

## POPOLAZIONE

Insieme di unità statistiche uguali rispetto ad un certo carattere.

## CAMPIONE

un qualunque sottoinsieme della popolazione contenente un certo numero di unità statistiche. Deve assomigliare il più possibile alla popolazione.

## CAMPIONE CASUALE

la procedura per ottenere un campione rappresentativo della popolazione è l'estrazione casuale delle unità.

	Popolazione	
	Finita	Infinita
Schema con reinserimento	<b>A</b>	<b>B</b>
Schema senza reinserimento	<b>C</b>	<b>D</b>

"Il campione casuale è un insieme di variabili casuali.  $(X_1, X_2, \dots, X_n)$  la cui funzione di densità dipende da quella della popolazione. La sua distribuzione di probabilità è definita come la distribuzione di probabilità o densità congiunta della n-pla di variabili casuali che compongono il campione casuale."

## CAMPIONE CASUALE SEMPLICE

-CON REPLICAZIONE da popolazione  $X \sim f(x, \theta)$  se le n variabili casuali sono IID (indipendenti, identicamente distribuite)

-SENZA REPLICAZIONE a popolazione  $X \sim f(x, \theta)$  se le n variabili casuali sono ID (identicamente distribuite)

## POP. FINITA

un insieme di unità realmente esistenti che possono essere oggetto di rilevazione totale oppure campionaria.

## POP. INFINITA

insieme potenziale delle osservazioni connesse alla ripetizione teoricamente illimitata di un esperimento casuale condotto nelle stesse condizioni.

## MODELLO DESCRITTIVO

è un modello matematico che esprime la probabilità o densità della v.c. che descrive l'esito della singola prova di un esperimento.

## SPAZIO CAMPIONARIO

consideriamo il campione casuale composto da n variabili IID, denominiamo **campione osservabile** una specifica realizzazione del campione casuale (ossia una n-pla indicata con  $X_1, \dots, X_n$ ); tutti i possibili campioni osservabili costituiscono lo spazio campionario.

## DISTRIBUZIONE CAMPIONARIA DELLA MEDIA

è la distribuzione delle medie campionarie ottenuti da tutti i possibili campioni della stessa ampiezza estratti dalla popolazione.

[ ☆ costruire spazio campionario

☆ attribuire la probabilità ad ogni campione

☆ su ogni campione si calcola la media campionaria

☆ costruire la distribuzione della media campionaria ]

$$\bar{X} = (1/n \cdot x_1) + \dots + (1/n \cdot x_n) = 1/n \cdot \sum X_i$$

**valore atteso**  $\rightarrow E(X) = \mu$  . **varianza**  $\rightarrow \sigma_x = \sigma/\sqrt{n}$  oppure  $\sigma_x = \sigma^2/n$   
 con fattore di correzione per pop. finite  
 $\sigma_{\bar{x}} = \sigma/\sqrt{n} \cdot \sqrt{(N-n)/(N-1)}$  oppure  $\sigma_{\bar{x}} = \sigma^2/n \cdot (N-n)/(N-1)$

**DISTRIBUZIONE CAMPIONARIA DELLA PROPORZIONE**

in una popolazione dicotomica un possibile campione è una successione di 0 e 1, la media applicata a queste successioni definisce la proporzione di successi sul campione. (Cioè è una media campionaria quando la popolazione è Bernulliana)

► CASO IID

$\hat{p} = X/n \rightarrow$  dove  $x$  è una v.c. e  $n$  è la dimensione del campione

$np = X \sim$  **binomiale**

$\rightarrow E(\hat{p}) = \mu p$

$\rightarrow \text{var}(\hat{p}) = (p \cdot q)/n$

► CASO ID

$\hat{p} = x/n \rightarrow$  dove  $x$  è una v.c. ipergeometrica e  $n$  è la dimensione del campione

$np = X \sim$  **ipergeometrica**

$\rightarrow E(\bar{X}) = p$

$\rightarrow \text{var}(\bar{x}) = (p \cdot q)/n \cdot (N-n)/(N-1)$

**TEOREMA DEL LIMITE CENTRALE**

sia  $[x_1, \dots, X_n]$  una successione di variabili casuali IID con  $E(x_i) = \mu$  e  $\text{var}(x_i) = \sigma^2$ .

sia  $[C_n = \sum \bar{X}_i/n]$  la successione delle medie aritmetiche fatte sulle prime  $n$  variabili casuali in cui  $E(\bar{X}_n) = \mu$  e  $\text{var}(\bar{X}_n) = \sigma^2/n$ .

Definita la v.c. standard  $Z_n = (\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ , per ogni numero reale  $z$  val la relazione:

$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$

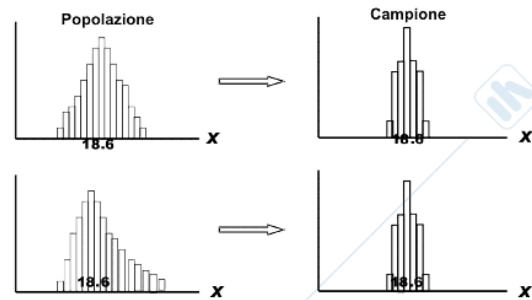
$n \rightarrow \infty$

dove  $\Phi()$  è la funzione di ripartizione della Normale standard.

[IN BREVE: al tendere di  $n$  all' infinito la funzione di ripartizione della v.c.  $Z_n$  tende alla normale standard a prescindere dal tipo di distribuzione di probabilità delle v.c. della successione, unico vincolo è che siano IID.]

RIEPILOGO:  $\frac{X_i - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$  .  $\frac{\hat{p} - p}{\sqrt{(pq)/n}} \sim N(0,1)$  .  $\frac{X - np}{\sqrt{npq}} \sim N(0,1)$

Anche nel caso di popolazione finita, se  $n > 30$  approssimo alla normale  $\rightarrow \frac{X_i - \mu}{(\sigma/\sqrt{n}) \cdot \sqrt{(N-n)/(N-1)}} \sim N(0,1)$



**DISTRIBUZIONE CAMPIONARIA DELLA VARIANZA**

► CASO IID

Se la media della popolazione  $E(X) = \mu$  è conosciuta, la varianza campionaria è data da:

$\hat{\sigma}^2 = 1/n \sum (X_i - \mu)^2 \rightarrow$  statistica corretta

Se la media della popolazione  $E(x) = \mu$  è sconosciuta, utilizzo la media campionaria e quindi la varianza campionaria è data da:

$S^2 = 1/(n-1) \sum (X_i - \bar{X})^2 \rightarrow$  statistica non corretta.

Per definizione può anche essere calcolata  $\tilde{S}^2 = 1/n \sum (X_i - \bar{X})^2 \rightarrow$  statistica corretta

Se la popolazione è Normale, la distribuzione delle tre varianze è:

$$\hat{\sigma}^2 \rightarrow \frac{n \cdot \sigma^2}{\sigma^2} \sim \chi^2(n) \quad S^2 \rightarrow \frac{n \cdot S^2}{\sigma^2} \sim \chi^2(n-1) \quad \check{S}^2 \rightarrow \frac{(n-1) \check{S}^2}{\sigma^2} \sim \chi^2(n-1)$$

$$E(\hat{\sigma}^2) = \sigma^2$$

$$E(S^2) = (n-1)/n \cdot \sigma^2$$

$$E(\check{S}^2) = \sigma^2$$

$$\text{Var}(\hat{\sigma}^2) = 2/n \cdot \sigma^4$$

$$\text{Var}(S^2) = 2(n-1)/n^2 \cdot \sigma^4$$

$$\text{Var}(\check{S}^2) = 2/(n-1) \cdot \sigma^4$$

► CASO ID

$$E(\hat{\sigma}^2) = \sigma^2$$

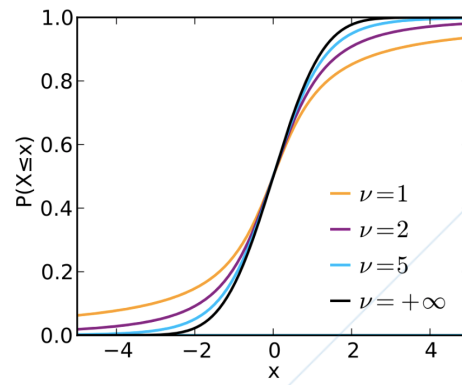
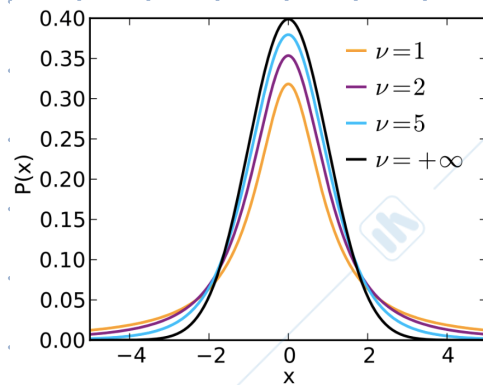
$$E(S^2) = ((n-1)/n)(N/N-1) \cdot \sigma^2$$

$$E(\check{S}^2) = (N/N-1)\sigma^2$$

**VARIABILE CASUALE T-STUDENT**

È una distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie indipendenti, la prima con distribuzione normale ( $Z \sim N(0,1)$ ) e la seconda, segue una distribuzione chi quadrato ( $U \sim \chi^2(r)$ ). la t-student ( $W$ ) con r GDL è definita da  $W = Z/\sqrt{U/r}$ . La incontriamo per la distribuzione campionaria della media quando non è nota la varianza e quindi non possiamo utilizzare la Normale.

Quindi in assenza della varianza della popolazione si utilizza la varianza campionaria  $S^2$ , allora la standardizzata della media campionaria è una v.c. t-student con le proprietà:  $\mu=0$  per  $r>1$  e  $S^2 = r/(r-2)$  per  $r>2$ , è simmetrica rispetto a 0 e tende alla Normale se  $r \rightarrow \infty$



**IN PRATICA:**

★ Se la popolazione è Normale e il campione IID, la v.c. media campionaria ha distribuzione  $N(\mu; \sigma^2/n)$ , pertanto la standardizzata è  $Z_{\bar{X}} = (\bar{X} - \mu) / (\sigma/\sqrt{n})$  ha distribuzione  $N(0,1)$ .

★ Se la popolazione non è Normale, la v.c. media campionaria ha distribuzione approssimativamente come la Normale purché n sia grande.

★ Se la varianza della popolazione non è nota e al posto di  $\sigma^2$  (in  $Z_{\bar{X}}$ ) xi metto la varianza campionaria ( $S^2$ ) ottengo allora una t-student.

Quindi i 3 casi sono:

$$\hat{\sigma}^2 \rightarrow W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim t(n).$$

$$S^2 \rightarrow W = \frac{\bar{X} - \mu}{S/\sqrt{n-1}} \sim t(n-1).$$

$$\check{S}^2 \rightarrow W = \frac{\bar{X} - \mu}{\check{S}/\sqrt{n}} \sim t(n-1).$$

Sia una popolazione Normale con media  $\mu$  incognita, siano  $\bar{X}$  e  $S^2$  media e varianza di un campione di ampiezza n, allora il rapporto  $T = (\bar{X} - \mu) / (S/\sqrt{n})$  è una v.c. t-student con n-1 GDL.

Nel caso in cui la popolazione di partenza non sia normale e abbia media  $\mu$ , allora estraggo media ( $\bar{X}$ ) e varianza ( $S^2$ ) di un campione casuale e se n è sufficientemente grande ( $\geq 40$ ) allora  $Z_{\bar{X}} = (\bar{X} - \mu) / (S/\sqrt{n})$  è una v.c. approssimata alla Normale standard (0,1)

## STIMA PUNTUALE DEI PARAMETRI STIMATORE

Dato un campione casuale  $(X_1, \dots, X_n)$ , per stimatore di un parametro  $\theta$  si intende la statistica campionaria  $T=t(X_1, \dots, X_n)$  utilizzata per stimare  $\theta$ .

### STIMA

è invece la singola determinazione dello stimatore, il valore,  $t=t(X_1, \dots, X_n)$ , che esso assume nel campione osservato inteso come n-pla di numeri effettivamente osservati.

### STIMA PUNTUALE

stimatore più conveniente per attribuire un valore al parametro di interesse  $\theta$ .

### PROPRIETÀ STIMATORE

per stabilire se lo stimatore è adeguato è cruciale studiare la variabile casuale ERRORE DI STIMA  $T-\theta$ .

★ la media degli errori di stima  $E(T-\theta)$  è preferibile sia nulla, se  $E(T-\theta)=0$  o  $E(T)=\theta$  lo stimatore è non distorto. La differenza  $D(T)=E(T)-\theta$  è detta distorsione. La **proprietà della non distorsione** non desidera né che  $E(T)<\theta$  né  $E(T)>\theta$  poiché il risultato produrrebbe stime mediamente al di sotto oppure al di sopra del parametro.

Sappiamo che questa proprietà vale per  $E(\bar{X})=\mu$ ,  $E(\hat{p})=p$ ,  $E(\hat{S}^2)=\sigma^2$

★ la media dei quadrati degli errori di stima  $MSE(T)=E(T-\theta)^2$  è preferibile più piccola possibile ed è anche detta ERRORE QUADRATICO MEDIO ed è la misura della distanza delle stime dal parametro incognito.

$$MSE(T)=\text{Var}(T)+[E(T)-\theta]^2$$

Se lo stimatore è non distorto, allora  $MSE(T)=\text{Var}(T)$ .

$$\text{Sappiamo quindi che } MSE(\bar{X})=\text{Var}(\bar{X})=\frac{\sigma^2}{n}, \quad MSE(\hat{p})=\text{Var}(\hat{p})=\frac{pq}{n}, \quad MSE(\hat{S}^2)=\frac{2\sigma^4}{n-1}$$

### EFFICIENZA STIMATORE

è una proprietà che si manifesta nel confronto di più stimatori impiegabili per stimare uno stesso parametro  $\theta$ .

Dati due stimatori di  $\theta$ ,  $T_1$  e  $T_2$ , si dice che  $T_1$  è più efficiente di  $T_2$  se

$$MSE(T_1)=E(T_1-\theta)^2 \leq MSE(T_2)=E(T_2-\theta)^2.$$

★ la media campionaria è uno stimatore blu: si definisca un qualunque stimatore lineare  $T=\sum[(1/n)+\delta_i]X_i$  ed è corretto se  $\sum \delta_i=0$  e la varianza di  $T$  è  $\sigma^2 \sum [(1/n)+\delta_i]^2$

### PROPRIETÀ ASINTOTICHE DEGLI STIMATORI

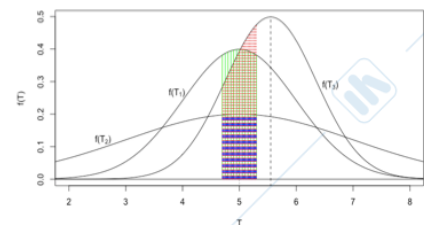
★ **Stimatore asintoticamente non distorto:** uno stimatore è detto asintoticamente non distorto se la sua distorsione è prossima allo 0 quando la dimensione del campione è molto grande  $\rightarrow D(T_n)=E(T_n)-\theta$

★ **Consistenza:** uno stimatore si dice consistente se vi è certezza che esso assuma un valore prossimo a  $\theta$  quando la dimensione del campione è elevata  $\rightarrow \lim P(|T_n-\theta|<\epsilon)=1$ .  
La distribuzione tende graficamente a concentrarsi attorno allo 0 laddove la differenza  $T_n-\theta$  tende a 0.

quindi abbiamo che:  $\lim P(|\bar{X}_n-\mu|<\epsilon)=1 \rightarrow$  lo stimatore  $T_n$   
 $\lim P(|\hat{p}_n-p|<\epsilon)=1$  è consistente

se  $\lim E(T_n)=\theta$  e  $\lim \text{Var}(T_n)=0$

$$\lim P(|S^2_n-\sigma^2|<\epsilon)=1$$



## STIMA PER INTERVALLO

Sia  $(X_1, \dots, X_n)$  un campione casuale proveniente da una popolazione della quale interessa stimare il parametro  $\Theta$ . Siano  $L_1=l_1(X_1, \dots, X_n)$  e  $L_2=l_2(X_1, \dots, X_n)$  con  $L_1 < L_2$  due statistiche campionarie tali che  $P(L_1 < \Theta < L_2) = 1 - \alpha$  dove  $\alpha$  è un numero molto piccolo che non dipende da  $\Theta$ .

★ l'intervallo  $[l_1, l_2]$  è detto intervallo di confidenza a livello  $100(1 - \alpha)\%$  per il parametro  $\Theta$ , dove  $(1 - \alpha)$  è definita livello di confidenza dell'intervallo dove  $\alpha$  è compreso tra 0 e 1.

L'intervallo che costruiamo attorno ad ogni campione facciamo in modo che il 95% contenga il parametro incognito e il 5% no, poi si estrae uno di questi intervalli a caso e abbiamo il 95% di probabilità che contenga  $\Theta$  e anche se non so se lo contenga veramente o meno, io sono confidente che ci sia.

★  $l_1$  e  $l_2$  sono detti rispettivamente Limite inferiore e Limite superiore dell'intervallo di confidenza. L'ampiezza o lunghezza dell'intervallo è data da  $A = l_2 - l_1$   
 [Formula generale intervallo di confidenza -> stima puntuale + - (fattore • errore standard)]

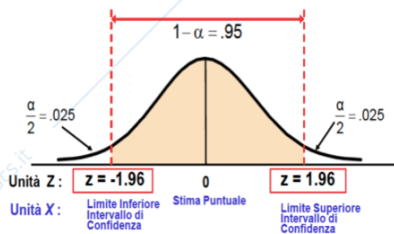
## STIMA PER INTERVALLO DELLA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA NOTA

Dato un campione casuale da una popolazione normale  $N(\mu, \sigma^2)$  e fissato un livello di probabilità molto alto, si ha che uno stimatore per intervallo di  $\mu$  ha estremi dati da:

$$L_1 = \bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{e} \quad L_2 = \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

dove  $Z_{\alpha/2}$  è il fattore di affidabilità, cioè il valore della normale standard che lascia in ciascuna coda  $\alpha/2$

Ad esempio se  $(1 - \alpha) = 0.95$  allora:



GENERALMENTE:

liv. confidenza	$1 - \alpha$	$Z_{\alpha/2}$
90%	0.90	1.645
95%	0.95	1.96
99%	0.99	2,58

### DIMOSTRAZIONE

poiché la variabile casuale:  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$  ha distribuzione normale segue che:

$$1 - \alpha = P\left(-Z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right) \rightarrow P\left(\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

FUNZIONE CARDINE O PIVOT - dipende dalle osservazioni campionarie da  $\Theta$ , la sua distribuzione è nota e indipendente dal  $\Theta$  ed è continua e monotona rispetto a  $\Theta$

### MARGINE DI ERRORE (ME)

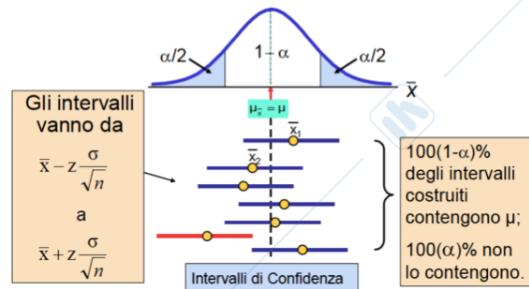
L'intervallo di confidenza può anche essere scritto come  $\bar{X} \pm ME$  dove  $ME = Z_{\alpha/2} \cdot \sigma / \sqrt{n}$

L'ME si riduce se la deviazione standard diminuisce ( $\sigma \downarrow$ ), l'ampiezza del campione aumenta ( $n \uparrow$ ), livello di confidenza diminuisce ( $1 - \alpha \downarrow$ ). L'ampiezza d'intervallo ( $l_2 - l_1$ ) è due volte l'ME.

### ★ Caso ID (della stima per intervallo della media)

se il campione è estratto senza ripetizione da una popolazione finita di dimensione  $N$ , la varianza della media è:  $\text{Var}(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$  e l'intervallo è dato:  $C = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sqrt{N-n} \cdot \sigma^2}{\sqrt{N-1} \cdot \sqrt{n}}$

### INTERVALLO DI CONFIDENZA DELLA DISTRIBUZIONE DELLA MEDIA CAMPIONARIA



### STIMA PER INTERVALLO DELLA PROPORZIONE NEL CASO DI GRANDI CAMPIONI

in questo caso la statistica da usare è:  $Z\hat{p} = \frac{\hat{p} - p}{\sqrt{(\hat{p} \cdot \hat{q})/n}}$  dove  $\hat{p}$ , stimatore di  $p$  è la v.c. proporzione.

Sia  $(X_1, \dots, X_n)$  un campione casuale proveniente da una popolazione Bernoulliana, allora, fissato livello di probabilità  $\alpha$ , si trova che uno stimatore per intervallo di  $p$  ha estremi dati da:  $L_1 = \hat{p} - Z_{\alpha/2} \cdot \sqrt{(\hat{p} \cdot \hat{q})/n}$  e  $L_2 = \hat{p} + Z_{\alpha/2} \cdot \sqrt{(\hat{p} \cdot \hat{q})/n}$

dove  $Z_{\alpha/2}$  è il valore della distribuzione normale standard che lascia  $\alpha/2$  su ogni coda.

### STIMA PER INTERVALLO DELLA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA NON NOTA

Sia  $(X_1, \dots, X_n)$  un campione casuale proveniente da una popolazione Normale  $N(\mu, \sigma)$  allora, fissato livello di probabilità  $\alpha$ , si trova che uno stimatore per intervallo di  $\mu$  ha gli estremi dati da:  $L_1 = \bar{X} - t_{\alpha/2} \cdot \frac{\tilde{S}}{\sqrt{n}}$  e  $L_2 = \bar{X} + t_{\alpha/2} \cdot \frac{\tilde{S}}{\sqrt{n}}$

dove  $t_{\alpha/2}$  è il valore della  $t$  di student con  $n-1$  gradi di libertà che lascia in ciascuna coda una probabilità di  $\alpha/2$

• FUNZIONE CARDINE  $T_{\bar{X}} = \frac{\bar{X} - \mu}{\tilde{S}/\sqrt{n}}$

un intervallo di confidenza con coefficiente  $1-\alpha$  ha estremi:

$$l_1 = \bar{x} - t_{\alpha/2} \cdot \frac{\tilde{s}}{\sqrt{n}} \quad \text{e} \quad l_2 = \bar{x} + t_{\alpha/2} \cdot \frac{\tilde{s}}{\sqrt{n}} \quad \text{dove } \bar{x} \text{ e } \tilde{s} \text{ sono i valori osservati di } \bar{X} \text{ e } \tilde{S}$$

### STIMA PER INTERVALLO DELLA MEDIA NEL CASO DI CAMPIONI GRANDI

Se la dimensione del campione è sufficientemente grande possiamo rifarci alla statistica  $Z_{\bar{X}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$  la cui distribuzione si approssima alla normale standard qualunque sia la

popolazione. L'intervallo avrà come estremi:  $L_1 = \bar{X} - z_{\alpha/2} \cdot \frac{\tilde{S}}{\sqrt{n}}$  e  $L_2 = \bar{X} + z_{\alpha/2} \cdot \frac{\tilde{S}}{\sqrt{n}}$

che sono variabili casuali, e l'intervallo racchiude al suo interno il parametro  $\mu$  con

probabilità  $1-\alpha$ . Da qui gli estremi dell'intervallo:  $l_1 = \bar{x} - z_{\alpha/2} \cdot \frac{\tilde{s}}{\sqrt{n}}$  e  $l_2 = \bar{x} + z_{\alpha/2} \cdot \frac{\tilde{s}}{\sqrt{n}}$

### INTERVALLO DI CONFIDENZA PER LA MEDIA

genericamente l'intervallo di confidenza è uguale campione  $\pm$  ME, cioè il margine di errore

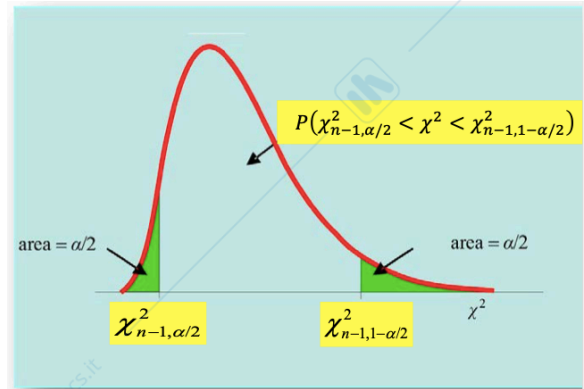
• determinazione ampiezza campionaria per la media

$$\bar{X} \pm ME = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{risolvere rispetto ad } n = \frac{Z_{\alpha/2}^2 \cdot \sigma^2}{ME^2}$$

• determinazione ampiezza campionaria per la proporzione  
 $\hat{p} \pm ME = Z_{\alpha/2} \cdot \frac{\sqrt{\hat{p} \cdot \hat{q}}}{\sqrt{n}} \rightarrow \hat{p} \cdot \hat{q} \text{ non può essere } > 0.25 \text{ quando } \hat{p} = 0.5$   
 sostituire 0.25 in  $\hat{p} \cdot \hat{q}$  e risolvere per  $n = \frac{0.25 \cdot Z^2 \alpha/2}{ME^2}$

### INTERVALLO DI CONFIDENZA PER LA VARIANZA

• ipotesi normalità della popolazione.  
 Se la popolazione è normale, sappiamo che la v.c.  $V = (n-1)S^2$  ha distribuzione chi-quadrato con n-1 gdl. Indichiamo con  $\chi^2_{n-1, \alpha/2}$  e  $\chi^2_{n-1, 1-\alpha/2}$  i quantili di livello  $\alpha/2$  e  $1-\alpha/2$  della v.c.  $\chi^2$  con n-1 gdl.  
 • Sia  $(X_1, \dots, X_n)$  un campione casuale proveniente da una popolazione Normale  $N(\mu, \sigma^2)$ .  
 •  $1-\alpha$  livello confidenza  $\rightarrow \chi^2_{n-1, \alpha/2}; \chi^2_{n-1, 1-\alpha/2}$



$$\left[ \frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}} \right]$$

### VERIFICA D'IPOTESI

• Sia  $\Theta$  il parametro d'interesse, ossia la costante caratteristica della popolazione oggetto di studio. L'ipotesi statistica è un'affermazione che riguarda il parametro.  
 • L'ipotesi sottoposta a verifica va sotto il nome di ipotesi nulla ( $H_0$ ), mentre è chiamata ipotesi alternativa ( $H_1$ ) l'affermazione contrapposta.

Quindi:

con riferimento al parametro  $\Theta$  supponiamo che sono  $H_0$  e  $H_1$  le due ipotesi (nulla e alternativa), con la verifica delle ipotesi si decide se rifiutare o no l'ipotesi nulla sulla base di una funzione di dati del campione casuale detta:

STATISTICA TEST  $\rightarrow$  [l'insieme di tutti i possibili valori che possono essere assunti dalla statistica test è detto spazio campionario, tale spazio viene diviso in due parti dette : regione di accettazione e regione di rifiuto].

• Regola decisionale:

bipartito o spazio parametrico ( $H_0$  e  $H_1$ ), bipartito lo spazio campionario, si accetta o rifiuta l'ipotesi nulla se la statistica test cade nella regione di accettazione o rifiuto.

Possibili Risultati Verifica di Ipotesi		
Decisione	Stato di Natura	
	$H_0$ Vera	$H_0$ Falsa
Non Rifiutare $H_0$	Decisione corretta (1 - $\alpha$ )	Errore di Secondo Tipo ( $\beta$ )
Rifiutare $H_0$	Errore di Primo Tipo ( $\alpha$ )	Decisione corretta (1 - $\beta$ )

• errore di primo tipo = rifiuto  $H_0$  quando vera  
 • errore secondo tipo = accetto  $H_0$  quando falsa  
 Teoria dei test in generale:  
 possiamo dire che più o meno tutti i procedimenti di verifica seguono il seguente criterio: si assegna ad  $\alpha$  un valore piccolo e si individua la regola di decisione che minimizza  $\beta$ , quindi introduco il test d'ipotesi per alcuni parametri di rilievo (medie e varianze) con approccio intuitivo.

### VERIFICA D'IPOTESI SULLA MEDIA DI UNA POPOLAZIONE CON VARIANZA NOTA

Sia  $H_0: \mu = \mu_0$

L'ipotesi alternativa contrapposta può assumere una delle tre configurazioni:

$H_1 : \mu < \mu_0$   
 $H_1 : \mu > \mu_0$

ipotesi alternativa unidirezionale

$H_1 : \mu \neq \mu_0$

ipotesi alternativa bidirezionale

★CASO  $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

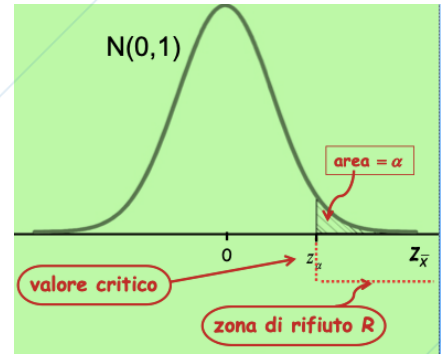
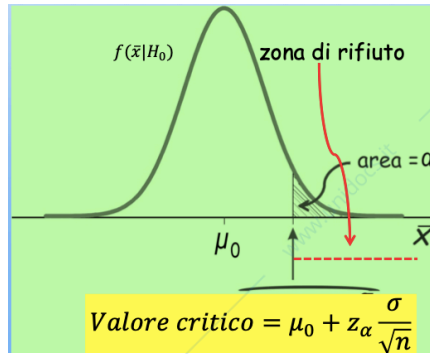
per la verifica di ipotesi in questione chiameremo statistica test  $Z_{\bar{x}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

La disuguaglianza che determina il rifiuto di  $H_0$  è data da:  $\bar{x} > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

regione di rifiuto ->

$R = \{x_1, x_2, \dots, x_n; \bar{x} > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\}$

Si rifiuta  $H_0$  per valori di  $\bar{X}$  molto più grandi di  $\mu_0$ , valori che, sotto  $H_0$ , sono stime poco «credibili» di  $\mu_0$ .



La disuguaglianza che determina il rifiuto di  $H_0$ , può essere, ovviamente, esplicitata in termini di valori standardizzati nel modo seguente:  $Z_x = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$ ,

Dunque, possiamo esprimere la zona di rifiuto del test in due modi equivalenti:

$R = \{x_1, x_2, \dots, x_n; Z_x > z_\alpha\} \Leftrightarrow R = \{x_1, x_2, \dots, x_n; \bar{x} > \mu_0 + z_\alpha \cdot (\sigma/\sqrt{n})\}$

• Se  $(x_1, x_2, \dots, x_n)$  è un campione casuale osservato, con media  $\bar{x}$ , la decisione di rifiutare o non rifiutare

l'ipotesi nulla  $H_0: \mu = \mu_0$  si basa sul confronto tra  $\bar{x}$  e  $\mu_0$ , o meglio, sulla differenza standardizzata  $Z_x = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$  -> Valore assunto nel campione osservato dalla v.c.  $Z_x = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Ammettendo che l'ipotesi nulla sia vera e che la popolazione generatrice sia normale,  $Z_x$  è una v.c. normale  $N(0,1)$ .

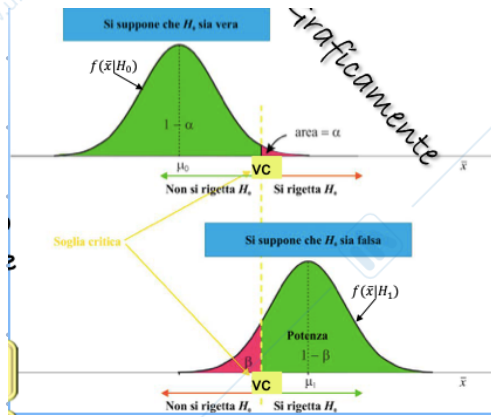
Quando la statistica test cade nella zona di rifiuto si dice che "il test è significativo", possiamo anche dire: "La media campionaria differisce significativamente da  $\mu_0$ ", "La differenza  $\bar{x} - \mu_0$  è significativamente diversa da 0", "Vi è sufficiente evidenza empirica contro l'ipotesi nulla".

### • Criteri di ottimizzazione nella verifica delle ipotesi

la teoria statistica consente di individuare il procedimento che, fissato il livello di  $\alpha$ , minimizza  $\beta$ , ovvero massimizza la potenza. Un tale procedimento è da considerarsi, ovviamente, ottimo. L'espressione della probabilità  $\beta$  per ipotesi alternativa:  $H_1 : \mu_1 > \mu_0$  è data da:

$$\beta = P(\bar{X} < \mu_0 + z_\alpha \cdot (\sigma/\sqrt{n}) \mid \mu = \mu_1)$$

$$\gamma = P(\bar{X} > \mu_0 + z_\alpha \cdot (\sigma/\sqrt{n}) \mid \mu = \mu_1) = 1 - \beta \rightarrow \text{chiamata potenza del test.}$$



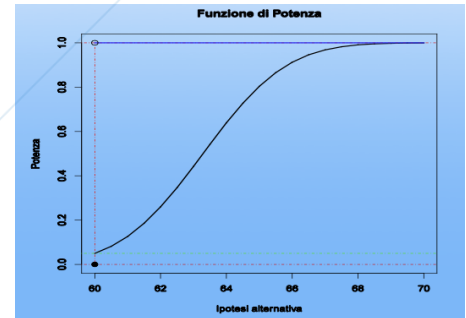
← sono illustrate le probabilità degli errori di prima e di seconda specie per la verifica delle ipotesi  $H_0: \mu = \mu_0$  contro  $H_1: \mu = \mu_1$ , con  $\mu_1$  maggiore di  $\mu_0$ .

$$\alpha = P(\bar{X} > VC \mid \mu = \mu_0)$$

$$\beta = P(\bar{X} < VC \mid \mu = \mu_1)$$

La potenza del test è influenzata da:

- 1- Dal livello di significatività.
- 2- Dalla specificazione dell'ipotesi alternativa.
- 3- dalla numerosità del campione.



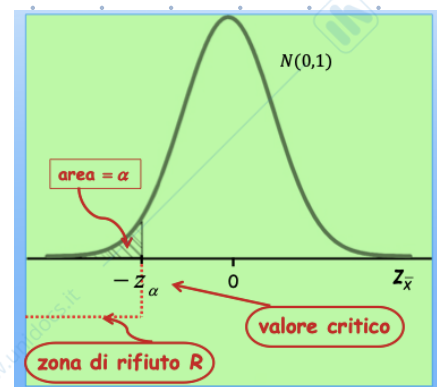
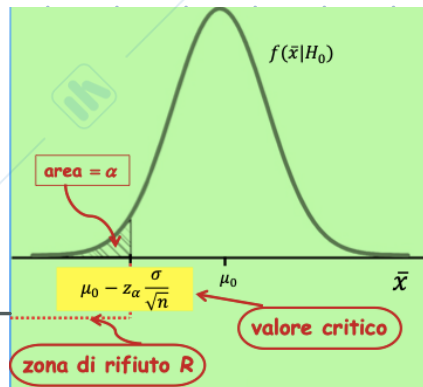
★ CASO  $H_0: \mu = \mu_0$

$H_1: \mu < \mu_0$

La disuguaglianza che determina il rifiuto di  $H_0$ :

$$\bar{x} < \mu_0 - z_\alpha \cdot (\sigma / \sqrt{n})$$

Si rifiuta  $H_0$  per valori della media campionaria di  $\mu_0$ , valori che,  $H_0$  molto più piccoli sotto  $H_0$ , sono stime poco «credibili» di  $\mu_0$ .



può essere, ovviamente, esplicitata in termini di valori standardizzati nel modo seguente

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

Dunque, possiamo esprimere la zona di rifiuto del test in due modi equivalenti:

$$R = \{x_1, x_2, \dots, x_n; z\bar{x} < -z_\alpha\} \Leftrightarrow R' = \{x_1, x_2, \dots, x_n; \bar{x} < \mu_0 - z_\alpha \cdot (\sigma / \sqrt{n})\}$$

• **Probabilità  $\beta$  e potenza del test**

$\beta$  è la probabilità che la statistica test  $\bar{X}$  assuma un valore appartenente alla zona di accettazione quando è vera l'ipotesi alternativa.

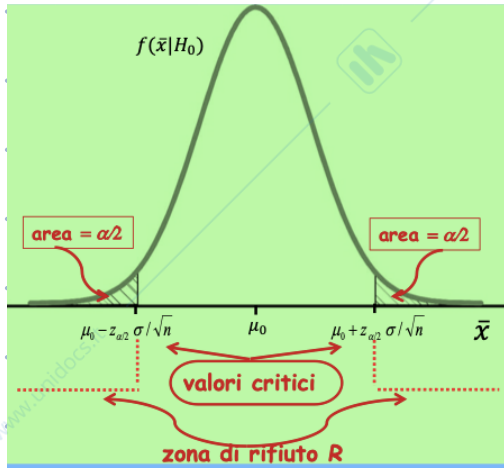
$$\beta = P(\bar{X} > \mu_0 - z_\alpha \cdot (\sigma / \sqrt{n}) \mid \mu = \mu_1)$$

$$1 - \beta = \gamma = P(\bar{X} > \mu_0 - z_\alpha \cdot (\sigma / \sqrt{n}) \mid \mu = \mu_1)$$

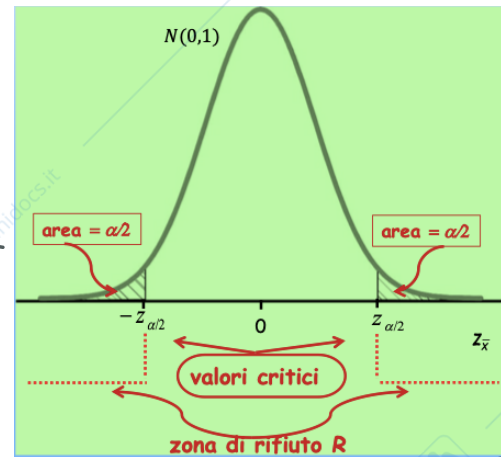
★ CASO  $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

In questo caso, si rifiuta l'ipotesi nulla per valori di  $z\bar{x}$  minori di  $-z\alpha/2$  o maggiori di  $z\alpha/2$ , oppure, equivalentemente per ogni  $\bar{X}$  minori di  $\mu_0 - z\alpha/2(\sigma/\sqrt{n})$  o maggiori di  $\mu_0 + z\alpha/2(\sigma/\sqrt{n})$



Si rifiuta  $H_0$  per valori di  $\bar{x}$  molto più piccoli di  $\mu_0 - z\alpha/2 \cdot (\sigma/\sqrt{n})$  o molto più grandi di  $\mu_0 + z\alpha/2(\sigma/\sqrt{n})$  valori che sotto  $H_0$  sono stime poco credibili di  $\mu_0$ .



Si rifiuta  $H_0$  per valori di  $z\bar{x}$  molto più piccoli o molto più grandi di zero, valori che, sotto  $H_0$ , sono stime poco credibili di  $\mu_0$ .

• **Probabilità  $\beta$  e potenza del test**

$\beta$  è la probabilità che la statistica test  $\bar{X}$  assuma un valore appartenente alla zona di accettazione quando è vera l'ipotesi alternativa

$$\beta = P(\mu_0 - z\alpha/2(\sigma/\sqrt{n}) < \bar{X} < \mu_0 + z\alpha/2(\sigma/\sqrt{n}) \mid \mu = \mu_1)$$

$$1 - \beta = \gamma = P(\mu_0 - z\alpha/2(\sigma/\sqrt{n}) \mid \mu = \mu_1) + P(\mu_0 + z\alpha/2(\sigma/\sqrt{n}) \mid \mu = \mu_1)$$

★ **Livello di significativita' osservato**

È la probabilità che la statistica test assuma un valore minore di quello ottenuto con i dati del campione (valore più estremo), ammesso che l'ipotesi nulla sia vera.

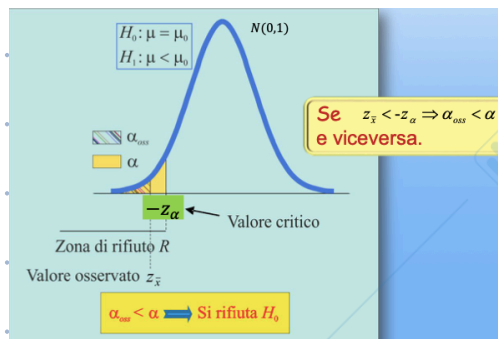
si rifiuta l'ipotesi nulla se il livello di significatività osservato è minore di  $\alpha$

• CASO  $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$

$$\alpha_{oss} = P(Z\bar{X} < z\bar{x} \mid \mu = \mu_0)$$

valore assunto dalla statistica test  $Z\bar{X}$



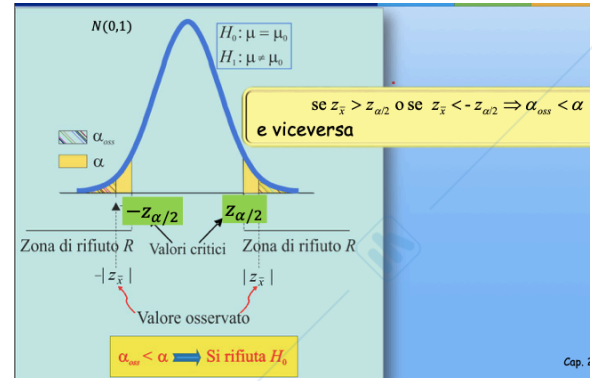
$$\alpha_{oss} = P(Z\bar{X} < -|z\bar{x}|) + P(Z\bar{X} > |z\bar{x}|) = 2P(Z\bar{X} > |z\bar{x}|)$$

Si ricorda che  $z\bar{x} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$  può assumere valori positivi o negativi per questo si prende il valore assoluto.

• CASO  $H_0: \mu = \mu_0$   
 $H_1: \mu \neq \mu_0$   
 $\alpha_{oss} = P(Z_{\bar{X}} < -|z_{\bar{X}}|) + P(Z_{\bar{X}} > |z_{\bar{X}}|)$   
 $= 2P(Z_{\bar{X}} > |z_{\bar{X}}|)$

È la probabilità che la statistica test assuma un valore più estremo di quello ottenuto con i dati del campione osservato (più grande di  $|z_{\bar{X}}|$  o più piccolo di  $-|z_{\bar{X}}|$ ), ammesso che l'ipotesi nulla sia vera.

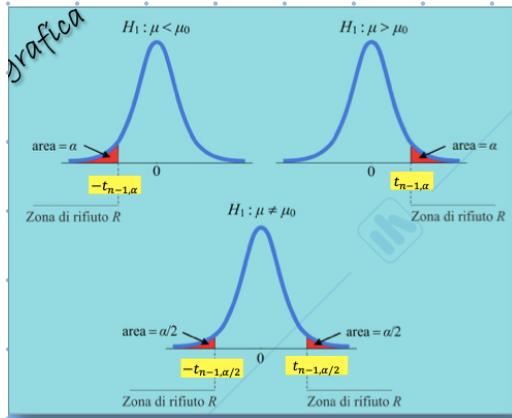
Si rifiuta l'ipotesi nulla se il livello di significatività osservato è minore di  $\alpha$ .



### VERIFICA D'IPOTESI SULLA MEDIA DI UNA POPOLAZIONE CON VARIANZA NON NOTA

La verifica dell'ipotesi si avvale della statistica test  $T_{\bar{X}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

che, sotto l'ipotesi di normalità della popolazione e se l'ipotesi nulla è vera, ha distribuzione t di Student con  $n - 1$  gradi di libertà.



$H_1: \mu < \mu_0 \rightarrow \alpha_{oss} = P(T_{\bar{X}} < t_{\bar{X}} | \mu = \mu_0)$

$H_1: \mu > \mu_0 \rightarrow \alpha_{oss} = P(T_{\bar{X}} > t_{\bar{X}} | \mu = \mu_0)$

$H_1: \mu \neq \mu_0 \rightarrow \alpha_{oss} = 2P(T_{\bar{X}} > |t_{\bar{X}}| | \mu = \mu_0)$

si rifiuta l'ipotesi nulla se il livello di significatività osservato è minore di  $\alpha$ .

### VERIFICA D'IPOTESI SULLA MEDIA NEL CASO DI GRANDI CAMPIONI

Se il campione ha dimensione sufficientemente grande, qualunque sia la popolazione generatrice, la statistica test da assumere è  $Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ , che è una v.c. la cui distribuzione può essere approssimata con una normale

ZONE DI RIFIUTO per le diverse ipotesi

Ipotesi $H_0$	Ipotesi $H_1$	Zona di rifiuto R
$\mu = \mu_0$	$\mu > \mu_0$	$\{z_{\bar{X}} : z_{\bar{X}} > z_{\alpha}\}$
	$\mu < \mu_0$	$\{z_{\bar{X}} : z_{\bar{X}} < -z_{\alpha}\}$
	$\mu \neq \mu_0$	$\{z_{\bar{X}} : z_{\bar{X}} > z_{\alpha/2}\}$

valore assunto dalla statistica test  $Z_{\bar{X}}$

#### • Livello di significatività osservato

Si calcola come nel caso della varianza nota con l'unica differenza che a denominatore di  $z_{\bar{X}}$  appare la deviazione standard campionaria e anziché quella della popolazione.

## VERIFICA DI IPOTESI SULLA MEDIA NEL CASO DI GRANDI CAMPIONI: POPOLAZIONI BERNOULLIANE

Se il campione casuale di ampiezza  $n$  (sufficientemente elevata) proviene da una popolazione bernoulliana, l'ipotesi nulla da verificare è  $H_0: p = p_0$ .

La frequenza relativa dei "successi" nel campione, ha, sotto l'ipotesi nulla, distribuzione prossima alla normale con media  $p_0$  e varianza  $p_0(1 - p_0)/n$ . La statistica test è;

$$Z_p = \frac{p - p_0}{\sqrt{(p_0 \cdot q_0)/n}}$$

- Livello di significatività osservato  $\rightarrow$  Si calcola come nei casi precedenti.

## VERIFICA DELLE IPOTESI SULLA VARIANZA

- Sistema di ipotesi:

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 > \sigma_0^2$$

- Campione osservato:  $(x_1, x_2, \dots, x_n)$ , proveniente da una popolazione normale con media e varianza entrambe incognite.

- statistica test:  $V = \frac{(N-1)S^2}{\sigma_0^2}$

Se l'ipotesi nulla è vera, la v.c.  $V$  ha distribuzione chi-quadrato con  $n - 1$  gradi di libertà.

### Calcolo del valore critico

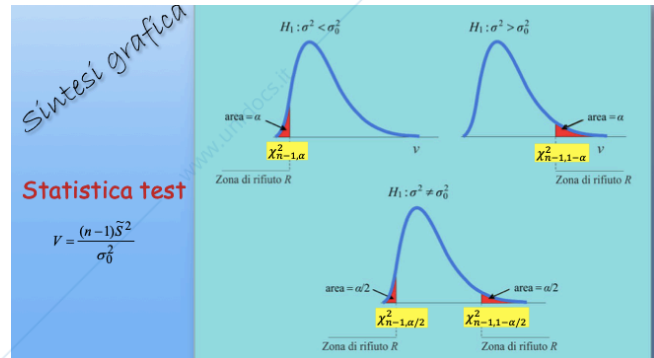
Si fissa  $\alpha$  e si determina quel valore VC tale che  $P(\tilde{S} > VC | \sigma^2 = \sigma_0^2) = \alpha$ ,

quindi  $P\left[\frac{(N-1)\tilde{S}^2}{\sigma_0^2} > \frac{(N-1)VC}{\sigma_0^2}\right] = \alpha$

dalle tavole  $\frac{(N-1)VC}{\sigma_0^2} = X^2_{n-1, 1-\alpha}$

per cui  $VC = \frac{\sigma_0^2}{(N-1)} \cdot X^2_{n-1, 1-\alpha}$

Ipotesi $H_0$	Ipotesi $H_1$	Statistica test	Zona di rifiuto $R$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$V = \frac{(n-1)\tilde{S}^2}{\sigma_0^2}$	$\{v: v > X^2_{n-1, 1-\alpha}\}$
	$\sigma^2 < \sigma_0^2$		$\{v: v < X^2_{n-1, \alpha}\}$
	$\sigma^2 \neq \sigma_0^2$		$\{v: v \notin (X^2_{n-1, \alpha/2}, X^2_{n-1, 1-\alpha/2})\}$
		$\alpha_{oss} = P(V > v   \sigma^2 = \sigma_0^2)$	$\alpha_{oss} = P(V < v   \sigma^2 = \sigma_0^2)$
$\alpha_{oss} = 2\min[P(V < v   \sigma^2 = \sigma_0^2), P(V > v   \sigma^2 = \sigma_0^2)]$			



### • Probabilità $\beta$ e potenza del test

Consideriamo, come esempio, il sistema di ipotesi:  $H_0: \sigma^2 = \sigma_0^2$  e  $H_1: \sigma^2 = \sigma_1^2$  (con  $\sigma_1^2 < \sigma_0^2$ ).

La probabilità dell'errore di seconda specie è data da  $\beta = P[\tilde{S}^2 > VC | \sigma^2 = \sigma_1^2]$

Dove VC è il valore critico inferiore. Sotto l'ipotesi alternativa  $(n-1) \tilde{S}^2 / \sigma_1^2$

quindi  $VC = \frac{\sigma_1^2}{n-1} \cdot X^2_{n-1, \alpha}$  per sostituzione  $\beta = P\left(\frac{(n-1)\tilde{S}^2}{\sigma_1^2} > \frac{\sigma_0^2}{\sigma_1^2} \cdot X^2_{n-1, \alpha}\right)$

**IN BREVE:**

**VERIFICA DI IPOTESI SU UNA MEDIA CON VARIANZA NOTA**

$$X \sim N(\mu, \sigma^2)$$

$$H_0 : \mu = \mu_0$$

**LA MIGLIOR REGIONE DI RIFIUTO E' DATA DA**

$$R = \left\{ \begin{array}{ll} \bar{x} < \mu_0 - z_{\alpha} \frac{\sigma}{\sqrt{n}} & \text{se } H_1 : \mu < \mu_0 \\ \bar{x} > \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}} & \text{se } H_1 : \mu > \mu_0 \\ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} & \text{se } H_1 : \mu \neq \mu_0 \end{array} \right.$$

**OPPURE**

$$R = \left\{ \begin{array}{ll} \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{\alpha} & \text{se } H_1 : \mu < \mu_0 \\ \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{\alpha} & \text{se } H_1 : \mu > \mu_0 \\ \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2} & \text{se } H_1 : \mu \neq \mu_0 \end{array} \right.$$

**DOVE**

$$z_{\gamma}(\gamma = \alpha, \alpha/2)$$

**E' TALE CHE**

$$P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_{\gamma}\right) = \gamma$$

**VERIFICA DI IPOTESI SU UNA MEDIA CON VARIANZA INCOGNITA**

$$X \sim N(\mu, \sigma^2)$$

$$H_0 : \mu = \mu_0$$

$$\tilde{S}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

**LA MIGLIOR REGIONE DI RIFIUTO E' DATA DA**

$$R = \left\{ \begin{array}{ll} \bar{x} < \mu_0 - t_{n-1, \alpha} \left(\frac{\tilde{s}}{\sqrt{n}}\right) & \text{se } H_1 : \mu < \mu_0 \\ \bar{x} > \mu_0 + t_{n-1, \alpha} \left(\frac{\tilde{s}}{\sqrt{n}}\right) & \text{se } H_1 : \mu > \mu_0 \\ |\bar{x} - \mu_0| > t_{n-1, \frac{\alpha}{2}} \left(\frac{\tilde{s}}{\sqrt{n}}\right) & \text{se } H_1 : \mu \neq \mu_0 \end{array} \right.$$

**OPPURE**

$$R = \left\{ \begin{array}{ll} \frac{\bar{x} - \mu_0}{\tilde{s}/\sqrt{n}} < -t_{n-1, \alpha} & \text{se } H_1 : \mu < \mu_0 \\ \frac{\bar{x} - \mu_0}{\tilde{s}/\sqrt{n}} > t_{n-1, \alpha} & \text{se } H_1 : \mu > \mu_0 \\ \left| \frac{\bar{x} - \mu_0}{\tilde{s}/\sqrt{n}} \right| > t_{n-1, \frac{\alpha}{2}} & \text{se } H_1 : \mu \neq \mu_0 \end{array} \right.$$

**DOVE**

$$t_{n-1, \gamma}(\gamma = \alpha, \alpha/2)$$

**E' TALE CHE**

$$P\left(\frac{\bar{X} - \mu_0}{\tilde{S}/\sqrt{n}} > t_{n-1, \gamma}\right) = \gamma$$

**VERIFICA DI IPOTESI SULLA PROPORZIONE**

$$X \sim Be(p)$$

$$H_0 : p = p_0$$

**SOTTO  $H_0$  E PER  $n$  GRANDE**

$$\hat{p} = \bar{X} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

**LA MIGLIOR REGIONE DI RIFIUTO E' DATA DA**

$$R = \left\{ \begin{array}{ll} \bar{x} < p_0 - z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} & \text{se } H_1 : p < p_0 \\ \bar{x} > p_0 + z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} & \text{se } H_1 : p > p_0 \\ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} & \text{se } H_1 : p \neq p_0 \end{array} \right.$$

**OPPURE**

$$R = \left\{ \begin{array}{ll} \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < -z_{\alpha} & \text{se } H_1 : p < p_0 \\ \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{\alpha} & \text{se } H_1 : p > p_0 \\ \left| \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right| > z_{\alpha/2} & \text{se } H_1 : p \neq p_0 \end{array} \right.$$

**DOVE**

$$z_{\gamma}(\gamma = \alpha, \alpha/2)$$

**E' TALE CHE**

$$P\left(\frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{\gamma}\right) = \gamma$$

**VERIFICA DI IPOTESI SULLA VARIANZA**

$$X \sim N(\mu, \sigma^2)$$

$$H_0 : \sigma^2 = \sigma_0^2$$

$$\tilde{S}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

**LA MIGLIOR REGIONE DI RIFIUTO E' DATA DA**

$$R = \left\{ \begin{array}{ll} \tilde{s}^2 < \frac{\sigma_0^2}{n-1} \chi_{n-1, \alpha}^2 & \text{se } H_1 : \sigma^2 < \sigma_0^2 \\ \tilde{s}^2 > \frac{\sigma_0^2}{n-1} \chi_{n-1, 1-\alpha}^2 & \text{se } H_1 : \sigma^2 > \sigma_0^2 \\ \text{oppure} & \\ \tilde{s}^2 > \frac{\sigma_0^2}{n-1} \chi_{n-1, \frac{\alpha}{2}}^2 & \text{se } H_1 : \sigma^2 \neq \sigma_0^2 \end{array} \right.$$

**OPPURE**

$$R = \left\{ \begin{array}{ll} \frac{(n-1)\tilde{s}^2}{\sigma_0^2} < \chi_{n-1, \alpha}^2 & \text{se } H_1 : \sigma^2 < \sigma_0^2 \\ \frac{(n-1)\tilde{s}^2}{\sigma_0^2} > \chi_{n-1, 1-\alpha}^2 & \text{se } H_1 : \sigma^2 > \sigma_0^2 \\ \text{oppure} & \\ \frac{(n-1)\tilde{s}^2}{\sigma_0^2} > \chi_{n-1, \frac{\alpha}{2}}^2 & \text{se } H_1 : \sigma^2 \neq \sigma_0^2 \end{array} \right.$$

**DOVE**

$$\chi_{n-1, \gamma}^2(\gamma = \alpha, \alpha/2, \gamma = 1 - \alpha, \gamma = 1 - \alpha/2)$$

**E' TALE CHE**

$$P\left(\frac{(n-1)\tilde{S}^2}{\sigma_0^2} < \chi_{n-1, \gamma}^2\right) = \gamma$$

## CONFRONTI TRA DUE POPOLAZIONI: INFERENZA SULLA DIFFERENZA DI MEDIE CON VARIANZE NOTE

### • Confronto tra medie di due pop. indipendenti

Consideriamo due popolazioni distribuite NORMALMENTE.

Supponiamo di estrarre un campione di ampiezza  $n_1$  dalla prima popolazione di ampiezza  $n_2$  dalla seconda popolazione INDIPENDENTI.

Siano  $\mu_1$  e  $\mu_2$  le medie che caratterizzano rispettivamente la prima e la seconda popolazione e si assumano NOTI i due scarti quadratici medi  $\sigma_1$  e  $\sigma_2$  delle due popolazioni.

### • Inferenza sulle medie di due popolazioni normali

Indichiamo con  $\mu = \mu_1 - \mu_2$ .

Sia  $\bar{X} = \bar{X}_1 - \bar{X}_2$  uno stimatore di  $\mu$

La v.c. differenza tra le medie campionarie di due campioni indipendenti provenienti da due popolazioni normali è distribuita normalmente  $\bar{X} \sim (\mu, \sigma^2_{\bar{X}})$ . Dove la media e la varianza sono date da:

$$E(\bar{X}) = E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = \mu,$$

$$\sigma^2(\bar{X}) = \sigma^2(\bar{X}_1 - \bar{X}_2) = \sigma^2_1/n_1 + \sigma^2_2/n_2$$

Il rapporto  $Z_{\bar{X}} = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$  oppure  $Z_{\bar{X}_1 - \bar{X}_2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$  è una v.c. normale standard.

Se le due popolazioni non hanno distribuzione normale la funzione cardine Z può essere utilizzata con ampiezze campionarie sufficientemente elevate ( $n \geq 30$ ).

• Int. di confidenza per la differenza di due medie -pop. normali (varianze note)

-> impiego lo stimatore  $\bar{X} = \bar{X}_1 - \bar{X}_2$   $l_1 = \bar{X} - z_{\alpha/2} \cdot \sigma_{\bar{X}}$

$P[-z_{\alpha/2} < (\bar{X} - \mu) / \sigma_{\bar{X}} < z_{\alpha/2}] = 1 - \alpha$  da cui  $P[\bar{X} - z_{\alpha/2} \cdot \sigma_{\bar{X}} < \mu < \bar{X} + z_{\alpha/2} \cdot \sigma_{\bar{X}}] = 1 - \alpha$   $l_2 = \bar{X} + z_{\alpha/2} \cdot \sigma_{\bar{X}}$

-> impiego la funzione  $Z_{\bar{X}_1 - \bar{X}_2}$

$P[-z_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} < z_{\alpha/2}] = 1 - \alpha$  da cui

$$P[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \cdot \sigma_{\bar{X}_1 - \bar{X}_2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \cdot \sigma_{\bar{X}_1 - \bar{X}_2}] = 1 - \alpha$$

$$P[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \cdot \sigma_{\bar{X}_1 - \bar{X}_2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \cdot \sigma_{\bar{X}_1 - \bar{X}_2}] = 1 - \alpha$$

### • Test per la differenza di due medie: pop. normali (varianze note)

Sia  $H_0: \mu = \mu_1 - \mu_2 = 0$

L'ipotesi alternativa contrapposta può assumere una delle tre configurazioni:

$H_1: \mu < 0$  ipotesi alternativa unidirezionale

$H_1: \mu > 0$  ipotesi alternativa unidirezionale

$H_1: \mu \neq 0$  ipotesi alternativa bidirezionale

Se l'ipotesi nulla è vera, il valore atteso della differenza  $\bar{X} = \bar{X}_1 - \bar{X}_2$  è nullo e la statistica:

$Z_{\bar{X}} = \frac{\bar{X}}{\sigma_{\bar{X}}}$  oppure  $Z_{\bar{X}_1 - \bar{X}_2} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$ . È una variabile casuale  $N(0,1)$ .

Ipotesi $H_0$	Ipotesi $H_1$	Statistica test	Zona di rifiuto R
$\mu = \mu_1 - \mu_2 = 0$	$\mu = \mu_1 - \mu_2 < 0$	$Z_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}}$	$\{z_{\alpha}; z_{\alpha}; < -z_{\alpha}\}$
	$\mu = \mu_1 - \mu_2 > 0$		$\{z_{\alpha}; z_{\alpha}; > z_{\alpha}\}$
	$\mu = \mu_1 - \mu_2 \neq 0$		$\{z_{\alpha}; z_{\alpha};  z_{\alpha}  > z_{\alpha/2}\}$

### • Livello di significatività osservato

Sia, per esempio,  $H_1: \mu = \mu_1 - \mu_2 > 0$

$$\alpha_{oss} = P[Z_{\bar{X}} > z_{\alpha} | \mu = 0]$$

$$\alpha_{oss} = P[Z_{\bar{X}_1 - \bar{X}_2} > z_{\alpha} | \mu = 0]$$

La logica sottostante a queste regole di decisione è che valori molto grandi (o molto piccoli) di  $Z(\bar{X}_1 - \bar{X}_2)$  non sono, verosimilmente, attribuibili al caso, ma al fatto che la media della prima popolazione è superiore (o inferiore) a quella della seconda, come contemplato dall'ipotesi alternativa.

Si rifiuta l'ipotesi nulla se il livello di significatività osservato è inferiore al livello di significatività  $\alpha$ .

## CONFRONTI TRA DUE POPOLAZIONI: DIFFERENZA DI MEDIE, VARIANZE NON NOTE

### • Confronto tra medie di due pop. indipendenti

Popolazioni distribuite NORMALMENTE.

Varianze delle due popolazioni NON NOTE ma UGUALI. Ipotesi di omogeneità della varianze ( $\sigma_{21} = \sigma_{22}$ ). (OMOSCHEDASTICITA')

Campioni INDIPENDENTI con numerosità piccole ( $n < 30$ ).

Per il confronto tra le medie di due popolazioni si utilizza il test t: t-test

### • Inferenza su due medie: piccoli campioni e varianze uguali

Si abbiano due campioni provenienti da due popolazioni normali aventi medie  $\mu_1$  e  $\mu_2$  e varianze incognite ma uguali, allora:

$$T_{\bar{X}} = \frac{\bar{X} - \mu}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}}$$

dove  $\bar{X} = \bar{X}_1 - \bar{X}_2$ ;  $\mu = \mu_1 - \mu_2$ ;  $\text{Sc} = \frac{\tilde{S}_1^2(n_1 - 1) + \tilde{S}_2^2(n_2 - 1)}{(n_1 + n_2 - 2)}$

ha distribuzione t di Student con  $n_1 + n_2 - 2$  gradi di libertà.

$$T(\bar{X}_1 - \bar{X}_2) = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}}$$

La formula precedente deriva dal seguente rapporto

### • Intervallo di confidenza per la differenza di due medie -piccoli campioni

$$P[-t_{\alpha/2} < \frac{\bar{X} - \mu}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}} < t_{\alpha/2}] = 1 - \alpha$$

$$l_1 = \bar{X} - t_{\alpha/2} \cdot \text{Sc}\sqrt{(1/n_1) + (1/n_2)}$$

$$l_2 = \bar{X} + t_{\alpha/2} \cdot \text{Sc}\sqrt{(1/n_1) + (1/n_2)}$$

oppure

$$P[-t_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}} < t_{\alpha/2}] = 1 - \alpha$$

$$l_1 = (\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} \cdot \text{Sc}\sqrt{(1/n_1) + (1/n_2)}$$

$$l_2 = (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} \cdot \text{Sc}\sqrt{(1/n_1) + (1/n_2)}$$

### • Test per la differenza di due medie: piccoli campioni e varianze uguali

statistica test da impiegare per la verifica di ipotesi:

$$T(\bar{X}_1 - \bar{X}_2) = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}} \quad \text{oppure} \quad T_{\bar{X}} = \frac{\bar{X} - \mu}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}}$$

zona di rifiuto  $R = \{t(\bar{X}_1 - \bar{X}_2) : t(\bar{X}_1 - \bar{X}_2) > t_{\alpha}\}$

$$T(\bar{X}_1 - \bar{X}_2) = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\text{Sc}\sqrt{(1/n_1) + (1/n_2)}}$$

Ipotesi $H_0$	Ipotesi $H_1$	Statistica test	Zona di rifiuto R
$\mu = 0$	$\mu < 0$	$T_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\text{Sc}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$\{t_{\bar{X}_1 - \bar{X}_2} : t_{\bar{X}_1 - \bar{X}_2} < -t_{\alpha}\}$
	$\mu > 0$		$\{t_{\bar{X}_1 - \bar{X}_2} : t_{\bar{X}_1 - \bar{X}_2} > t_{\alpha}\}$
	$\mu \neq 0$		$\{t_{\bar{X}_1 - \bar{X}_2} :  t_{\bar{X}_1 - \bar{X}_2}  > t_{\alpha/2}\}$

$T_{\alpha/2}$  e  $t_{\alpha}$  sono i quantili di una distribuzione t di Student con  $n_1 + n_2 - 2$  gradi di libertà

**Inferenza sulla differenza di due medie: varianze non assunte uguali grandi campioni**

statistica test da impiegare per la verifica di ipotesi:

$$Z_x = \frac{\bar{X} - \mu}{\sigma_x} = Z(\bar{X}_1 - \bar{X}_2) = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S^2_{x1}/n_1 + S^2_{x2}/n_2}}$$

la statistica  $Z_x$  è approssimativamente una normale  $N(0,1)$  se i campioni sono grandi

Limiti intervallo di confidenza			
		$l_1 = (\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \hat{\sigma}_{\bar{x}}$	
		$l_2 = (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \hat{\sigma}_{\bar{x}}$	
Ipotesi $H_0$	Ipotesi $H_1$	Statistica test	Zona di rifiuto R
$\mu = \mu_1 - \mu_2 = 0$	$\mu < 0$	$Z_{\bar{x}} = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{x}}}$	$\{z_{\bar{x}} : z_{\bar{x}} < -z_{\alpha}\}$
	$\mu > 0$		$\{z_{\bar{x}} : z_{\bar{x}} > z_{\alpha}\}$
	$\mu \neq 0$		$\{z_{\bar{x}} :  z_{\bar{x}}  > z_{\alpha/2}\}$

**CONFRONTI TRA DUE POPOLAZIONI: DIFFERENZA DI PROPORZIONI (GRANDI CAMPIONI).**

**Intervallo di confidenza per la differenza di due proporzioni (grandi campioni)**

Proporzioni delle due popolazioni:  $p_1$  e  $p_2$  ;

Proporzione dei due campioni:  $\hat{p}_1$  e  $\hat{p}_2$ ;

I due campioni sono indipendenti.

Si indichi con  $p = p_1 - p_2$  e con  $\hat{p} = \hat{p}_1 - \hat{p}_2$

si ha che:

$$E(\hat{p}) = E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 = p$$

$$Var(\hat{p}) = \sigma^2_{\hat{p}} = [p_1(1-p_1)/n_1] + [p_2(1-p_2)/n_2]$$

Per costruire l'intervallo di confidenza sulla differenza di proporzioni si usa la statistica:

$$Z_{\hat{p}} = \frac{\hat{p} - p}{\sigma_{\hat{p}}} \quad \text{oppure} \quad Z(\hat{p}_1 - \hat{p}_2) = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma^2_{\hat{p}_1 - \hat{p}_2}}$$

la cui distribuzione è prossima alla normale  $N(0, 1)$ , perché i campioni hanno grandi dimensioni.

Dalla proposizione

$$P[-z_{\alpha/2} < \frac{\hat{p} - p}{\sigma_{\hat{p}}} < z_{\alpha/2}] \approx 1 - \alpha \quad \text{oppure} \quad P[-z_{\alpha/2} < \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma^2_{\hat{p}_1 - \hat{p}_2}} < z_{\alpha/2}] \approx 1 - \alpha$$

con l'usuale procedimento, si ottengono gli estremi dell'intervallo di confidenza per  $p - p$  :

$$l_1 = p - z_{\alpha/2} \cdot \sigma_{\hat{p}} \quad \text{oppure} \quad l_1 = (\hat{p}_1 - \hat{p}_2) - z_{\alpha/2} \cdot \sigma^2_{\hat{p}_1 - \hat{p}_2}$$

$$l_2 = p + z_{\alpha/2} \cdot \sigma_{\hat{p}} \quad \text{oppure} \quad l_2 = (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2} \cdot \sigma^2_{\hat{p}_1 - \hat{p}_2}$$

**Test per la differenza di due proporzioni**

Sia  $H_0 : p = p_1 - p_2 = 0$

L'ipotesi alternativa contrapposta può assumere una

delle tre configurazioni:

$H_1 : p < 0$  ipotesi alternativa unidirezionale

$H_1 : p > 0$  ipotesi alternativa unidirezionale

$H_1 : p \neq 0$  ipotesi alternativa bidirezionale

Sia  $H_0 : p = p_1 - p_2 = 0$ , sotto  $H_0$   $E(\hat{p}) = 0$  e

$Var(\hat{p}) = \sigma^2_{\hat{p}} = Poiché$  sotto  $H_0$ ,  $p_1$  e  $p_2$  sono uguali, diciamo  $p_0 = p_0 \cdot q_0 \cdot (1/n_1 + 1/n_2)$

Se  $H_0$  è vera, i due campioni provengono da

una stessa popolazione bernoulliana con parametro  $p_0$ . Si può parlare di un unico campione di numerosità  $n_1 + n_2$ . Lo stimatore di massima verosomiglianza di  $p_0$  è

dato da:

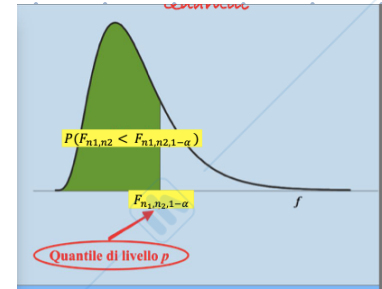
$$\hat{p}_C = \frac{\sum X_i + \sum Y_i}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

Ipotesi $H_0$	Ipotesi $H_1$	Statistica test	Zona di rifiuto R
$p = p_1 - p_2 = 0$	$p = p_1 - p_2 < 0$	$Z_{\hat{p}_1 - \hat{p}_2} = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}}$	$\{z_{\hat{p}_1 - \hat{p}_2} : z_{\hat{p}_1 - \hat{p}_2} < -z_{\alpha}\}$
	$p = p_1 - p_2 > 0$		$\{z_{\hat{p}_1 - \hat{p}_2} : z_{\hat{p}_1 - \hat{p}_2} > z_{\alpha}\}$
	$p = p_1 - p_2 \neq 0$		$\{z_{\hat{p}_1 - \hat{p}_2} :  z_{\hat{p}_1 - \hat{p}_2}  > z_{\alpha/2}\}$

## CONFRONTI TRA VARIANZE

### • Distribuzione di probabilità F di Fisher

Siano  $X_1$  e  $X_2$  due variabili casuali indipendenti aventi distribuzioni di probabilità chi-quadrato con  $n_1$  e  $n_2$  gradi di libertà. Allora, la distribuzione di probabilità del rapporto  $Y = \frac{X_1/n_1}{X_2/n_2}$  è una variabile casuale F di Fisher con  $n_1$  e  $n_2$  gdl



### • Test per le varianze di due popolazioni normali

Siano  $C_1$  e  $C_2$  due campioni indipendenti provenienti dalle popolazioni normali  $P_1$  e  $P_2$  aventi la stessa varianza. Siano  $S_1^2$  e  $S_2^2$  le varianze del primo e del secondo campione. Allora, il rapporto:

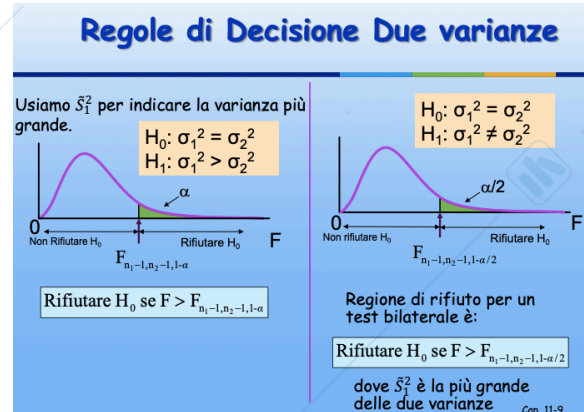
$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$  ha distribuzione F di Fisher con  $n_1 - 1$  gradi di libertà del numeratore e  $n_2 - 1$  gradi di libertà del denominatore.

Il rapporto:  $F = \frac{S_1^2/\tilde{S}_1^2}{S_2^2/\tilde{S}_2^2}$  è una appropriata statistica test per sottoporre a verifica l'ipotesi di uguaglianza delle varianze di due popolazioni normali  $P_1$  e  $P_2$ .

Se  $\sigma_1^2$  è la varianza di  $P_1$  e  $\sigma_2^2$  è la varianza di  $P_2$ , le zone di rifiuto del test per la verifica dell'ipotesi nulla  $H_0: \sigma_1^2 = \sigma_2^2$

Sono quelle indicate nella tabella

Ipotesi $H_0$	Ipotesi $H_1$	Statistica test	Zona di rifiuto R
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{\tilde{S}_1^2}{\tilde{S}_2^2}$	$\{f: f < f_{n_1-1, n_2-1, \alpha}\}$
	$\sigma_1^2 > \sigma_2^2$		$\{f: f > f_{n_1-1, n_2-1, 1-\alpha}\}$
	$\sigma_1^2 \neq \sigma_2^2$		$\{f: f \notin (f_{n_1-1, n_2-1, \alpha/2}, f_{n_1-1, n_2-1, 1-\alpha/2})\}$



## TEST SULLA BONTA' DI ADATTAMENTO

(o test  $X^2$ )

Il test chi quadrato è una tecnica particolarmente utile per verificare se i dati osservati sono coerenti (consistenti) con una particolare distribuzione.

Sia dato un campione casuale di  $n$  unità nella quale si osserva il carattere A. distribuzione frequenza campione osservato:

categoria	$C_1$	$C_2$	...	$C_k$	tot
frequenza	$n_1$	$n_2$	...	$n_k$	$n$

modello descrittivo della popolazione generatrice del campione:

categoria	$C_1$	$C_2$	...	$C_k$	tot
frequenza	$P_1$	$P_2$	...	$P_k$	1

Sia  $H_0 = p_i = p_{i0}$

Se l'ipotesi nulla è vera, i valori attesi delle frequenze  $n_1 \dots n_k$  sono dati da  $np_{10}, np_{20}, \dots, np_{k0}$

perché la generica frequenza della seconda tabella è una v.c. binomiale con parametri  $n$  e  $p_{i0}$ . ci chiediamo:

I dati campionari conformano con la distribuzione di frequenza ipotizzata?

I dati campionari conformano con specificate probabilità tecniche?

La verifica dell'ipotesi nulla consiste nel misurare una "distanza" tra la distribuzione di frequenze osservata e la distribuzione di frequenze attese sotto l'ipotesi nulla.

Il termine "bontà di adattamento" si riferisce a quanto bene le frequenze osservate si adattano alle frequenze teoriche.

Le ipotesi del test sono:

In altri termini,

$H_0: F(x) = F_0(x) \rightarrow$  distribuzione popolazione = distribuzione teorica

$H_1: F(x) \neq F_0(x) \rightarrow$  distribuzione popolazione  $\neq$  distribuzione teorica

La distribuzione della popolazione,  $F_0(x)$ , può essere interamente specificata oppure solo parzialmente specificata. In quest'ultimo caso le ipotesi del test sono:

$H_0: F(x) \in \Omega$

$H_1: F(x) \neq \Omega$

Dove  $\Omega$  è una famiglia di funzioni di distribuzione con forma analitica definita (ad esempio famiglia di distribuzioni normali).

### • il test e la regione di rifiuto:

La distribuzione di  $\chi^2$  è approssimativamente una variabile casuale chi-quadro su campioni sufficientemente grandi.

### • regola costruzione test:

$$\chi^2 = \sum_{i=1}^k \frac{(\text{frequenze Osservate} - \text{frequenze Attese})^2}{\text{frequenze Attese}} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

-Specificare  $H_0$  completamente sia nella forma funzionale che nei parametri.

-Riproporre  $N$  determinazioni del campione casuale:  $(x_1, x_2, \dots, x_N)$  sotto forma di distribuzione di frequenza discreta o continua a seconda dell'ipotesi specificata:

$(x_i, n_i) \quad i=1, \dots, k \quad n_i = \text{frequenza osservata}$

o continua a seconda dell'ipotesi specificata:

$(x_{i-1}, x_i, n_i) \quad i=1, \dots, k \quad n_i = \text{frequenza osservata}$

-Sulla base dell'ipotesi nulla calcolare le probabilità,  $p_i$ , che in una prova si presenti la modalità  $x_i$  o la probabilità di appartenere alla classe  $(x_{i-1}, x_i)$ :

$p_i = P(X=x_i | H_0)$  oppure  $P(X \in (x_{i-1}, x_i) | H_0)$

-Calcolare le frequenze attese moltiplicandopi per il totale delle osservazioni,  $N$ :  $E_i = Np_i$

-Calcolare il chi-quadro:  $\chi^2 = \sum \frac{(n_i - Np_i)^2}{Np_i}$

-Calcolare il valore teorico  $\chi^2_{gdl, 1-\alpha}$  dove i gradi di libertà sono  $k-1$  o  $k-1-m$

-Se  $\chi^2 > \chi^2_{gdl, 1-\alpha}$  si rifiuta l'ipotesi che il campione osservato provenga da una popolazione avente una distribuzione specificata dall'ipotesi nulla.

### Ci sono dei suggerimenti:

- Il campione deve essere sufficientemente grande.
- Il numero delle modalità o degli intervalli di modalità non dovrà essere troppo piccolo (almeno 5, 6)
- La frequenza  $n_i$  assegnata ad ogni modalità (classi) dovrà essere in pratica non inferiore a 5.

• Rifiutare  $H_0$  se  $\chi^2 > \chi^2_{gdl, 1-\alpha}$

con  $gdl = k - 1 - m$ ,  
 $m =$  numero dei parametri stimati sul campione

