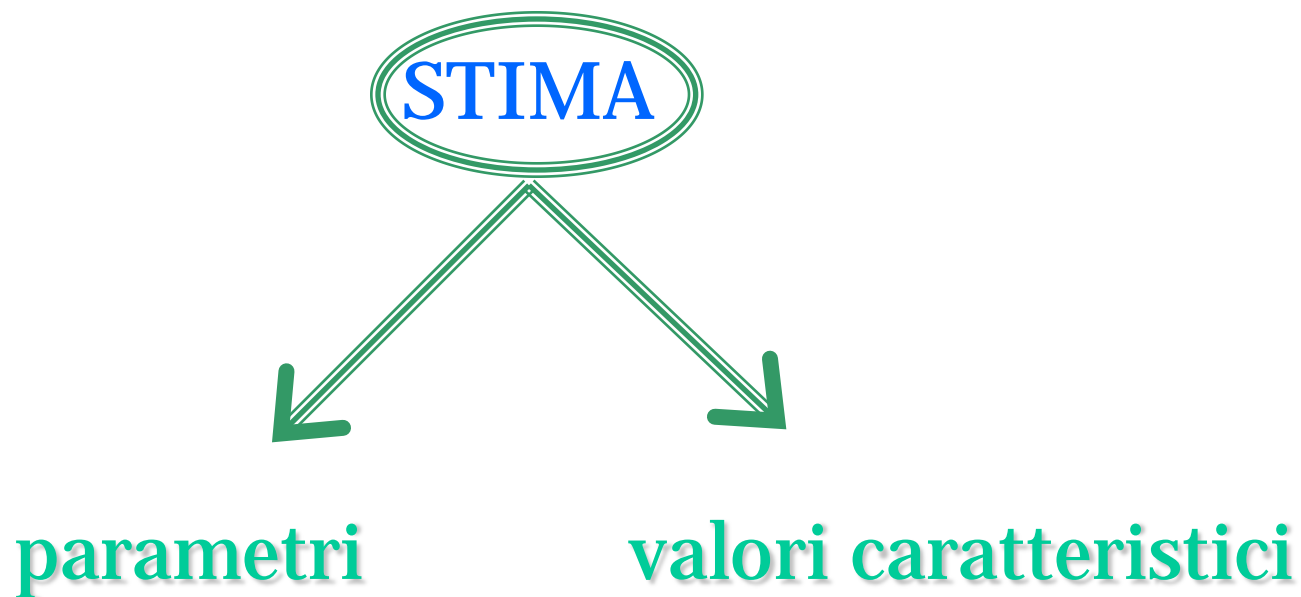
**Caratteristiche qualsiasi non coincidenti con
parametri**



Inferenza non parametrica

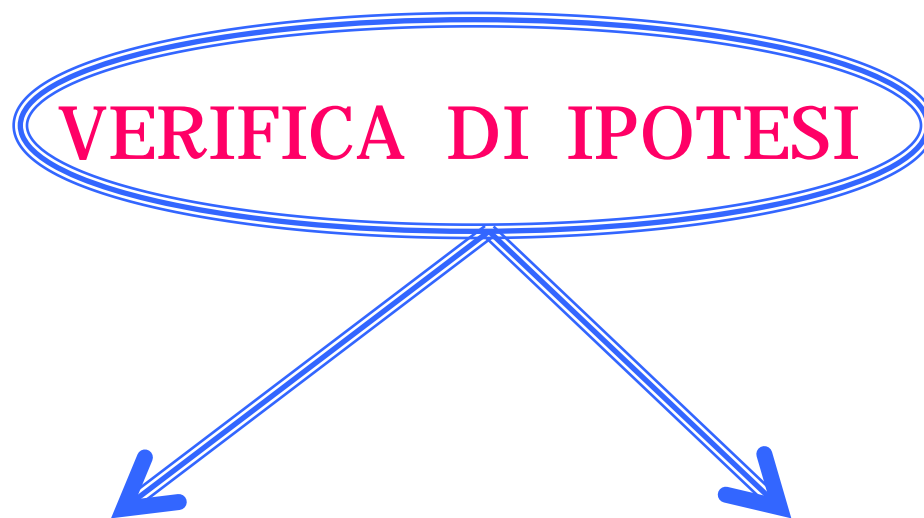
Introduzione all'inferenza statistica

Tecniche inferenziali



Introduzione all'inferenza statistica

Tecniche inferenziali



parametrica

non parametrica

Quali problemi inferenziali affronteremo?

- ❑ Ci interesseremo all'inferenza su :
 - media di una popolazione normale: varianza della popolazione nota e ignota;
 - proporzione in una popolazione dicotomica;
 - relazioni fra due variabili con distribuzione ignota.

Quali problemi inferenziali affronteremo?

- ❑ Nel primo e nel secondo caso faremo riferimento a tecniche di inferenza ***parametrica***.
- ❑ Nel terzo caso faremo riferimento a tecniche di inferenza ***non parametrica***.

Quali tecniche inferenziali considereremo?

- ❑ Stima dei parametri (nel primo e nel secondo caso).
- ❑ Verifica delle ipotesi (in tutti e tre i casi).
- ❑ Determinazione della precisione delle stime:
intervallo di confidenza (nel primo e nel secondo caso).

Inferenza sulle medie: la stima

- Indichiamo con X la variabile della popolazione a cui siamo interessati e scriviamo:

$$X \sim \text{Norm}(\mu, \sigma^2)$$

- Della variabile X conosciamo la varianza ma non la media per cui abbiamo necessità di stimarla.
- Estraiamo dalla popolazione di ampiezza N un campione bernoulliano di numerosità n .

Inferenza sulla media: la stima

- Il campione che estrarremo potrà assumere un qualunque valore per ognuno dei seguenti n :

$$X_1, X_2, X_3, \dots, X_i, \dots, X_n$$

- Il campione estratto sarà il seguente:

$$x_1, x_2, x_3, \dots, x_i, \dots, x_n$$

- Sarà il campione che utilizzeremo per determinare il valore per il parametro incognito.

Inferenza sulla media: la stima

- Per stimare la media della popolazione, quindi, faremo riferimento alla seguente funzione teorica:

$$\bar{X} = \frac{\sum_i^n X_i}{n}$$

- Questa funzione ha, anch'essa una distribuzione di probabilità normale: $\bar{X} \sim \text{Norm}(\mu, \frac{\sigma^2}{n})$

dove $\sigma^2 = \frac{\sum_i^N (x_i - \mu)^2}{N} = \text{varianza popolazione}$

Introduzione all'inferenza statistica

Inferenza sulla media: la stima

- Applicando la funzione teorica precedente ai dati del campione estratto: $x_1, x_2, x_3, \dots, x_i, \dots, x_n$, calcoliamo la stima della media della popolazione:

$$\bar{x} = \hat{\mu} = \frac{\sum_i^n x_i}{n}$$

- Quella che abbiamo ottenuto è la stima per la media della popolazione. Pertanto estendiamo questo risultato all'intera popolazione (facciamo inferenza!)

Inferenza sulla media

- Avendo stimato il parametro incognito, a partire dal valore ottenuto potremo:
- eseguire una verifica di ipotesi mediante la quale sarà possibile accettare o rifiutare delle affermazioni formulate sui valori dei parametri della popolazione.
- Determinare la precisione della stima mediante l'intervallo di confidenza, ossia determinare un intervallo di valori entro cui con un livello di probabilità prefissato (generalmente il 95%) cadrà il parametro μ

Inferenza sulla media: verifica delle ipotesi

- L'ipotesi statistica è un'assunzione che viene formulata sul valore di un parametro della popolazione (ad es. la media).
- Le ipotesi formulate sono due:

H_0 : ipotesi nulla o ipotesi principale
 H_1 : ipotesi alternativa

Inferenza sulla media: verifica delle ipotesi

- Lo strumento che permette di decidere riguardo all'attendibilità dell'ipotesi (rifiutare o non rifiutare l'ipotesi) prende il nome di test statistico ed è funzione delle osservazioni campionarie.
- In particolare il test è una procedura che, sulla base di dati campionari e con un certo grado di probabilità, consente di decidere se è ragionevole accettare l'ipotesi nulla H_0 oppure respingerla (propendendo implicitamente per l'ipotesi alternativa H_1).

Inferenza sulla media: verifica delle ipotesi

- Nella verifica d'ipotesi si procede come segue:
 - Costruiamo il sistema d'ipotesi, definendo diverse ipotesi alternative.
- 1. Caso di ipotesi nulla semplice ed alternativa semplice:

$\begin{cases} H_0: \text{la media della popolazione è uguale ad un certo valore} \\ H_1: \text{la media della popolazione è uguale ad un altro valore} \end{cases}$

Inferenza sulla media: verifica delle ipotesi

2. Caso di ipotesi nulla semplice ed alternativa complessa, caso bidirezionale:

$\begin{cases} H_0: \text{la media della popolazione è uguale ad un certo valore} \\ H_1: \text{la media della popolazione è diversa da quel valore} \end{cases}$

3. Caso di ipotesi nulla semplice ed alternativa complessa, caso unidirezionale destro (o positivo):

$\begin{cases} H_0: \text{la media della popolazione è uguale ad un certo valore} \\ H_1: \text{la media della popolazione è maggiore di quel valore} \end{cases}$

Inferenza sulla media: verifica delle ipotesi

4. Caso di ipotesi nulla semplice ed alternativa complessa, caso unidirezionale sinistro (o negativo):

$\begin{cases} H_0: \text{la media della popolazione è uguale ad un certo valore} \\ H_1: \text{la media della popolazione è minore di quel valore} \end{cases}$

- Le ipotesi nulle sono sempre semplici mentre quelle alternative più frequentemente sono complesse

Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi

□ Quanto abbiamo visto si scrive come segue:

$$1. \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu = \mu_1 \end{cases} \quad 2. \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

$$3. \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases} \quad 4. \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi

□ Dopo aver formulato le ipotesi passiamo a verificare se possiamo accettare l'ipotesi nulla o rifiutarla.

□ Consideriamo, dapprima il caso, in cui conosciamo la varianza della popolazione. Pertanto calcoliamo il test statistico:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

□ Il test statistico è il test Z che abbiamo ottenuto standardizzando la variabile: $\bar{X} \sim \text{Norm}(\mu, \frac{\sigma^2}{n})$ dove σ / \sqrt{n} è lo scarto quadratico medio della media campionaria.

Inferenza sulla media: verifica delle ipotesi

- ❑ Dopo aver calcolato il test sulla base del risultato ottenuto dobbiamo decidere se accettare o rifiutare l'ipotesi. Come facciamo?
- ❑ Dobbiamo stabilire una soglia che ci permetta di decidere.
- ❑ Questa soglia la fissiamo in base ad una probabilità che indichiamo con α . Di solito α è molto piccolo: 5% o 1%.
- ❑ Qual è il ruolo di α ?

Inferenza sulla media: verifica delle ipotesi

- ❑ Alla probabilità α corrispondono dei valori di z che possono essere trovati sulle tavole della variabile normale standardizzata.

- ❑ Questi valori sono:
 1. nel caso unidirezionale: per $\alpha = 0,05$, $z = 1,64$;
 2. nel caso bidirezionale: per $\alpha = 0,025$, $z = 1,96$.

Inferenza sulla media: verifica delle ipotesi

❑ Decidiamo di accettare o rifiutare l'ipotesi nulla se risulta:

1) Unidirezionale destro:

$z \leq z_\alpha$ accettiamo l'ipotesi nulla; $z > z_\alpha$ rifiutiamo l'ipotesi nulla

2) Unidirezionale sinistro:

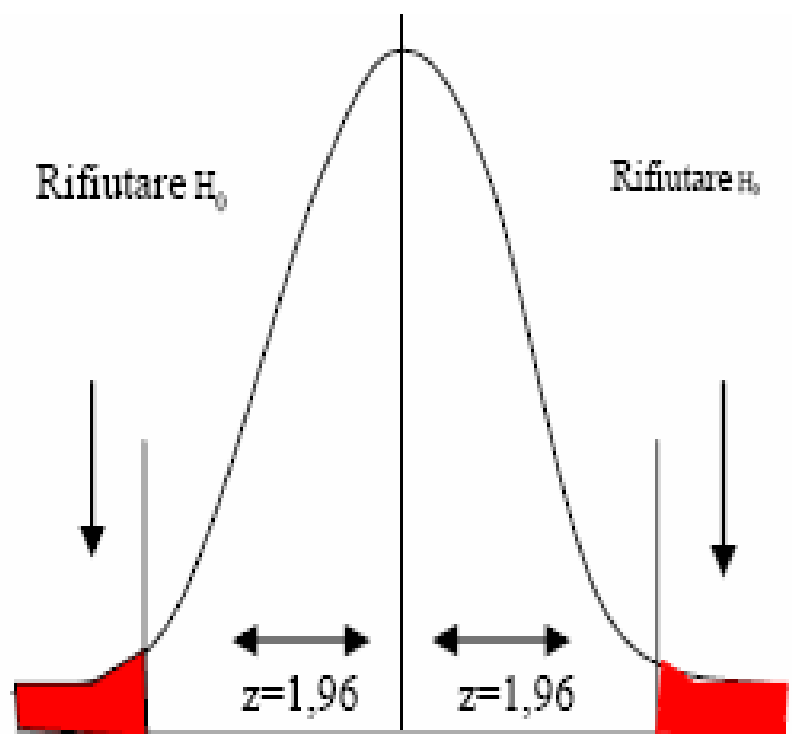
$z \geq z_\alpha$ accettiamo l'ipotesi nulla; $z < z_\alpha$ rifiutiamo l'ipotesi nulla

3) Bidirezionale:

$z_{\alpha/2} \leq z \leq z_{\alpha/2}$ accettiamo l'ipotesi nulla; $z < z_{\alpha/2}$ o $z > z_{\alpha/2}$
rifiutiamo l'ipotesi nulla

Introduzione all'inferenza statistica

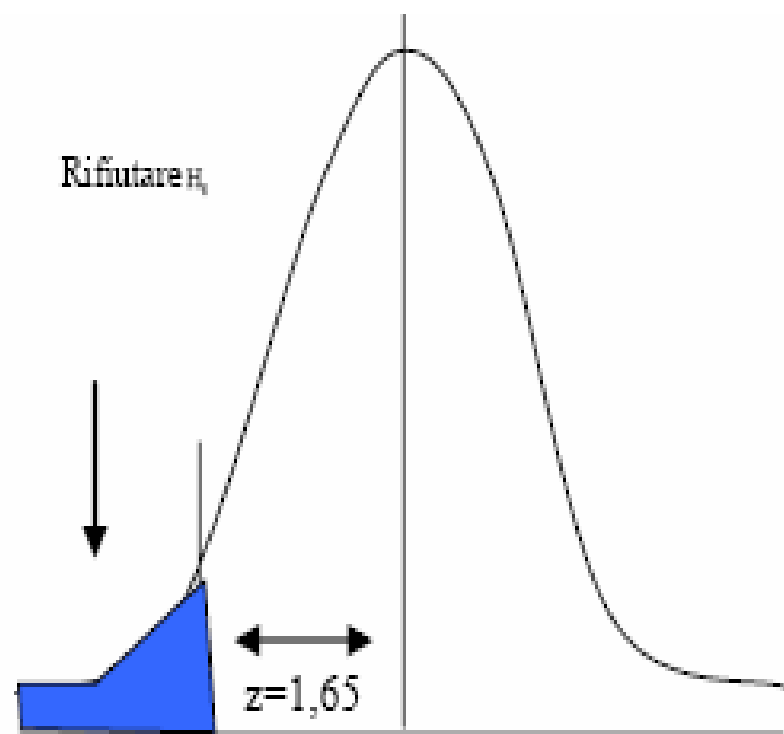
Inferenza sulla media: verifica delle ipotesi



0,025

0,025

$\alpha=0,05$

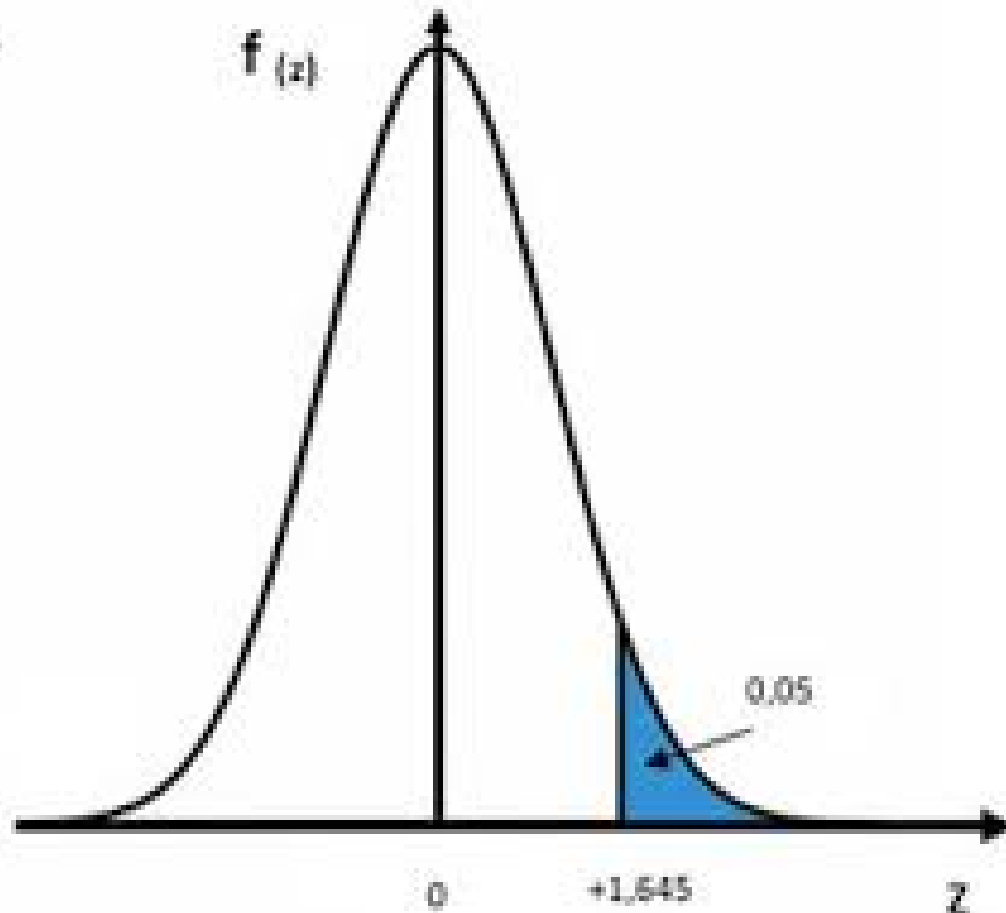


0,05

$\alpha=0,05$

Introduzione all'inferenza statistica

- Regione di rifiuto nel caso di verifica di ipotesi unidirezionale destra



Inferenza sulla media: verifica delle ipotesi

Allorché abbiamo preso la decisione siamo certi del risultato?

- Operiamo in situazioni di incertezza, pertanto possiamo aver sbagliato *sia se abbiamo rifiutato* un'ipotesi nulla, che invece potrebbe essere **vera**, *sia se abbiamo accettato* l'ipotesi nulla perché, al contrario, potrebbe essere **falsa** dato che operiamo con dati campionari, quindi con informazioni parziali.

Inferenza sulla media: verifica delle ipotesi

- Pertanto, possiamo distinguere gli errori che potremmo commettere, allorché prendiamo una decisione, in due categorie:
 - ✓ **errore di prima specie** che consiste nel rifiutare un'ipotesi nulla che potrebbe essere vera;
 - ✓ **errore di seconda specie** che consiste nell'accettare un'ipotesi nulla che potrebbe essere falsa.

Inferenza sulla media: verifica delle ipotesi

- La probabilità di poter commettere questi errori è:
 - ✓ $\Pr(\text{errore di prima specie}) = \alpha$. Tale probabilità viene fissata a priori.
 - ✓ $\Pr(\text{errore di seconda specie}) = \beta$. Tale probabilità può essere calcolata.
 - ✓ Il complemento di β , $1 - \beta$ è detto **potenza del test**.

Introduzione all'inferenza statistica

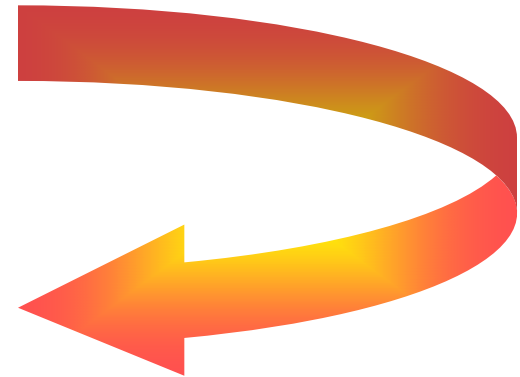
Inferenza sulla media: verifica delle ipotesi

🌸 **Probabilità d'errore di II specie:**

$$\beta = \int_{-\infty}^{\bar{x}_\alpha} f_1(\bar{x}) d\bar{x}$$

🌸 **Potenza del test**

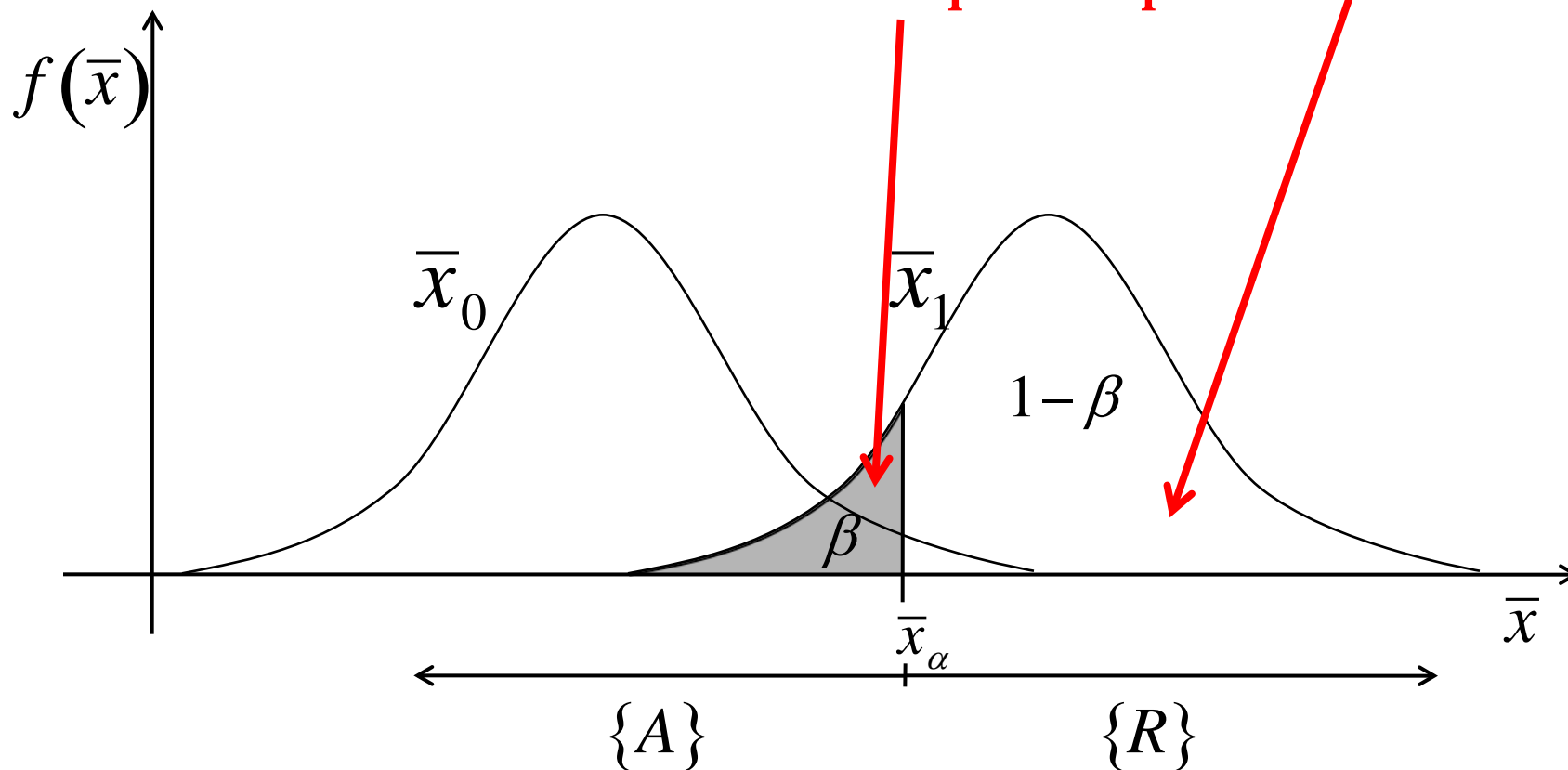
$$1 - \beta = \int_{\bar{x}_\alpha}^{\infty} f_1(\bar{x}) d\bar{x}$$



Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi

Probabilità dell'errore di II specie e potenza del test



Inferenza sulla media: verifica delle ipotesi

□ *P-Value*

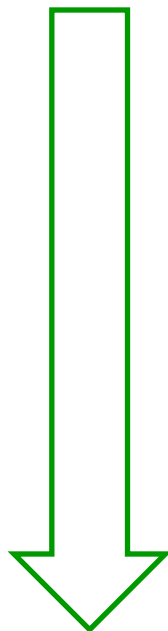
- 😊 probabilità di osservare, sotto l'ipotesi nulla, un valore della statistica test almeno uguale a quello ottenuto dal campione
- 😊 il più piccolo livello di significatività al di sopra del quale si rifiuta l'ipotesi nulla
- 😊 definito anche *livello di significatività osservato*

Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi

P-Value

Come si usa?



Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi

P-Value

- Prefissato α , se:

$$\alpha > p - value$$

Rifiutiamo H_0

$$\alpha < p - value$$

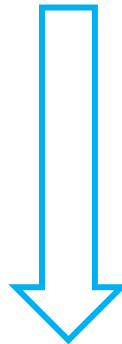
Accettiamo H_0

Introduzione all'inferenza statistica

➤ **Un p -value basso porta a rifiutare l'ipotesi nulla.**

✎ Si osservi che il **p -value** è una probabilità osservata, diversa per ogni valore della statistica campionaria ottenuta in quanto dipende dall'informazione campionaria che viene presa in considerazione in quel momento.

➤ Nei lavori di ricerca se vi sono risultati empirici spesso viene usata questa classificazione:



Introduzione all'inferenza statistica

$0,00 < \text{p-value} < 0,01$  forte evidenza empirica contro H_0

$0,01 < \text{p-value} < 0,05$  evidenza empirica contro H_0

$0,05 < \text{p-value} < 0,10$  scarsa evidenza empirica contro H_0

$0,10 < \text{p-value} < 1,00$  non c'è evidenza empirica contro H_0

□ Verifica di ipotesi per la media di una popolazione

📁 **Esempio:** Un campione di $n = 11$ studenti, estratto da una quinta classe di un Liceo scientifico avente varianza pari a $0,25$, ha ottenuto i seguenti voti allo scritto di matematica (x):

Voto in Matematica	6	6	7	8	9	7	6	5	3	7	5
--------------------	---	---	---	---	---	---	---	---	---	---	---

Ipotizzando che il voto di matematica si distribuisca normalmente:

1. stimare il voto medio;
2. verificare l'ipotesi che il voto medio in matematica della popolazione degli studenti sia pari a 7 contro le seguenti alternative: a) che sia diverso da 7; b) che sia >7 ; c) che sia <7 , posto $\alpha = 0,01$.

Introduzione all'inferenza statistica

□ Verifica di ipotesi per la media di una popolazione

1. Stimiamo il voto medio come segue:

$$\hat{\mu} = \frac{\sum_i^n x_i}{n} = \frac{69}{11} = 6,27$$

2. Costruiamo il sistema di ipotesi:

$$a) \begin{cases} \mu = 7 \\ \mu \neq 7 \end{cases}$$


$$b) \begin{cases} \mu = 7 \\ \mu < 7 \end{cases}$$

$$c) \begin{cases} \mu = 7 \\ \mu > 7 \end{cases}$$

Verifichiamo l'ipotesi considerando il seguente test:

□ Stima e verifica di ipotesi per la media di una popolazione

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{6,27 - 7}{\frac{0,5}{\sqrt{11}}} = \frac{-0,73}{1,5076} = -0,15$$

- a) Consideriamo il valore ottenuto in valore assoluto e lo confrontiamo con il valore assoluto della normale standardizzata che, per $\alpha/2 = 0,025$, è pari a **1,96**  **0,15 < 1,96**, quindi accettiamo l'ipotesi che il voto medio in matematica sia pari a 7.
- b) Essendo $-0,15 > -1,64$ (valore di z per $\alpha = 0,05$) accettiamo l'ipotesi che il voto medio in matematica sia pari a 7.
- c) Essendo $-0,15 < 1,64$ (valore di z per $\alpha = 0,05$) accettiamo l'ipotesi che il voto medio in matematica sia pari a 7.

Inferenza sulla media: verifica delle ipotesi (varianza ignota)

□ Abbiamo considerato il problema della verifica delle ipotesi sulla media partendo dall'ipotesi che la varianza della popolazione sia nota. Se invece non la conosciamo (caso molto più frequente), partendo dalle informazioni del campione estratto, dobbiamo stimarla come segue :

$$s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n-1}$$

Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi (varianza ignota)

- Pertanto i problemi precedenti saranno risolti mediante il seguente test:

$$t = \frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}}$$

Inferenza sulla media: verifica delle ipotesi (varianza ignota)

➔ Regole di decisione

- $t \leq t_{\alpha;n-1}$ si accetta H_0 ;
 $t > t_{\alpha;n-1}$ si rifiuta H_0 ;
- $t \geq t_{\alpha;n-1}$ si accetta H_0 ;
 $t < t_{\alpha;n-1}$ si rifiuta H_0 ;

Introduzione all'inferenza statistica

Inferenza sulla media: verifica delle ipotesi (varianza ignota)

3. $- t_{\alpha/2;n-1} \leq t \leq t_{\alpha/2;n-1}$ si accetta H_0

$- t < - t_{\alpha/2;n-1}$ o $t > t_{\alpha/2;n-1}$ si rifiuta H_0

- ✓ dove $t_{\alpha/2}$ è il valore della T di Student, con $n-1$ gradi di libertà (g.d.l.) e va ricercato sulle tavole di questa variabile casuale in corrispondenza di $\alpha/2$ e di $n-1$.
- ✓ Se n è sufficientemente elevato ($n > 30$) i valori della v. c. T di Student coincidono con quelli della normale standardizzata.

□ Verifica di ipotesi per la media di una popolazione

📁 **Esempio:** Consideriamo lo stesso campione dell'esercizio precedente ma, in questo caso, non conosciamo la varianza della popolazione

Voto in Matematica	6	6	7	8	9	7	6	5	3	7	5
--------------------	---	---	---	---	---	---	---	---	---	---	---

Ipotizzando che il voto di matematica si distribuisca normalmente:

1. stimare il voto medio e la varianza;
2. verificare l'ipotesi che il voto medio in matematica della popolazione degli studenti sia pari a 7 contro le seguenti alternative: a) che sia diverso da 7; b) che sia >7 ; c) che sia <7 , posto $\alpha = 0,01$.

Introduzione all'inferenza statistica

□ Verifica di ipotesi per la media di una popolazione

1. Stimiamo il voto medio come segue:

$$\hat{\mu} = \frac{\sum_i^n x_i}{n} = \frac{69}{11} = 6,27 \qquad s^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n-1} = \frac{28,8}{11-1} = \frac{28,8}{10} = 2,88$$

2. Costruiamo il sistema di ipotesi:

$$a) \begin{cases} \mu = 7 \\ \mu \neq 7 \end{cases}$$

$$b) \begin{cases} \mu = 7 \\ \mu < 7 \end{cases}$$

$$c) \begin{cases} \mu = 7 \\ \mu > 7 \end{cases}$$

Verifichiamo l'ipotesi considerando il seguente test:

Introduzione all'inferenza statistica

□ Stima e verifica di ipotesi per la media di una popolazione

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{6,27 - 7}{\frac{1,6971}{\sqrt{11}}} = \frac{-0,73}{0,5117} = -1,43$$

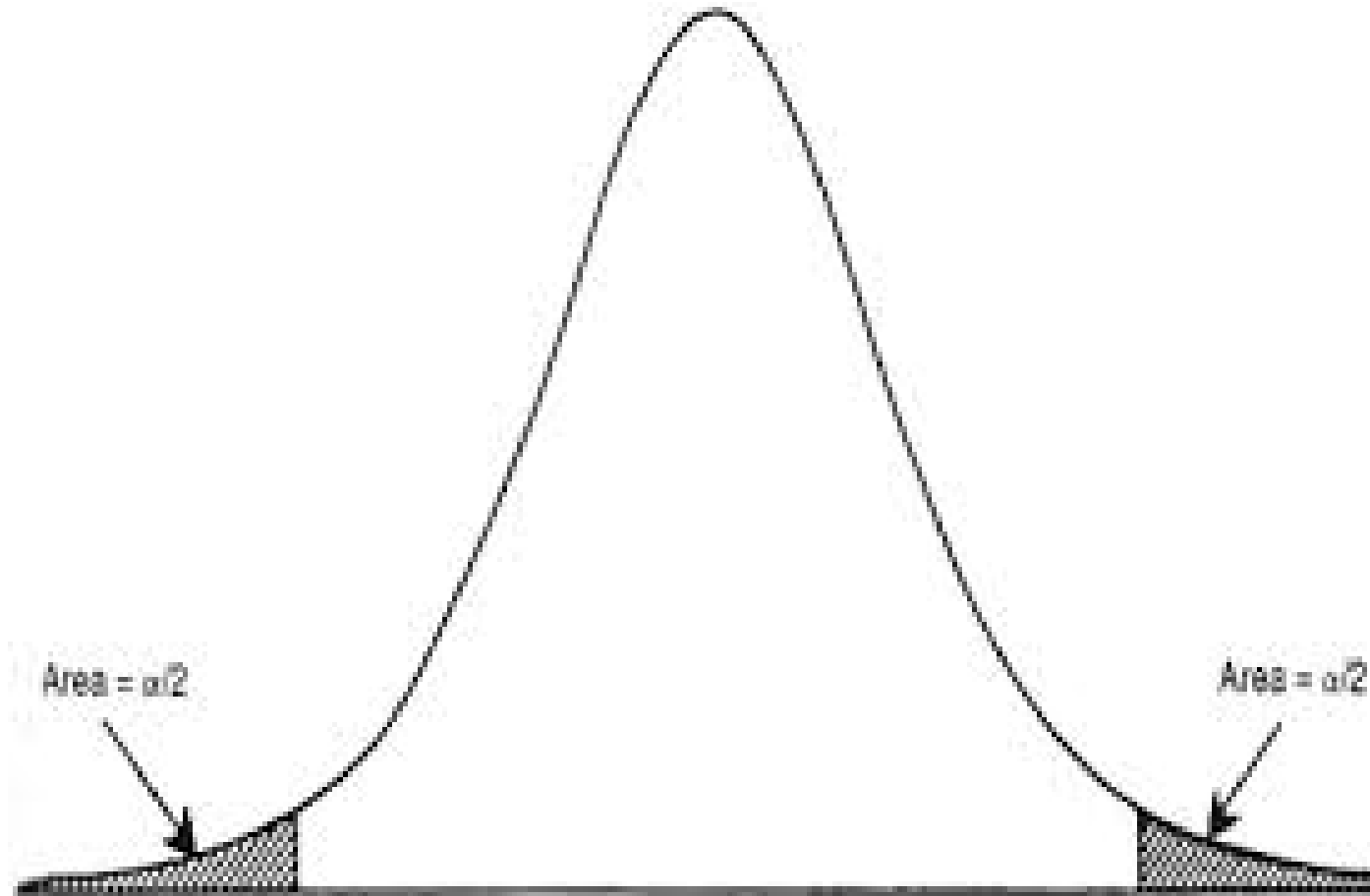
- a) Cerchiamo sulle tavole della T di Student il valore in corrispondenza di **n-1 (11-1)** gradi di libertà e **$\alpha/2 = 0,025$** e quello con **n-1 (11-1)** e **$\alpha = 0,05$** ;
- b) I valori sono, rispettivamente, **2,228** e **1,812**;
- c) Nel caso bidirezionale confrontiamo il valore assoluto del test con quello della T: **1,43 < 2,228** per cui accettiamo l'ipotesi che il voto medio in matematica sia pari a 7.
- d) Nel caso unidirezionale sinistro : **-1,43 > -1,812** per cui accettiamo l'ipotesi che il voto medio in matematica sia pari a 7.
- e) Nel caso unidirezionale destro: **-1,43 < 1,812** per cui accettiamo l'ipotesi che il voto medio in matematica sia pari a 7.

Introduzione all'inferenza statistica

□ Osservazioni

- ✓ La T di Student è simmetrica ed ha forma campanulare per cui le regioni di rifiuto graficamente sono analoghe a quanto visto per la normale standardizzata.
- ✓ Per campioni la cui dimensione è ≥ 30 , pur non conoscendo la varianza della popolazione, si può usare il test z in quanto la T converge verso una normale standardizzata.

□ T di Student: regioni di rifiuto nel caso bidirezionale



Inferenza sulla media: intervallo di confidenza

- ❑ Per procedere con le tecniche inferenziali è necessario considerare la **stima puntuale** ossia fornire **un valore** per il parametro incognito mediante i dati campionari.
- ❑ Il dato ottenuto ci permette di operare con le tecniche inferenziali.

Inferenza sulla media: intervallo di confidenza

- ❑ **Osservazione.** Non viene definita alcuna probabilità che questo dato ha di verificarsi.
- ❑ Pertanto diventa necessario costruire un intervallo nel quale ci aspettiamo sia contenuto il parametro da stimare con un elevato grado di fiducia.
- ❑ Questa “fiducia” è espressa in termini probabilistici e viene detta confidenza.

Inferenza sulla media: intervallo di confidenza

- Un tale intervallo si dice **intervallo di confidenza** e la probabilità (che indicheremo con $1-\alpha$) assegnata viene detta **livello di significatività** (o livello di confidenza). Usualmente si sceglie come livello di significatività il 95% o il 99%.

- Costruiremo l'intervallo di confidenza per la media in due situazioni:
 - a. Caso di varianza della popolazione nota
 - b. Caso di varianza della popolazione ignota, quindi, stimata.

Inferenza sulla media: intervallo di confidenza

- Consideriamo il caso **a. varianza nota**. L'intervallo di confidenza verrà scritto come segue:

$$1 - \alpha = \Pr \left\{ \bar{x} - z_A \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_B \frac{\sigma}{\sqrt{n}} \right\}$$

dove

$$z_A = z_B = z_{\frac{\alpha}{2}}$$

Introduzione all'inferenza statistica

Inferenza sulla media: intervallo di confidenza

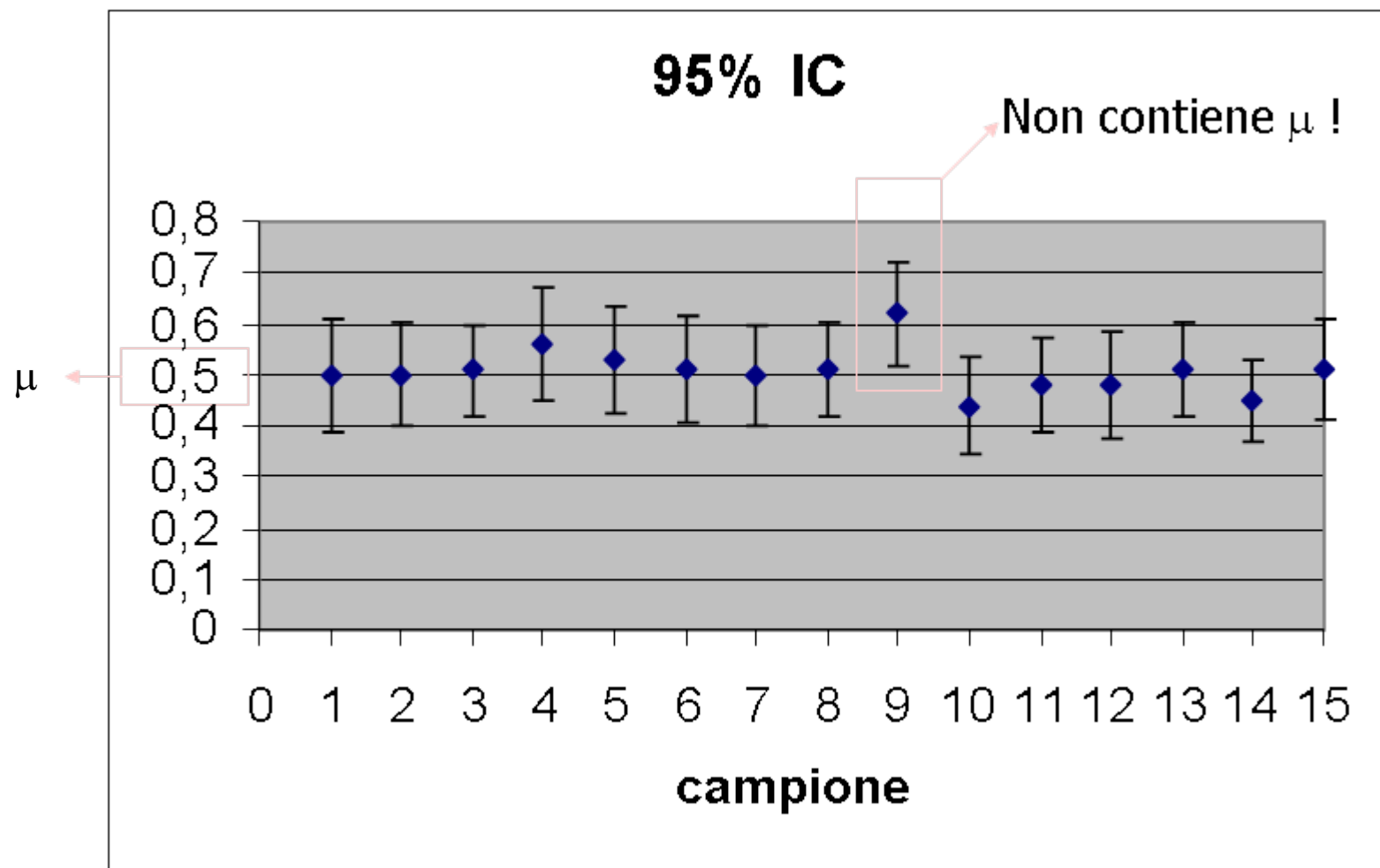
- Come già detto il livello di confidenza è, frequentemente: **95%** o **99%**.
- Pertanto i corrispondenti valori di $z_{\alpha/2}$ sono:

$$1-\alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow z_{\alpha/2} = z_{0,025} = 1,96$$

$$1-\alpha = 0,99 \rightarrow \alpha = 0,01 \rightarrow z_{\alpha/2} = z_{0,005} = 2,58$$

Introduzione all'inferenza statistica

Inferenza sulla media: intervallo di confidenza



Introduzione all'inferenza statistica

Intervallo di confidenza, varianza della popolazione nota: esempio

- Riprendiamo l'esempio precedente e determiniamo l'intervallo di confidenza al 95% per la media della popolazione. Consideriamo il seguente intervallo teorico:

$$1 - \alpha = \Pr \left\{ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

- ricordando che per $\alpha = 0,05$, $\alpha/2 = 0,025$, $z = 1,96$, sostituendo le altre quantità, si ha:

Introduzione all'inferenza statistica

Intervallo di confidenza, varianza della popolazione

nota: esempio

$$\begin{aligned} 0,95 &= \Pr\left\{6,27 - 1,96 \frac{0,5}{\sqrt{11}} \leq \mu \leq 6,27 + 1,96 \frac{0,5}{\sqrt{11}}\right\} = \\ &= \Pr\{6,27 - 1,96 * 0,1508 \leq \mu \leq 6,27 + 1,96 * 0,1508\} = \\ &= \Pr\{5,97 \leq \mu \leq 6,56\} \end{aligned}$$

Con probabilità del 95% il voto medio è compreso fra 5,97 e 6,56!

Inferenza sulla media: intervallo di confidenza

- Consideriamo il caso **b. varianza ignota**. L'intervallo di confidenza verrà scritto come segue:

$$1 - \alpha = \Pr \left\{ \hat{\mu} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \hat{\mu} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right\}$$

dove $t_{\alpha/2}$ è il valore della T di Student, con $n-1$ gradi di libertà (g.d.l.) e va ricercato sulle tavole di questa variabile casuale in corrispondenza di $\alpha/2$ e di $n-1$.

Introduzione all'inferenza statistica

Intervallo di confidenza, varianza della popolazione ignota: esempio

- Riprendiamo l'esempio già visto per la verifica delle ipotesi e consideriamo il seguente intervallo:

$$1 - \alpha = \Pr \left\{ \hat{\mu} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \hat{\mu} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right\}$$

- Consideriamo, nuovamente, $1 - \alpha = 0,95$ e cerchiamo sulle tavole della T di Student, il valore in corrispondenza di 0,025 e di 10. Questo valore è: **2,228**

Introduzione all'inferenza statistica

Intervallo di confidenza, varianza della popolazione ignota: esempio

□ Sostituiamo:

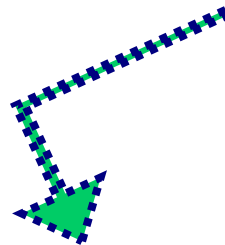
$$0,95 = \Pr\left\{6,27 - 2,228 \frac{1,6971}{\sqrt{11}} \leq \mu \leq 6,27 + 2,228 \frac{1,6971}{\sqrt{11}}\right\}$$

$$0,95 = \Pr\{6,27 - 2,228 * 0,5117 \leq \mu \leq 6,27 + 2,228 * 0,5117\}$$

$$0,95 = \Pr\{5,13 \leq \mu \leq 7,41\}$$

Con probabilità del 95% il voto medio è compreso fra 5,13 e 7,41! Questo intervallo è leggermente più ampio perché risente della variabilità campionaria!

- ☺ Dimensione campionaria nel caso di stima per intervallo per la media μ
- ❖ Obiettivo è quello di determinare la dimensione del campione necessaria affinché la precisione delle stime della media sia contenuta in limiti accettabili. Pertanto



Introduzione all'inferenza statistica

❖ E' necessario prefissare

a. il grado di fiducia $1-\alpha$;

b. l'ampiezza massima dell'intervallo di confidenza (**2d**).

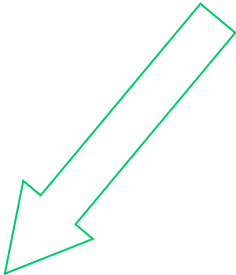


E' necessario, inoltre, che si conosca il valore di σ^2 .

Introduzione all'inferenza statistica

 Considerando l'intervallo di confidenza

$$1 - \alpha = \Pr \left\{ \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\}$$

$$n = \frac{z_{\frac{\alpha}{2}}^2 \sigma^2}{d^2}$$


Introduzione all'inferenza statistica

☺ Dimensione campionaria nel caso di verifica di ipotesi sul valore di μ

 conoscere, almeno approssimativamente, σ^2 ;

 prefissare α e β ;

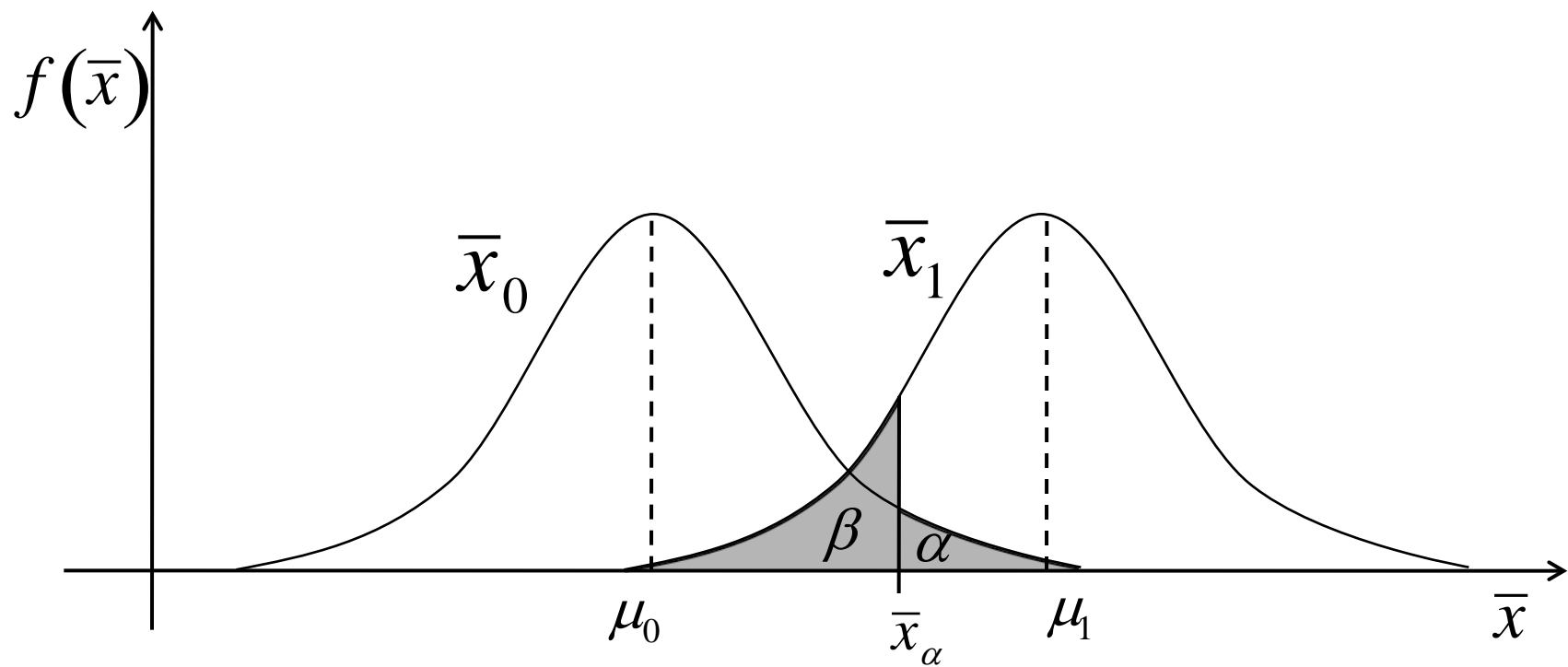
 prefissare il parametro di non centralità (δ) o un minimo per esso (δ_{\min}).

Introduzione all'inferenza statistica

 Consideriamo il caso più semplice:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases} \quad \text{con } \mu_1 > \mu_0$$

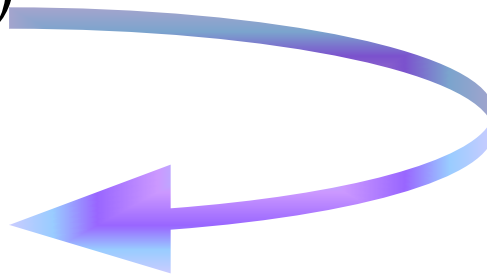
Introduzione all'inferenza statistica



Introduzione all'inferenza statistica



$$n = \frac{\sigma^2}{\delta^2} (z_\beta - z_\alpha)^2$$



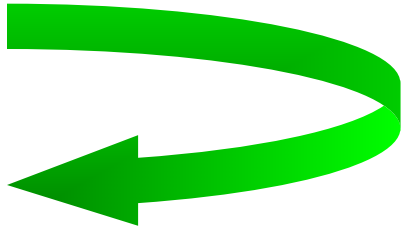
- Allo stesso risultato si giunge anche in caso di ipotesi alternativa semplice con $\mu_1 < \mu_0$.
- Se l'ipotesi è complessa unidirezionale si dovrà sostituire nell'espressione a δ , δ_{\min}

Introduzione all'inferenza statistica

Ipotesi alternativa bidirezionale

$$H_1 : \mu \neq \mu_0$$

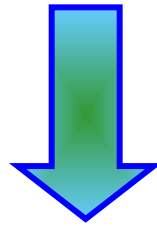
Posto $\delta'_{\min} = \delta''_{\min}$

$$n = \frac{\sigma^2}{\delta^2} \left(z_{\frac{\beta}{2}} - z_{\frac{\alpha}{2}} \right)^2$$


Introduzione all'inferenza statistica

 Modello dicotomico

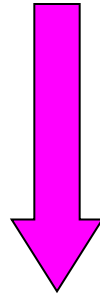
 Variabile X che presenta due sole modalità
- incompatibili e esaustive



SI, NO; presente, assente; bianco, altro
colore; <10 , ≥ 10 ; ecc. ...

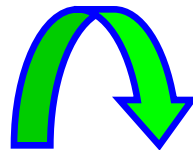
Introduzione all'inferenza statistica

↪ Tali modalità possono essere codificate assegnando i valori **0** e **1**.

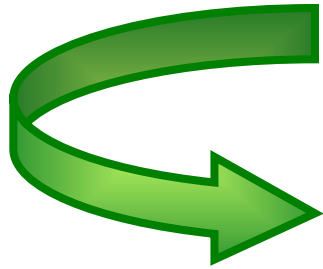


$$\pi = \Pr\{X = x = 1\}$$

$$1 - \pi = \Pr\{X = x = 0\}$$



Introduzione all'inferenza statistica



$$p(x) = \begin{cases} 1 - \pi & \text{per } x = 0 \\ \pi & \text{per } x = 1 \end{cases} \quad x = 0, 1$$


$$E(X) = \pi \quad \text{Var}(X) = \pi(1 - \pi)$$

-Proprietà di X:

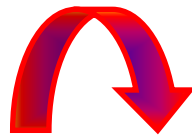
a) $X=0, 1$; b) un solo parametro, π .

Introduzione all'inferenza statistica

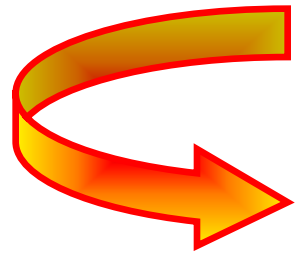
❖ Il parametro da stimare è π .

 Dato un campione casuale semplice di ampiezza n , la modalità 1 , potrà presentarsi X volte ($X = 0, 1, \dots, n$).

 La probabilità che $X = 0, 1, 2, \dots, n$ è



Introduzione all'inferenza statistica

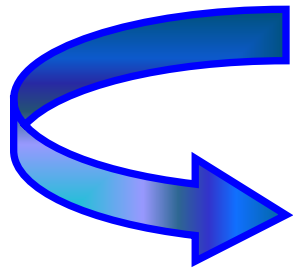


$$p(X) = \binom{n}{X} \pi^X (1 - \pi)^{(n-X)} \quad X = 0, 1, \dots, n$$

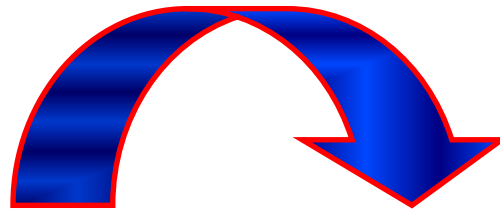


Funzione di probabilità di una v. c. binomiale

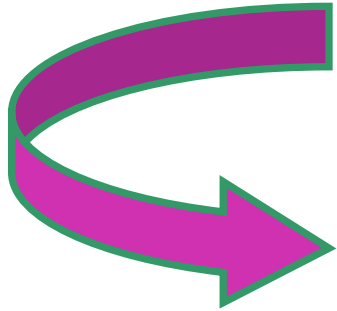
↘ dovendo stimare π facciamo ricorso al metodo della massima verosimiglianza.



Considerando \mathbf{P} quale stimatore di π e scrivendo la $p(\mathbf{X})$ come funzione di \mathbf{P}



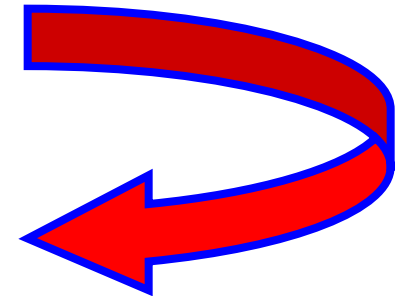
Introduzione all'inferenza statistica



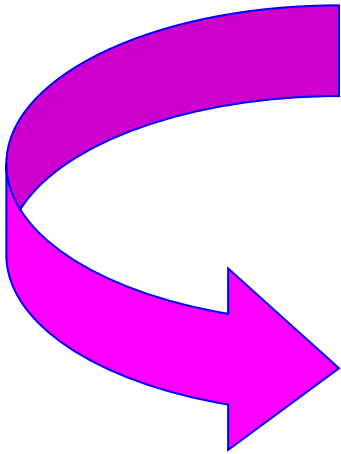
$$P = \frac{X}{n}$$

X = numero di volte che si presenta $X = 1$
dato un campione di ampiezza n

$X \sim \text{Binomiale}$



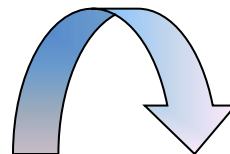
Introduzione all'inferenza statistica



$$E(P) = \pi \quad \text{Var}(P) = \frac{\pi(1 - \pi)}{n}$$

Lo stimatore **P** è:

- 1) corretto; 2) efficiente;
- 3) consistente; 4) sufficiente.

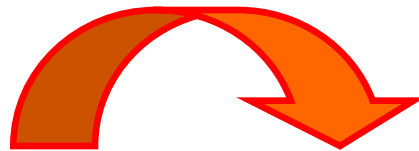


Introduzione all'inferenza statistica

 **Stima puntuale:**

$$\hat{\pi} = p = \frac{x}{n}$$

Dove x è il numero di unità che presentano la modalità $\mathbf{x = 1}$.



Introduzione all'inferenza statistica

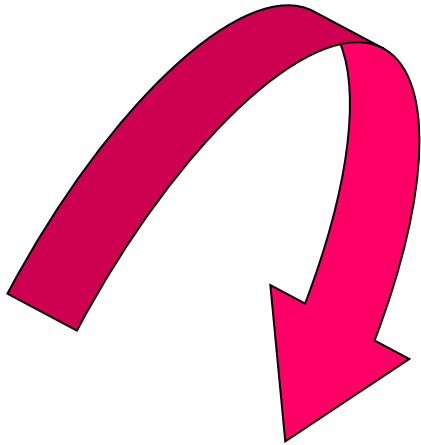
○ Grandi campioni

- ❖ La dimensione del campione può ritenersi elevata allorché risulta :

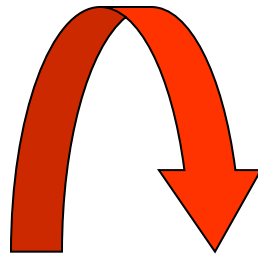
$$\begin{cases} n\pi \geq 20 & \text{se } \pi < 0,50 \\ n(1 - \pi) \geq 20 & \text{se } \pi > 0,50 \end{cases}$$



Introduzione all'inferenza statistica



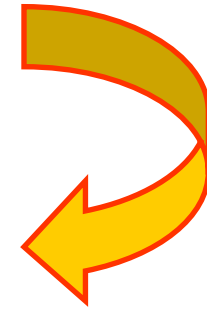
$$P \approx \text{Norm} \left(\pi; \frac{\pi(1-\pi)}{n} \right)$$



Introduzione all'inferenza statistica

Stima per intervallo

$$Z = \frac{P - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}}$$



$$\Pr \left\{ p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right\} = 1 - \alpha$$

Introduzione all'inferenza statistica

Esempio:

Su un campione di 400 bibliotecari, 335 sono donne. Costruiamo l'intervallo di confidenza al 95% della proporzione dei bibliotecari che sono donne.

Soluzione:

Dobbiamo considerare il seguente intervallo

$$\Pr \left\{ p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right\} = 1 - \alpha$$

Introduzione all'inferenza statistica

- Stimiamo dapprima π mediante lo stimatore $P = X/n$
- La stima sarà: $p = \frac{x}{n} = \frac{335}{400} = 0,84$
- Avendo posto $1 - \alpha = 0,95 \implies \alpha = 0,05 \implies z_{\alpha/2} = z_{0,025} = 1,96$
- Costruiamo l'intervallo di confidenza

$$\Pr \left\{ 0,84 - 1,96 \sqrt{\frac{0,84(1-0,84)}{400}} \leq \pi \leq 0,84 + 1,96 \sqrt{\frac{0,84(1-0,84)}{400}} \right\} = 0,95$$

$$\Pr\{0,84 - 1,96 * 0,37 \leq \pi \leq 0,84 + 1,96 * 0,37\} = 0,95$$

$$\Pr\{0,82 \leq \pi \leq 0,86\} = 0,95$$


Introduzione all'inferenza statistica

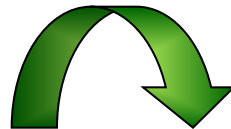
☺ Verifica di ipotesi

$$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi > \pi_0 \end{cases}$$

$$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi < \pi_0 \end{cases}$$

$$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi \neq \pi_0 \end{cases}$$

Vera H_0  $P \approx \text{Norm} \left[\pi_0, \frac{\pi_0(1-\pi_0)}{n} \right]$



Introduzione all'inferenza statistica

 Il test è

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

Dopo aver fissato α le regole di decisione sono quelle già viste per la media di una popolazione.

Introduzione all'inferenza statistica

Esempio:

Un servizio di ambulanze dichiara che il 45% delle sue chiamate riguardano situazioni con pericolo di vita. A fine di verificare se questa dichiarazione sia vera, viene selezionato dagli archivi un campione di 200 chiamate di cui 70 riguardano situazioni con pericolo di vita. Posto un livello di significatività del 5% verifichiamo se è possibile ritenere credibile la dichiarazione.

Soluzione:

- Consideriamo il seguente sistema d'ipotesi: $\begin{cases} H_0: \pi = 0,45 \\ H_1: \pi < 0,45 \end{cases}$
- - Stimiamo dapprima π mediante lo stimatore $P = X/n$

Introduzione all'inferenza statistica

- La stima sarà: $p = \frac{x}{n} = \frac{70}{200} = 0,35$
- Passiamo a costruire il test z:

$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} = \frac{0,35 - 0,45}{\sqrt{\frac{0,45(1 - 0,45)}{200}}} = \frac{-0,15}{0,035} = -4,29$$

- La regola di decisione è:
- $z \geq z_\alpha$ accettiamo l'ipotesi nulla;
- $z < z_\alpha$ rifiutiamo l'ipotesi nulla

Introduzione all'inferenza statistica

- Il valore di z per $\alpha = 0,05$ è $z = 1,64$ pertanto:
- Essendo $z = -4,29 < z_{\alpha = -1,64}$ rifiutiamo l'ipotesi nulla



Quindi non possiamo ritenere vera la dichiarazione del servizio di ambulanze secondo cui le chiamate in oggetto sono pari al 45%.

- Se avessimo scelto $\alpha = 0,01$, z sarebbe stato pari a 2,33 pertanto, **avremmo nuovamente rifiutato l'ipotesi nulla.**



Verifica d'ipotesi sul numero di successi nella popolazione

$$z = \frac{x - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}}$$

Introduzione all'inferenza statistica

Esempio:

Se avessimo voluto risolvere l'esercizio precedente considerando direttamente il numero di successi e non le proporzioni il test sarebbe stato il seguente:

$$z = \frac{x - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}} = \frac{70 - 200 \cdot 0,45}{\sqrt{200 \cdot 0,45(1 - 0,45)}} = \frac{-20}{7,0356} = -2,84$$

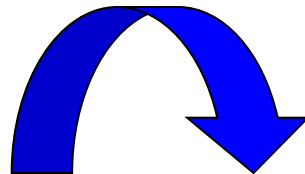
Pertanto, essendo $z_{0,05} = -1,64$ e $z_{0,01} = -2,33$ **avremmo rifiutato l'ipotesi nulla ed avremmo considerata falsa la dichiarazione del servizio di ambulanza.**

Introduzione all'inferenza statistica

⌘ Dimensione campionaria

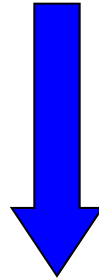
$$\sigma_P^2 = \frac{\pi(1-\pi)}{n} \quad d = z_{\alpha/2} \sigma_P = z_{\alpha/2} \sqrt{\frac{\pi(1-\pi)}{n}}$$

$$n = \frac{z_{\alpha/2}^2 \pi(1-\pi)}{d^2}$$



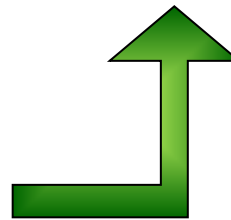
Introduzione all'inferenza statistica

↪ Non conoscendo il valore di π si prende quel valore per cui



$$\pi (1 - \pi) = \textit{massimo}$$

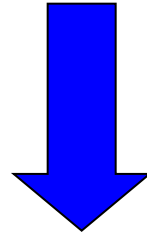
$$\pi = 0,5$$



Introduzione all'inferenza statistica

↩ Più frequentemente si fissa $1-\alpha = 0,95$

$$Z_{\alpha/2} = 1,96$$



$$\max n = \frac{1,96^2 \cdot 0,5^2}{d^2} = \frac{1,96^2}{4d^2}$$

Introduzione all'inferenza statistica


⌘ Verifica d'ipotesi sulla uguaglianza di più distribuzioni

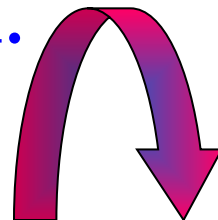
□ t popolazioni classificate secondo le modalità di un carattere Y ;

□ dalle popolazioni estraiamo t campioni casuali classificati secondo le modalità del carattere Y .

Introduzione all'inferenza statistica

 **Dati classificati in una tabella a doppia entrata.**

 **Se la distribuzione di Y non varia al variare di X (gruppi) le frequenze relative di ogni modalità y_i dovrebbero restare invariate al variare delle modalità x_h di X .**



Introduzione all'inferenza statistica

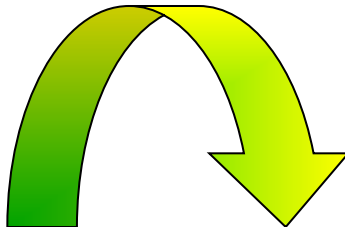
$$\pi_{1/1} = \pi_{1/2} = \cdots = \pi_{1/h} = \cdots = \pi_{1/t} = \cdots = \pi_1$$

.....

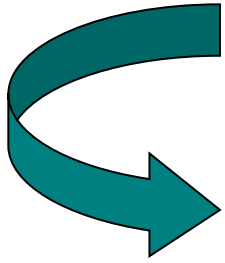
$$\pi_{i/1} = \pi_{s/2} = \cdots = \pi_{i/h} = \cdots = \pi_{i/t} = \cdots = \pi_i$$

.....

$$\pi_{s/1} = \pi_{s/2} = \cdots = \pi_{s/h} = \cdots = \pi_{s/t} = \cdots = \pi_s$$



Introduzione all'inferenza statistica

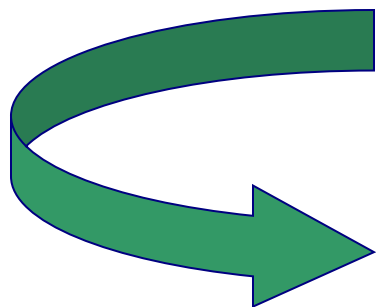


L'ipotesi di uguaglianza delle t distribuzioni Y_h va formulata sulle $\pi_{i|h}$

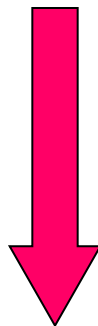


$$\begin{cases} H_0 : \pi_{i|h} = \pi_{i|h'} = \pi_i & \text{per } h = h' = 1, \dots, t; i = 1, \dots, s \\ H_1 : \text{almeno una } \pi_{i|h} \neq \pi_{i|h'} & \text{per } h = h' = 1, \dots, t; i = 1, \dots, s; h \neq h' \end{cases}$$

Introduzione all'inferenza statistica



$\begin{cases} H_0 : \text{indipendenza distributiva di } Y \text{ da } X \\ H_1 : \text{dipendenza distributiva di } Y \text{ da } X \end{cases}$

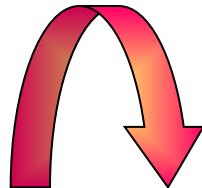


Introduzione all'inferenza statistica

⇒ Non essendo note le π_i^* bisognerà stimarle.

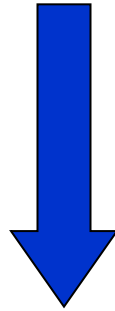
⇒ Vera l'ipotesi nulla, la migliore stima sarà:

$$\hat{\pi}_i^* = \frac{n_{i.}}{n..}$$



Introduzione all'inferenza statistica

➤ Sotto l'ipotesi nulla



$$\hat{n}_{ih}^* = Fr\{Y = y_i \mid X = x_h\} = \hat{\pi}_i^* \cdot n_{.h}$$

Dove \hat{n}_{ih}^* sono le frequenze attese per ogni y_i all'interno di ognuna delle t distribuzioni.

Introduzione all'inferenza statistica

 Se è vera l'ipotesi nulla

$$n_{ih} - \hat{n}_{ih}^* = \text{componente accidentale}$$

 Se l'ipotesi nulla è falsa

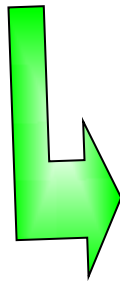
$$n_{ih} - \hat{n}_{ih}^* = \text{comp. acc.} + \text{comp. sist.}$$

Introduzione all'inferenza statistica

➔ Dovremo considerare, quindi, la quantità

$$\sum_h^t \sum_i^s \frac{(n_{ih} - \hat{n}_{ih}^*)^2}{\hat{n}_{ih}^*} \approx \chi^2$$

Il test sarà



$$\chi^2 = \sum_h^t \sum_i^s \frac{(n_{ih} - \hat{n}_{ih}^*)^2}{\hat{n}_{ih}^*}$$

Introduzione all'inferenza statistica

Regole di decisione

$$\chi^2 \leq \chi_{\alpha}^2$$

si accetta H_0

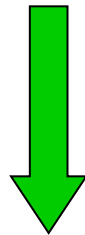
$$\chi^2 > \chi_{\alpha}^2$$

si rifiuta H_0

Introduzione all'inferenza statistica

↘ Gradi di libertà

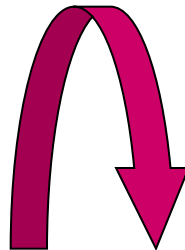
- ✓ t distribuzioni di Y , ognuna con $s-1$ gradi di libertà



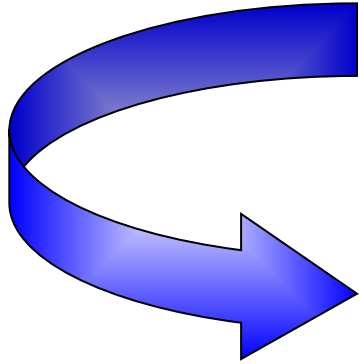
$t(s-1)$ = gradi di libertà relativi a tutte le t distribuzioni

Introduzione all'inferenza statistica

- ✓ Per stimare le π_i^* sono stati imposti $s-1$ vincoli in quanto,
- s sono le modalità di Y ;
 - 1 vincolo è determinato dal fatto che è sufficiente conoscere $s-1$ stime di π_i^* in quanto, essendo in combinazione lineare tra loro, l'ultima può essere determinata per differenza.



Introduzione all'inferenza statistica



$$t(s - 1) - (s - 1) = (t - 1)(s - 1)$$

✎ Per poter procedere nella verifica d'ipotesi è necessario che sia soddisfatta la seguente condizione

$$n_{ih}^* \geq 5 \quad \forall i e h$$

Introduzione all'inferenza statistica

❖ Calcolo abbreviato del χ^2

$$\chi^2 = n \left(\sum_h^t \sum_i^s \frac{n_{ih}^2}{n_{i.} \cdot n_{.h}} - 1 \right)$$

Introduzione all'inferenza statistica

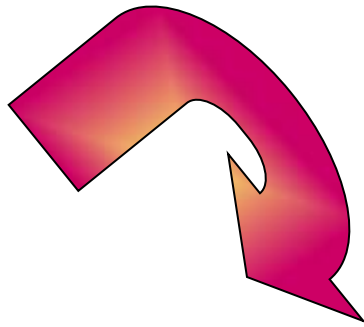


Se si rifiuta l'ipotesi di indipendenza distributiva, è possibile calcolare un indice che ci dà una misura relativa del grado di dipendenza di Y rispetto a X .



Indice di connessione di Y rispetto a X

Introduzione all'inferenza statistica



$$\eta = \frac{\sum_h^t \sum_i^s |n \cdot n_{ih} - n_{i.} \cdot n_{.h}|}{2 \left(n^2 - \sum_i^s n_{i.}^2 \right)}$$

0

1

Introduzione all'inferenza statistica

➤ $\eta = 0$ → **indipendenza**

➤ $\eta = 1$ → **Massima dipendenza o
massima connessione**



**Ad ogni modalità di Y corrisponde
solo una o più di X.**

Introduzione all'inferenza statistica

Esempio:

La direzione di una fabbrica vuole determinare se l'esperienza del dipendente (classificata in tre sottogruppi di anzianità di lavoro: < 2 anni; 2-5 anni; > 5 anni) influenza il numero di unità prodotte in un'ora. I risultati ottenuti su un campione di 150 dipendenti sono stati i seguenti:

Esperienza	Unità prodotte/ora			Totale
	1	2	3	
< 2 anni	10	30	10	50
2 – 5 anni	10	20	20	50
> 5 anni	10	10	30	50
Totale	30	60	60	150

Introduzione all'inferenza statistica

Soluzione:

Il problema di fronte al quale ci troviamo è quello dell'uguaglianza fra più distribuzioni. Infatti se tutte le distribuzioni delle unità prodotte, condizionate agli anni di anzianità, fossero uguali la fabbrica dovrebbe ritenere che l'anzianità non contribuisce ad incrementare la produzione che, forse, dipende da altri fattori. Invece, se anche una sola distribuzione fosse differente la fabbrica potrebbe concludere che l'anzianità lavorativa contribuisce ad incrementare la quantità prodotta. Pertanto il sistema di ipotesi che verrà costruito per verificare questa ipotesi è il seguente:

$$\begin{cases} H_0 : \pi_{i|h} = \pi_{i|h'} = \pi_i & \text{per } h = h' = 1, \dots, t; i = 1, \dots, s \\ H_1 : \text{almeno una } \pi_{i|h} \neq \pi_{i|h'} & \text{per } h = h' = 1, \dots, t; i = 1, \dots, s; h \neq h' \end{cases}$$

Introduzione all'inferenza statistica

Posto $\alpha = 0,05$, consideriamo il seguente test:

$$\chi^2 = \sum_h^t \sum_i^s \frac{(n_{ih} - \hat{n}_{ih}^*)^2}{\hat{n}_{ih}^*}$$

Che calcoleremo mediante la formula abbreviata:

$$\chi^2 = n \left(\sum_h^t \sum_i^s \frac{n_{ih}^2}{n_{i.} \cdot n_{.h}} - 1 \right)$$

Introduzione all'inferenza statistica

$n_{ih}^2/n_{i.} \cdot n_{.h}$	Unità prodotte/ora			Totale
	1	2	3	
< 2 anni	0,0667	0,3000	0,0333	
2 – 5 anni	0,0667	0,1333	0,1333	
> 5 anni	0,0667	0,0333	0,3000	
Totale				1,1333

$$\chi^2 = 150(1,1333 - 1) \cong 20,0$$

Avendo posto $\alpha = 0,05$ e considerato che i gradi di libertà sono:

$$(s-1)(t-1) = (3 - 1)(3 - 1) = 4 \quad \longrightarrow \quad \chi_{4;0,05}^2 = 9,49$$

Essendo il valore del test chi-quadro maggiore del valore critico rilevato sulle tavole possiamo concludere che, effettivamente, le distribuzioni non sono tutte uguali, pertanto **l'anzianità influisce sul numero di unità prodotte/ora.**

Introduzione all'inferenza statistica



Associazione fra due caratteri



Test non parametrico su tabelle di contingenza.



Campione di n unità classificato sulla base delle modalità di due caratteri.

Introduzione all'inferenza statistica

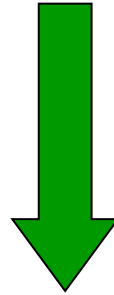
- Se a determinate modalità di un carattere corrispondono più frequentemente alcune modalità dell'altro carattere e viceversa, i due caratteri sono associati



$$\pi_{ih} = \Pr\{Y = y_i \cap X = x_h\}$$

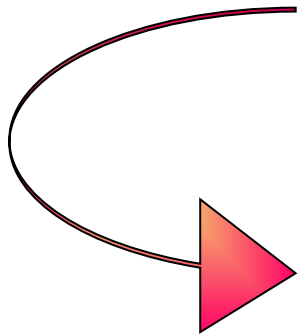
Introduzione all'inferenza statistica

- Se le distribuzioni di un carattere, condizionate alle modalità dell'altro carattere, restano invariate e viceversa.



caratteri non associati

Introduzione all'inferenza statistica



indipendenza

$$\pi_{i/h} = \pi_i \quad \forall \quad h = 1, \dots, t$$

$$\pi_{h/i} = \pi_h \quad \forall \quad i = 1, \dots, s$$

oppure $\pi_{ih} = \pi_i \cdot \pi_h$

Introduzione all'inferenza statistica

➔ Il problema si traduce in una verifica d'ipotesi che ha come ipotesi nulla l'indipendenza fra i due caratteri e come ipotesi alternativa la loro associazione

$$\begin{cases} H_0 : \pi_{ih} = {}_0\pi_{ih} & ({}_0\pi_{ih} = \pi_i \cdot \pi_h) \\ H_1 : \pi_{ih} \neq {}_0\pi_{ih} & \text{per almeno una } i \text{ o una } h; i=1, \dots, s; h=1, \dots, t \end{cases}$$

Introduzione all'inferenza statistica



$$\hat{\pi}_i = \frac{n_{i.}}{n_{..}} \quad \hat{\pi}_h = \frac{n_{.h}}{n_{..}} \quad \hat{\pi}_{ih} = \hat{\pi}_i \cdot \hat{\pi}_h = \frac{n_{i.} \cdot n_{.h}}{n_{..}^2}$$

➤ Le frequenze attese, sotto l'ipotesi nulla

$$\hat{n}_{ih}^* = \hat{\pi}_{ih} \cdot n_{..} = \frac{n_{i.} \cdot n_{.h}}{n_{..}^2} n_{..} = \frac{n_{i.} \cdot n_{.h}}{n_{..}}$$

Introduzione all'inferenza statistica

😊 **Test**

$$\chi^2 = \sum_h^t \sum_i^s \frac{(n_{ih} - \hat{n}_{ih}^*)^2}{\hat{n}_{ih}}$$

✓ **Regola di decisione**

$\chi^2 \leq \chi_\alpha^2$ **si accetta l'ipotesi nulla**

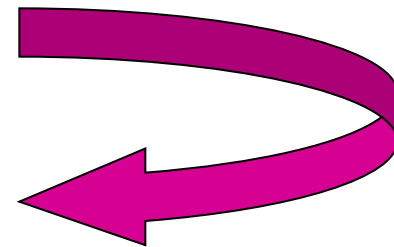
$\chi^2 > \chi_\alpha^2$ **si rifiuta l'ipotesi nulla**

Introduzione all'inferenza statistica

⇒ Gradi di libertà

✓ Per calcolare le frequenze attese sono state stimate

$$\hat{\pi}_i \quad \text{e} \quad \hat{\pi}_h$$

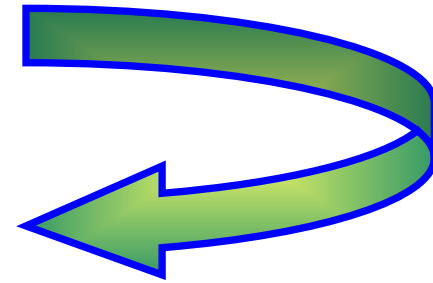


I vincoli sono $k = (s-1) + (t-1)$

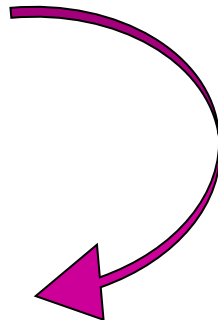
Introduzione all'inferenza statistica

- ✓ le distribuzioni congiunte sono st

$$g = st - 1 - k = (s-1)(t-1)$$



👋 Se si rifiuta l'ipotesi nulla è possibile calcolare il grado di associazione fra i due caratteri



Introduzione all'inferenza statistica

↪ Una misura relativa è

$$A = \sqrt{\frac{\chi^2}{n \cdot \min(s-1, t-1)}}$$

Il denominatore rappresenta il massimo valore che può essere assunto dal χ^2 , data una tabella di contingenza di dimensione st e un campione di ampiezza n .

Introduzione all'inferenza statistica

Esempio:

Un'agenzia di viaggi seleziona un campione casuale di individui tra la sua clientela e chiede loro se si sono rivolti ad un'agenzia di viaggi per prenotare l'ultimo biglietto aereo acquistato. Al termine della rilevazione i risultati vengono classificati per fasce di età in quanto l'agenzia è interessata a conoscere se esiste associazione fra l'età degli intervistati e la loro scelta. I risultati ottenuti sono:

Età (anni)	Ha usato un'agenzia per prenotare l'ultimo biglietto aereo?		Totale
	Si	No	
Meno di 30	15	30	45
30 - 39	20	42	62
40 - 49	47	42	89
50 - 59	36	50	86
60 e più	45	20	65
Totale	163	184	347

Introduzione all'inferenza statistica

Soluzione:

■ Il problema di fronte al quale ci troviamo è quello in cui il campione estratto è stato classificato in base a due caratteri e si vuole verificare se a determinate modalità di un carattere corrispondono più frequentemente alcune modalità dell'altro carattere e viceversa. Il problema si traduce in una verifica d'ipotesi che ha come ipotesi nulla l'**indipendenza** fra i due caratteri e come ipotesi alternativa la loro **associazione**. Pertanto il sistema di ipotesi che verrà costruito per verificare questa ipotesi è il seguente:

$$\begin{cases} H_0 : \pi_{ih} = {}_0\pi_{ih} & ({}_0\pi_{ih} = \pi_i \cdot \pi_h) \\ H_1 : \pi_{ih} \neq {}_0\pi_{ih} & \text{per almeno una } i \text{ o una } h; i=1, \dots, s; h=1, \dots, t \end{cases}$$

Introduzione all'inferenza statistica

Posto $\alpha = 0,05$, consideriamo il seguente test:

$$\chi^2 = \sum_h^t \sum_i^s \frac{(n_{ih} - \hat{n}_{ih}^*)^2}{\hat{n}_{ih}^*}$$

Che calcoleremo mediante la formula abbreviata:

$$\chi^2 = n \left(\sum_h^t \sum_i^s \frac{n_{ih}^2}{n_{i.} \cdot n_{.h}} - 1 \right)$$

Introduzione all'inferenza statistica

$n_{ih}^2 / n_{i.} \cdot n_{.h}$	Ha usato un'agenzia per prenotare l'ultimo biglietto aereo?		Totale
	Si	No	
Meno di 30	0,0307	0,1087	
30 - 39	0,0396	0,1546	
40 - 49	0,1523	0,1077	
50 - 59	0,0925	0,1580	
60 e più	0,1911	0,0334	
Totale			1,0686

$$\chi^2 = 347(1,0686 - 1) = 23,80$$

Avendo posto $\alpha = 0,05$ e considerato che i gradi di libertà sono:

$$(s-1)(t-1) = (5-1)(2-1) = 4 \quad \longrightarrow \quad \chi_{4;0,05}^2 = 9,49$$



Introduzione all'inferenza statistica

Essendo il valore del test chi-quadro maggiore del valore critico rilevato sulle tavole possiamo concludere che **non vi è indipendenza** fra l'età e la scelta dell'agenzia per prenotare il biglietto aereo ma, al contrario, **i due caratteri sono associati**. Per completezza, calcoliamo la misura relativa per vedere quanto è forte l'associazione fra i due caratteri (**solo a fini descrittivi!**)

$$A = \sqrt{\frac{\chi^2}{n \cdot \min(s-1, t-1)}} = \sqrt{\frac{23,80}{347 \cdot \min(5-1, 2-1)}} = \sqrt{\frac{23,80}{347 \cdot 1}} = \sqrt{0,0686} = 0,26 \cong 26\%$$



L'intensità del legame tra i due caratteri è solo del 26% circa

Considerazioni finali

- ❑ L'inferenza statistica è un insieme di tecniche particolarmente importanti in ambito sperimentale.
- ❑ Vengono utilizzate in tutti quei campi applicativi in cui si opera con dei campioni. Ad esempio nella sperimentazione di un farmaco, nelle analisi di mercato, ecc.
- ❑ Viene utilizzata in tutte le situazioni in cui si deve decidere in condizioni di incertezza!
- ❑ Abbiamo fatto solo un breve cenno a quello che è l'inferenza statistica.

Introduzione all'inferenza statistica

Considerazioni finali

- ❑ Infatti ci sono tanti argomenti che non è stato possibile considerare. Ad esempio l'inferenza sulla varianza di una popolazione normale, su altri parametri di una variabile normale, sui parametri di una variabile normale doppia, sui parametri di un modello dicotomico nel caso di piccoli campioni, di quella non parametrica si è fatto solo un breve cenno, ecc.
- ❑ Quello che va ricordato, comunque, è che nella vita di tutti i giorni, allorché siamo chiamati a scegliere in situazioni di incertezza, facciamo «*inferenza*»!
- ❑ *Facciamo STATISTICA!!!!*

