FORMULARIO INFERENZA STATISTICA

Probabilità:

- definizione classica:

$$P(E) = \frac{numero\ di\ risultati\ favorevoli\ all'evento\ E}{numero\ di\ risultati\ possibili} = \frac{n_E}{n_S}$$

- definizione frequentista:

$$P(E) = \lim_{n \to \infty} \frac{n_E}{n}$$

- Postulato 3 della teoria assiomatica

$$A \cap B = \emptyset \quad \rightarrow \quad P(A \cup B) = P(A) + P(B)$$

- probabilità condizionata:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- principio delle probabilità composte:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

- indipendenza statistica:

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

- teorema sull'indipendenza statistica

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(A \cap \bar{B}) = P(A) \cdot P(\bar{B})$$

$$P(\bar{A} \cap B) = P(\bar{A}) \cdot P(B)$$

$$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B})$$

- teorema probabilità totali:

$$P(A) = \sum_{i=1}^{n} P(E_i)P(A|E_i)$$

VARIABILI CASUALI:

- funzione di probabilità:

$$P(x_i) = P(X = x_i)$$

- funzione di ripartizione:

$$F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} P(x)$$

VARIABILI CASUALI DISCRETE:

- valore atteso

$$E(X) = \sum_i x_i P(x_i) = \mu_X$$

- varianza:

$$V(X) = \sigma_X^2 = \sum_i (x_i - \mu_X)^2 P(x_i) \quad V(X) = E(X^2) - [E(X)]^2 = \mu_{X^2} - \mu_X^2$$

VARIABILI CASUALI CONTINUE

- valore atteso:

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x) dx$$

- varianza

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx$$

DISTRIBUZIONE DI PROBABILITA' DI VARIABILI CASUALI DISCRETE

- distribuzione variabile casuale uniforme discreta

$$P(x) = \frac{1}{s}$$

con valore atteso e varianza pari a

$$E(X) = a + \frac{s-1}{2} \qquad V(X) = \frac{s^2 - 1}{12}$$

- distribuzione di bernoulli:

$$P(x) = \pi^x (1 - \pi)^{1-x}$$

con valore atteso e varianza

$$E(X) = \pi$$

$$V(X) = \pi(1 - \pi)$$

- distribuzione binomiale:

$$P(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

con valore atteso e varianza pari a :

$$E(X) = n\pi$$

$$V(X) = n\pi(1-\pi)$$

Coefficiente binomiale: calcolo

Il coefficiente binomiale fornisce il numero di possibili **combinazioni** di n elementi presi x alla volta e si legge «n su x». In pratica, restituisce il numero di sequenze con x successi in n prove:

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

- $n!$ → rappresenta il prodotto dei primi n numeri naturali $n! = 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n$
- $0! = 1$ → per definizione

Esempio

Calcolare il numero di combinazioni di 5 $(n = 5)$ elementi presi 2 alla volta $(x = 2)$

$$\binom{5}{2} = \frac{5!}{2!\,(5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 3 \cdot 2 \cdot 1} = \frac{5 \cdot 4 \cdot \cancel{3!}}{2 \cdot 1 \cdot \cancel{3!}} = 10$$

<u>DISTRIBUZIONE DI VARIABILI CASUALI CONTINUE</u>

- distribuzione uniforme continua

$$f(x) = \begin{cases} \dfrac{1}{b-a} & se\ a \le x \le b \\ 0 & altrove \end{cases}$$

$$E(X) = \frac{a+b}{2} \qquad V(X) = \frac{(b-a)^2}{12}$$

- distribuzione normale

$$E(X) = \mu$$

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

$$V(X) = \sigma^2$$

standardizzazione della distribuzione normale

$$Z = \frac{X - \mu}{\sigma} \qquad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

per funzioni di ripartizione standardizzate di valori z negativi sfrutto le simmetrie:

$$P(Z \leq -z) = P(Z \geq z)$$

$$\Phi(-z) = 1 - \Phi(z)$$

Teorema del limite centrale:
somma:

$$S_n \sim N(n\mu, n\sigma^2) \quad \rightarrow Z_n = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \overset{d}{\rightarrow} Z \sim N(0,1)$$

media:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \rightarrow Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{d}{\rightarrow} Z \sim N(0,1)$$

Campionamento casuale semplice:

Campioni	**Estrazione**	
	con ripetizione	*senza ripetizione*
ordinati	N^n	$\frac{N!}{(N-n)!}$
non ordinati	$\frac{(N+n-1)!}{n!\,(N-1)!}$	$\frac{N!}{n!\,(N-n)!} \rightarrow \binom{N}{n}$

<u>DISTRIBUZIONI CAMPIONARIE</u>

- Statistica media campionaria:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

valori della distribuzione media campionaria→ campionamento con ripetizione:

il **valore atteso**: $E(X) = \mu_{\bar{X}} = \mu$

la **varianza**: $V(\bar{X}) = \sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}$

lo **scarto quadratico medio** (deviazione standard, errore standard o standard error):

$$SD(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

valori della distribuzione media campionaria→ campionamento senza ripetizione:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

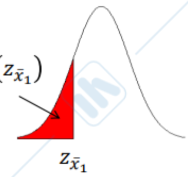
$$V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

standardizzazione della statistica media campionaria→ se la popolazione è finita e con distribuzione normale

$$X \sim N(\mu, \sigma^2) \;\rightarrow\; \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \;\rightarrow\; Z_{\bar{X}} \sim N(0,1)$$

dove: $Z_{\bar{X}} = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$P(Z_{\bar{X}} < z_{\bar{x}_1}) = F(z_{\bar{x}_1}) = \Phi(z_{\bar{x}_1})$



Media campionaria con popolazione non normale o incognita e n di grandi dimensioni:

$$X \sim ?\,(\mu, \sigma^2) \xrightarrow{n \to \infty} \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z_{\bar{X}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{n \to \infty} N(0,1)$$

$$\lim_{n \to \infty} P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z\right) = P(Z < z) = \Phi(z)$$

- Proporzione campionaria:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = P$$

con valori…

il **valore atteso**: $E(P) = \pi$

la **varianza**: $V(P) = \sigma_P^2 = \dfrac{\pi\,(1-\pi\,)}{n}$

lo **scarto quadratico medio** (deviazione standard, errore standard o standard error):

$$SD(P) = \sigma_P = \sqrt{\frac{\pi\,(1-\pi\,)}{n}}$$

Forma della distribuzione della proporzione campionaria

$$X \sim Ber(\pi) \xrightarrow{n \to \infty} P \sim N\left(\pi, \frac{\pi(1-\pi\,)}{n}\right)$$

$$Z_P = \frac{P-\pi}{\sqrt{\frac{\pi\,(1-\pi\,)}{n}}} \xrightarrow{n \to \infty} N(0,1)$$

Condizioni per l'applicazione del teorema del limite centrale

$$n > 20$$

$$n\pi > 5$$

$$n(1-\pi) > 5$$

<u>STIMA PUNTUALE</u>

stima puntuale e stimatori

Campione casuale
$X_1, X_2, \dots, X_n$
↓
$T_n = t(X_1, X_2, \dots, X_n)$
Stimatore puntuale
↓
variabile casuale

Campione osservato
$x_1, x_2, \dots, x_n$
↓
$t_n = t(x_1, x_2, \dots, x_n)$
Stima puntuale
↓
valore numerico

Proprietà degli stimatori per n finito:
- Correttezza:

$$E(T) = \theta$$

Distorsione:

$$B(T) = E(T) - \theta$$

- Errore di stima → errore quadratico medio

$$MSE(T) = E[(T-\theta)^2]$$

Scomposizione dell'errore quadratico medio:

$$MSE(T) = Var(T) + B(T)^2$$

dove $Var(T) = E\left[\left(T - E(T)\right)^2\right]$ è la **varianza di T**.

se lo stimatore è corretto…

$$MSE(T) = Var(T)$$

efficienza di uno stimatore:

$$MSE(T_1) < MSE(T_2) \qquad Var(T_1) < Var(T_2)$$ se stimatori corretti

Proprietà degli stimatori per n→infinito

- Consistenza:

$$\lim_{n\to\infty} MSE(T_n) = \lim_{n\to\infty} E(T_n-\theta)^2 = 0$$

se stimatore corretto…

$$\lim_{n\to\infty} Var(T_n) = 0$$

- Correttezza asintotica:

$$\lim_{n\to\infty} E(T_n) = \theta$$

Stima puntuale della varianza della popolazione: stimatore varianza campionaria corretta

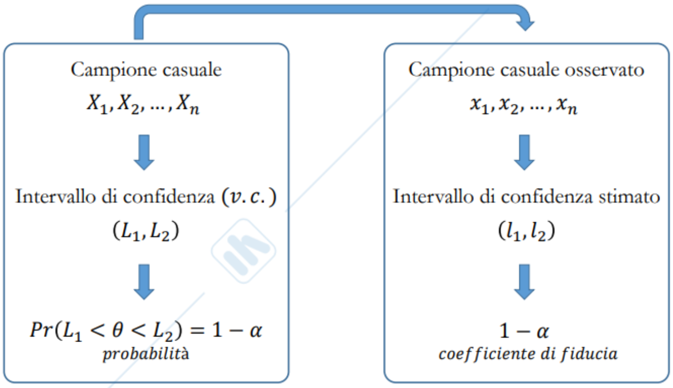
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

<u>STIMA PER INTERVALLO</u>

- margine di errore:

  $\theta = stima\ puntuale\ \pm\ margine\ di\ errore\ (\delta)$

- stimatore intervallo di confidenza e intervallo di confidenza stimato



INTERVALLO DI CONFIDENZA PER LA MEDIA DATA UNA POPOLAZIONE NORMALE CON VARIANZA NOTA:

- stimatore intervallo di confidenza:

$$Pr\left[\underbrace{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L_1} \le \mu \le \underbrace{\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L_2}\right] = 1 - \alpha$$

- intervallo di confidenza stimato:

$$\left[\underbrace{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{l_1}; \underbrace{\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{l_2}\right]$$

INTERVALLO DI CONFIDENZA PER LA MEDIA DATA UNA POPOLAZIONE NORMALE CON VARIANZA INCOGNITA:

Stimatore varianza corretta:

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

standardizzazione t-student:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

stimatore intervallo di confidenza:

$$Pr\left[\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

intervallo di confidenza stimato:

$$\left[\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}; \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right]$$

INTERVALLO DI CONFIDENZA DELLA MEDIA PER UNA POPOLAZIONE INCOGNITA CON VARIANZA INCOGNITA→ GRANDI CAMPIONI

Stimatore intervallo di confidenza:

$$P\left[\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right] = 1 - \alpha$$

intervallo di confidenza stimato:

$$\left[\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}; \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

INTERVALLO DI CONFIDENZA DELLA PROPORZIONE CON UNA POPOLAZIONE DI BERNOULLI:

Stimatore intervallo di confidenza:

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

intervallo di confidenza stimato

$$\left[p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}; p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$$

INTERVALLO DI CONFIDENZA PER LA MEDIA DI UNA POPOLAZIONE FINITA CON CAMPIONAMENTO SENZA RIPETIZIONE E VARIANZA NOTA:

Stimatore intervallo di confidenza:

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

intervallo di confidenza stimato:

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right]$$

DETERMINAZIONE DELLA NUMEROSITA' CAMPIONARIA:

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

<u>TEST D'IPOTESI:</u>

Sistema di ipotesi:

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$ ipotesi nulla semplice / ipotesi **alternativa bidirezionale**

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$ ipotesi alternativa **unilaterale a sinistra**

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$ ipotesi alternativa **unilaterale a destra**

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta = \theta_1 \end{cases}$$ due ipotesi semplici

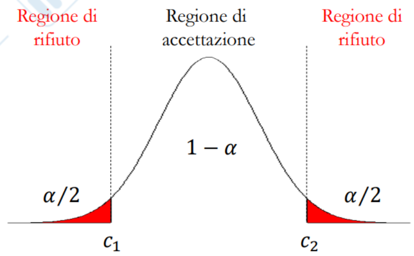
dove $\boldsymbol{\theta_0}$ è un **valore dato**, specificato dal problema.

REGIONE DI ACCETTAZIONE E RIFIUTO:

Test bilaterale

A seconda della statistica-test utilizzata, si hanno i seguenti **due valori critici** per la regione di accettazione e di rifiuto in un **test bilaterale:**

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

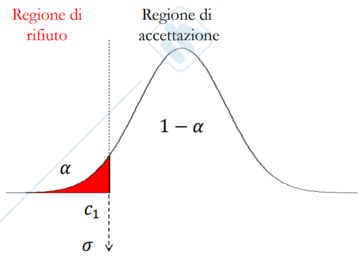


statistica-test $\bar{X}$	$\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	μ_0	$\mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
statistica-test $Z_{\bar{X}} = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$-z_{\alpha/2}$	0	$+z_{\alpha/2}$

Test unilaterale a sinistra

Invece, in un **test unilaterale a sinistra**, si ha un unico **valore critico** che separa la regione di accettazione da quella di rifiuto:

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

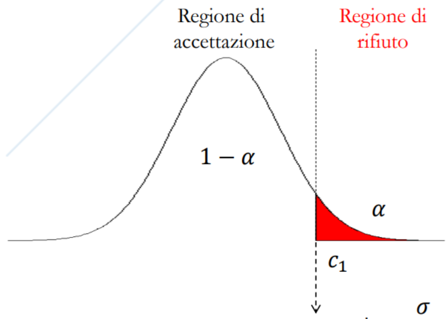


statistica-test $\bar{X}$	$\mu_0 - z_{\alpha}\frac{\sigma}{\sqrt{n}}$	μ_0
statistica-test $Z_{\bar{X}} = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$-z_{\alpha}$	0

Test unilaterale a destra

Infine, anche in un **test unilaterale a destra** si avrà un unico **valore critico** che separa la regione di accettazione da quella di rifiuto:

$$\begin{cases} H_0\colon \mu = \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$$



statistica-test $\bar{X}$ μ_0 $\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$

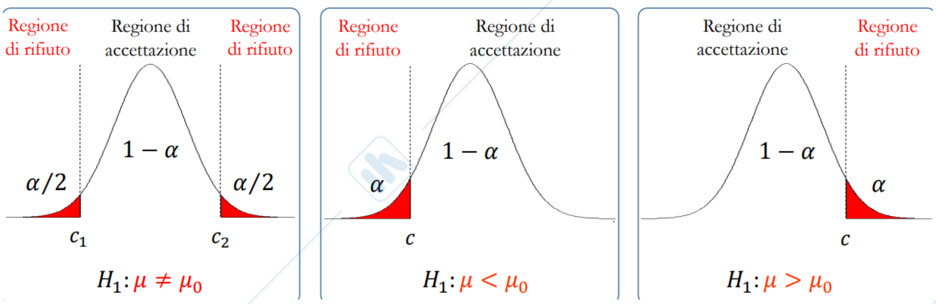
statistica-test $Z_{\bar{X}} = \frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ 0 $+z_\alpha$

TEST D'IPOTESI PER LA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA NOTA

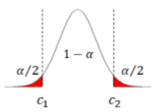
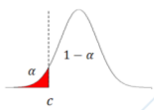
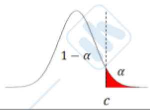
METODO DEL VALORE CRITICO

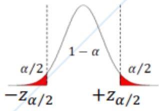
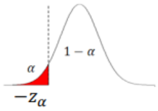
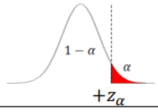
$$se\ t_n \in regione\ di\ accettazione \ \rightarrow\ H_0\ accettata$$

$$se\ t_n \in regione\ di\ rifiuto \ \rightarrow\ H_0\ rifiutata$$



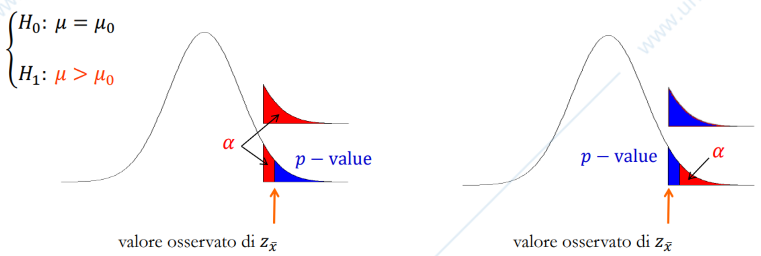
Le **regioni di rifiuto** quando si utilizza la **statistica-test $\bar{X}$** sono:

Ipotesi alternativa		Regione di rifiuto
$H_1\colon \mu \neq \mu_0$		$\bar{X} \le \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ o $\bar{X} \ge \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
$H_1\colon \mu < \mu_0$		$\bar{X} \le \mu_0 - z_{\alpha}\frac{\sigma}{\sqrt{n}}$
$H_1\colon \mu > \mu_0$		$\bar{X} \ge \mu_0 + z_{\alpha}\frac{\sigma}{\sqrt{n}}$

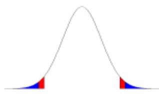
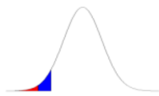
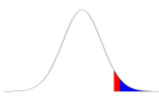
Ipotesi alternativa		**Regione di rifiuto**
H_1: $\mu \neq \mu_0$		$\|Z\| \geq z_{\alpha/2}$
H_1: $\mu < \mu_0$		$Z \leq -z_{\alpha}$
H_1: $\mu > \mu_0$		$Z \geq +z_{\alpha}$

METODO DEL P-VALUE

$$se\ \ p\text{−}value < \alpha \ \ \rightarrow \ \ rifiuto\ H_0$$



$p\text{−}value < \alpha \ \rightarrow \ \boldsymbol{rifiuto\ H_0}$ $\qquad$ $p\text{−}value > \alpha \ \rightarrow \ \boldsymbol{accetto\ H_0}$

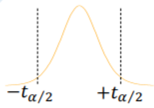
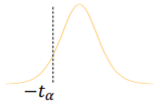
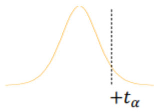
Ipotesi alternativa		***p*-value**
H_1: $\mu \neq \mu_0$		$2[1 - \Phi(z)]$
H_1: $\mu < \mu_0$		$\Phi(z)$
H_1: $\mu > \mu_0$		$1 - \Phi(z)$

TEST D'IPOTESI PER LA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA INCOGNITA

Statistica-test T:

$$T_{\bar{X}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Le **regioni di rifiuto** quando si utilizza la **statistica-test *T*** sono:

Ipotesi alternativa		Regione di rifiuto
H_1: $\mu \neq \mu_0$		$\lvert T\rvert \geq t_{\alpha/2}$
H_1: $\mu < \mu_0$		$T \leq -t_{\alpha}$
H_1: $\mu > \mu_0$		$T \geq +t_{\alpha}$

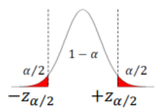
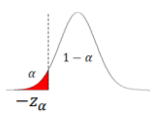
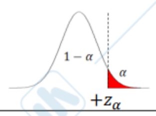
<u>TEST D'IPOTESI PER LA MEDIA DI UNA POPOLAZIONE NON NORMALE CON VARIANZA INCOGNITA</u>

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim N(0,1)$$

<u>TEST D'IPOTESI PER UNA PROPORZIONE</u>

$$Z_P = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$

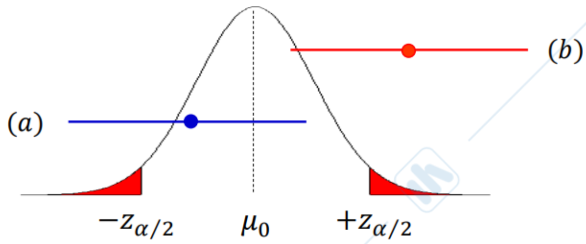
Le **regioni di rifiuto** per la **statistica-test Z_P** sono:

Ipotesi alternativa		Regione di rifiuto
H_1: $\pi \neq \pi_0$		$\lvert Z\rvert \geq z_{\alpha/2}$
H_1: $\pi < \pi_0$		$Z \leq -z_{\alpha}$
H_1: $\pi > \pi_0$		$Z \geq +z_{\alpha}$

<u>METODO DELL'INTERVALLO DI CONFIDENZA:</u>

$se\ l'intervallo\ di\ confidenza\ al\ livello\ (1-\alpha)\ \boldsymbol{comprende}\ \mu_0 \rightarrow \boldsymbol{accetto\ H_0}$

$altrimenti \rightarrow rifiuto\ H_0$



ERRORI DEL PRIMO E SECONDO TIPO

	Decisione	
	$\boldsymbol{Accetto\ H_0}$	$\boldsymbol{Rifiuto\ H_0}$
$\boldsymbol{H_0}$ è $\boldsymbol{vera}$	Corretta $1-\alpha$	Errore del I tipo α
$\boldsymbol{H_0}$ è $\boldsymbol{falsa}$	Errore del II tipo β	Corretta $1-\beta$

TEST D'INDIPENDENZA:

Quando **due variabili sono indipendenti**, la probabilità congiunta p_{ij} può essere espressa in termini delle probabilità marginali $p_{i.}$ e $p_{.j}$:

$$\begin{cases} H_0\colon p_{ij} = p_{i.} \cdot p_{.j} \\ H_1\colon p_{ij} \neq p_{i.} \cdot p_{.j} \quad \text{per almeno una combinazione } (i,j) \end{cases}$$

L'ipotesi di indipendenza può essere espressa anche **in termini di frequenze assolute**, calcolando le **frequenze teoriche in caso di indipendenza**, indicate con n^*_{ij}:

$$n^*_{ij} = \frac{n_{1.} \cdot n_{.1}}{n}$$

STATISTICA TEST: CHI QUADRO

$$\chi^2 = \sum_{i=1}^{H} \sum_{j=1}^{K} \frac{\left(n_{ij} - n^*_{ij}\right)^2}{n^*_{ij}}$$

$$se\ \ \chi^2_{osservato} < \chi^2_{(H-1)(K-1);\alpha} \rightarrow accetto\ H_0$$

$$se\ \ \chi^2_{osservato} > \chi^2_{(H-1)(K-1);\alpha} \rightarrow rifiuto\ H_0$$