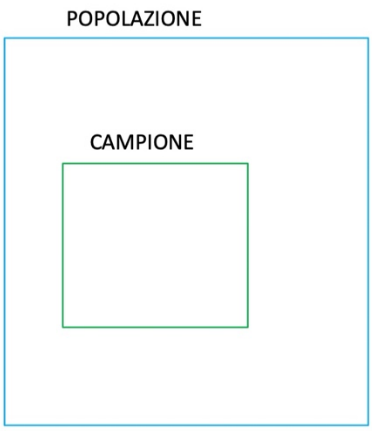
STATISTICA
FORMULARIO
INFERENZA

BY ANDREA IL MATEMATICO

INDICE

CAMPIONAMENTO

Selezioniamo con opportune tecniche un CAMPIONE all'interno di una popolazione
Il nostro scopo è quello di fare una stima su un certo PARAMETRO θ oppure effettuare un TEST DI IPOTESI.
Per farlo utilizziamo i dati dello STIMATORE S del camopione



POPOLAZIONE
N individui

CAMPIONE
n individui

θ = PARAMETRO

S = STIMATORE

S è lo stimatore (noto) del campione
che serve per stimare il parametro
θ della popopolazione (ignoto)

PARAMETRO, STIMATORE E STIMA

Il **PARAMETRO** è il vero dato che ci interessa conoscere di una popolazione ed è un valore che sintetizza i dati della popolazione.
Esempi di parametri sono: media, proporzione e varianza della popolazione.
Il parametro viene stimato mediante l'utilizzo di uno stimatore.

Lo **STIMATORE** è una variabile casuale descritta da una funzione che associa ad ogni possibile campione un valore del parametro da stimare.

La **STIMA** è il valore che assume lo stimatore in un particolare campione

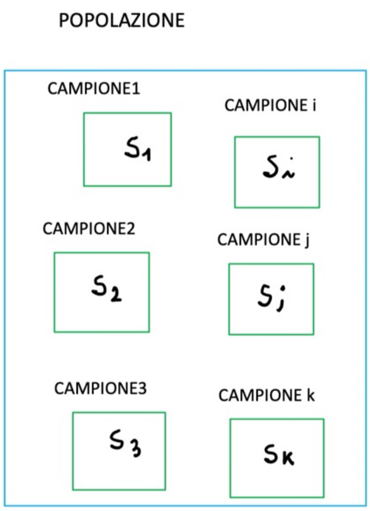
TANTI CAMPIONI

All'interno di una popolazione con N elementi possiamo avere tanti campioni con numerosità n in cui conta l'ordine.
In generale quando l'estrazione avviene con reimmissione vi sono:
N^n campioni
Ad esempio su una popolazione di 20 elementi possiamo estrarre oltre 3 milioni di possibili campioni

Ogni campione ha un suo stimatore S specifico per stimare il parametro θ
In generale diciamo che uno stimatore risulta corretto o non distorto quando il suo valore medio è uguale al parametro da stimare.
$E(S) = \theta$
Nel caso in cui non sia definiamo lo stimatore distorto.
Possiamo misurare inoltre la distorsione $B(t)$ come la differenza tra il valore medio dello stimatore e il parametro da stimare

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

STIMATORI DISTORTI E NON DISTORTI



θ = Vero PARAMETRO da stimare

$S_1, S_2 \ldots S_K$ = K possibili STIMATORE del parametro

STIMATORE

NON DISTORTO

$E(S_i) = \theta$

$\frac{\sum S_i}{K} = \theta$

$\frac{S_1 + S_2 + \cdots + S_K}{K} = \theta$

DISTORTO

$E(S_i) \neq \theta$

$\downarrow$

$B(S)$ = distorsione di S

$= E(S) - \theta$

Devo correggerlo

$\hat{S}$ = stimatore corretto

$E(\hat{S}) = \theta$

Nel caso della presenza di una distorsione possiamo valutare lo stimatore asintoticamente corretto quando all'aumentare della numerosità campionaria il valore della distorsione si riduce a zero

$\lim_{n \to \infty} B(t) = 0$

Tra due stimatori corretti preferiamo quello con varianza minima

$E(T_1) = E(T_2) = \theta$: $var(T_1) < var(T_2) \to T_1 \succ T_2$

ALCUNI STIMATORI SPECIFICI E STIME PUNTUALI

Si definisce stima puntuale è il procedimento che permette di assegnare ad un determinato parametro θ un solo valore

MEDIA CAMPIONARIA

$\bar{x} = \frac{\sum x_i}{n}$ è stimatore puntuale per la vera media μ

VARIANZA CAMPIONARIA CORRETTA

$s_C^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x_i}{n-1} - \bar{x}^2 \cdot \frac{n}{n-1}$ è stimatore per la varianza σ^2

PROPORZIONE CAMPIONARIA

$p = \frac{x}{n}$ è lo stimatore puntuale per la vera proporzione π

BETA DI REGRESSIONE (COEFFICIENTE ANGOLARE)

$B_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum x_i y_i - n\,\bar{x}\,\bar{y}}{\sum x_i^2 - n\,\bar{x}^2}$ è stimatore puntale per β_1

STIMA INTERVALLARE – INTERVALLI DI CONFIDENZA

Per parametri specifici come media, varianza, proporzione, chi quadrato e beta è possibile determinare delle stime intervallari mediante gli intervalli di confidenza.
In particolare un intervallo di confidenza è un intervallo entro cui abbiamo una certa probabilità $\mathbf{1-\alpha}$ di trovare il vero valore del parametro.
Sarebbe più corretto dire che estratto un numero sempre maggiori di campioni con una stessa numerosità una percentuale pari $\mathbf{1-\alpha}$ avrà la stima del vero parametro θ all'interno di tale intervallo.

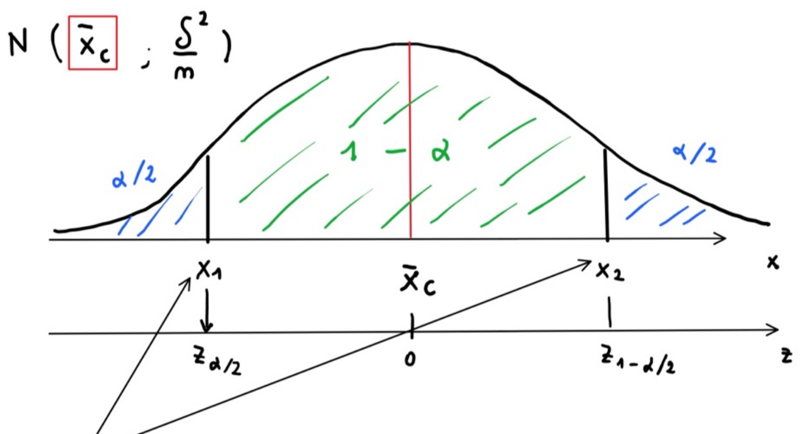
STIMA INTERVALLARE DI UNA MEDIA

Quando vogliamo cercare un intervallo di confidenza per una media la prima cosa da chiederci è se **conosciamo la vera varianza** σ^2 della popolazione.
Nel caso in cui la conosciamo usiamo la funzione **normale standardizzata** per determinare l'intervallo.
Se invece la **varianza è ignota** utilizziamo come sostituto la varianza campionaria e per stimare l'intervallo ricorriamo alla distribuzione t-student con $n-1$ gradi di libertà

VARIANZA NOTA
L'intervallo di confidenza a livello $1-\alpha$ per la vera media μ è dato da:

$$IC_\mu(1-\alpha) = (x_1, x_2) = \left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \;, \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Dove $\frac{\sigma}{\sqrt{n}}$ rappresenta lo standard error SE delle medie campionarie di ampiezza n
Per trovare i valori della normale standardizzata utilizziamo le apposite tavole statistiche



$$X_{1,2} = \bar{x}_c \pm z_{1-\frac{\alpha}{2}} \frac{\delta}{\sqrt{m}}$$

$$P(x_1 < \mu < x_2) = 1 - \alpha$$

VARIANZA IGNOTA
Quando la varianza è ignota riproponiamo lo stesso calcolo, ma:
al posto della varianza σ^2 usiamo la **varianza campionaria corretta**
$s_C^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x_i}{n-1} - \bar{x}^2 \cdot \frac{n}{n-1}$
E al posto della normale standardizzata utilizziamo la **t-student** con $\boldsymbol{n-1}$ **gradi di libertà**
L'intervallo di confidenza per la vera media risulta essere:

$$IC_\mu(1-\alpha) = (x_1, x_2) = \left(\bar{x} - t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s_C}{\sqrt{n}} \ , \bar{x} + z_{1-\frac{\alpha}{2},t-1} \cdot \frac{s_C}{\sqrt{n}} \right)$$

Per trovare i valori della t-student utilizziamo le apposite tavole statistiche

STIMA INTERVALLARE DI UNA PROPORZIONE

Per trovare l'intervallo di confidenza a livello $1-\alpha$ per la proporzione π usiamo sempre la distribuzione **normale standardizzata**

$$IC_\pi(1-\alpha) = (p_1, p_2) = \left(p_c - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_c \cdot (1-p_c)}{n}} \ , p_c + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Dove $\sqrt{\frac{p_c \cdot (1-p_c)}{n}}$ rappresenta lo standard error (SE) delle proporzioni campionarie di ampiezza n

STIMA INTERVALLARE DI UNA VARIANZA

Per trovare l'intervallo di confidenza a livello $1-\alpha$ per la proporzione π usiamo la **distribuzione chi-quadrato con** $\boldsymbol{n-1}$ **gradi di liberta**

$$IC_\pi(1-\alpha) = (\sigma_1^2, \sigma_2^2) = \left(s_c^2 \cdot \frac{n-1}{\chi^2_{1-\frac{\alpha}{2},n-1}} \ , s_c^2 \cdot \frac{n-1}{\chi^2_{\frac{\alpha}{2},n-1}} \right)$$

Per i valori della distribuzione chi-quadrato usiamo le apposite tavole
Attenzione che la distribuzione chi-quadrato non è simmetrica!

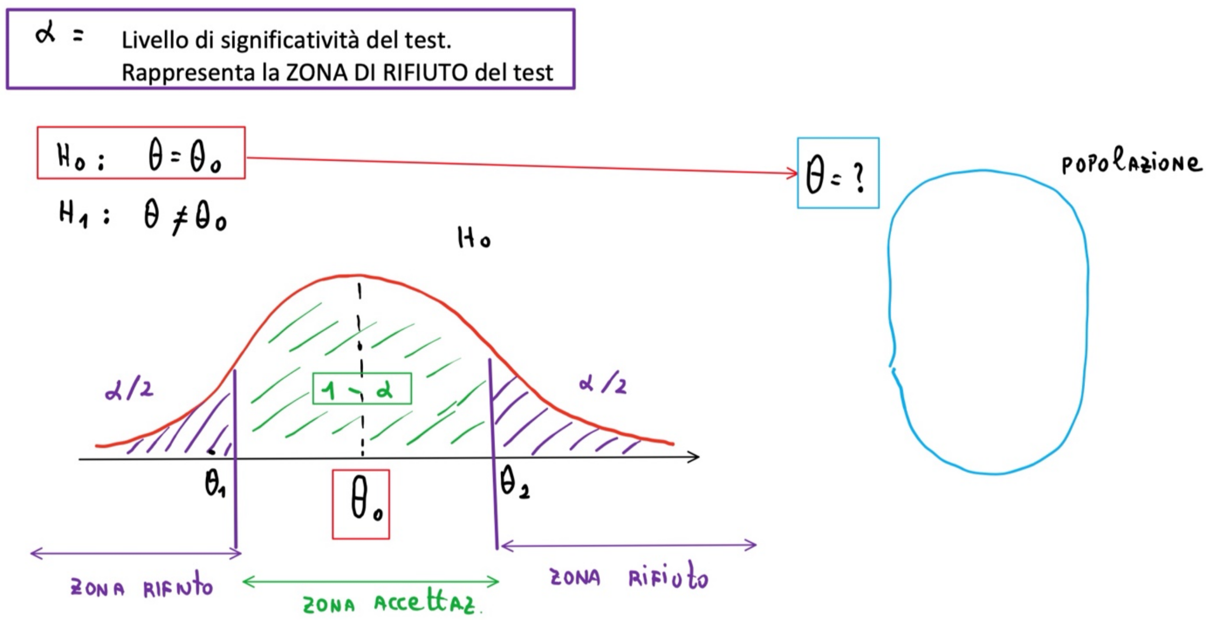
TEST DI IPOTESI

I **test di ipotesi** sono test statistici dove si contrappongono **due ipotesi**.
Dal un lato troviamo l'**ipotesi nulla $\boldsymbol{H_0}$** secondo cui il valore del parametro ignoto θ è uguale a quello ipotizzato θ_0
Molto spesso l'ipotesi nulla rappresenta una credenza comunemente accettata ed è quindi uno *status quo* delle cose
In contrapposizione troviamo l'**ipotesi alternativa $\boldsymbol{H_1}$** secondo la quale il parametro assumerebbe valori difformi: diversi, minori o maggiori rispetto a quelli ipotizzati in H_0.
In questo senso H_1 rappresenta la novità o la rivoluzione.

TEST DI IPOTESI: H_0: $\theta = \theta_0$ $\qquad H_1$: $\theta \neq \theta_0$, $\theta < \theta_0$, $\theta > \theta_0$

A seconda che l'ipotesi H_1 richieda che il parametro sia diverso, minore o maggiore del parametro ipotizzato parliamo rispettivamente di **test bilaterale**, **monolaterale sinistro** e **monolaterale destro**.
I test di ipotesi si svolgono rispetto ad un certo valore di significatività α che rappresenta quella soglia che stabilisce se un risultato può essere statisticamente significativo.
La zona associata ad α è quindi quella associata all'evidenza dell'ipotesi alternativa H_1

TEST BIDIREZIONALE
La zona di rifiuto cade metà sulla parte destra e metà sulla parte sinistra.
NON sono accettati ne valori troppo grandi ne valori troppo piccoli



TEST DI IPOTESI SU UNA MEDIA

Nel test di ipotesi su una media si contrappongono due ipotesi
L'ipotesi nulla H_0 che sostiene che un valore della media μ è quello ipotizzato μ_0

H_0: $\mu = \mu_0$

A seconda che il test sia bilaterale, monolaterale destro o monolaterale sinistro l'ipotesi alternativa H_1 risulta:
H_1: $\mu \neq \mu_0$ H_1: $\mu < \mu_0$ H_1: $\mu > \mu_0$

Il test è posto su un certo **valore di significatività α** cui è associata la regione di rifiuto dell'ipotesi nulla

La prima distinzione la facciamo a seconda che la varianza σ^2 della popolazione sia nota o ignota.

TEST SU UNA MEDIA

→ VARIANZA DELLA POPOLAZIONE NOTA → Uso la Normale

$N(\mu_0; \frac{\sigma^2}{n})$

→ VARIANZA DELLA POPOLAZIONE NON NOTA → Usiamo la Normale T-Student

$t_{n-1}(\mu_0; \frac{\hat{S}^2}{n})$

n-1 gradi di libertà

Usiamo la **VARIANZA CAMPIONARIA CORRETTA**

↓

Se facciamo i calcoli a mano, e quindi abbiamo bisogno di utilizzare le tavole

→ CAMPIONE PICCOLO
n<30 (n<40) (n<60)
(dipende dalle tavole di cui si dispone

↓

Usiamo la t-Student, con n-1 gradi di liberta (g.d.l.)

$N_T(\mu_0; \frac{\hat{S}^2}{n})$

n-1 gradi di libertà

→ CAMPIONE GRANDE
n>30 (n>40) (n>60)
(dipende dalle tavole di cui si dispone

↓

Usiamo la NORMALE

$N(\mu_0; \frac{\hat{S}^2}{n})$

TEST BILATERALE CON VARIANZA NOTA

Quando la **varianza è nota** usiamo la distribuzione normale standardizzata.

Nel test bilaterale la **zona di accettazione** dell'ipotesi nulla in questo caso è:

$$I_{H_0} = (x_1, x_2) = \left(\mu_0 - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \;, \mu_0 + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Accettiamo l'ipotesi nulla se la media campionaria $\bar{x}_C$ appartiene a questo intervallo

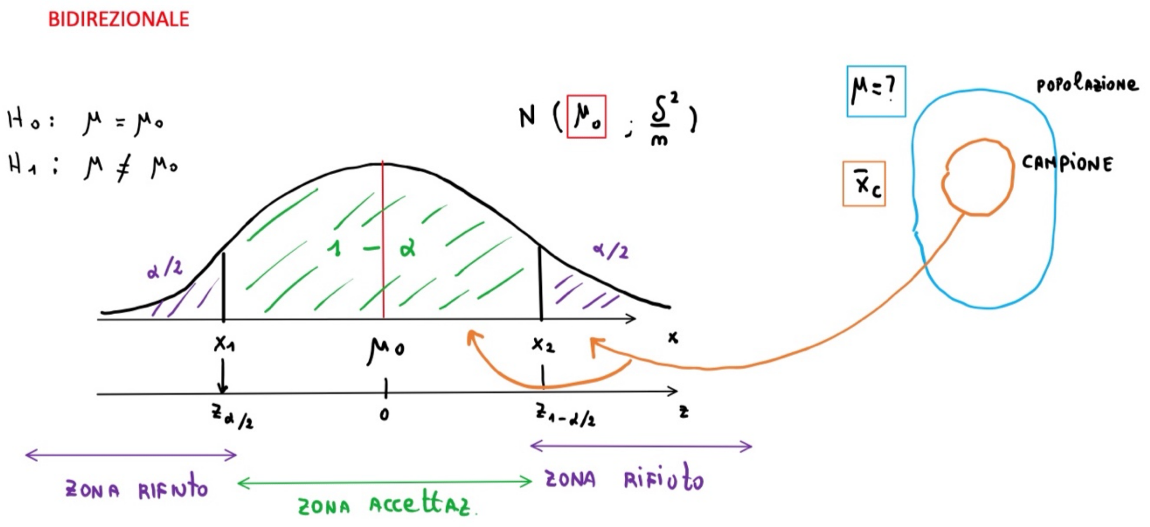
$$\bar{x}_C \in (x_1, x_2) = \left(\mu_0 - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \;, \mu_0 + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

In altre parole la distanza (valore assoluto) tra la media campionaria $\bar{x}_C$ e la media ipotizzata μ_o non deve eccedere z volte l'errore standard ($SE = \frac{\sigma}{\sqrt{n}}$)

$$|\bar{x}_C - \mu_0| < z_{1-\frac{\alpha}{2}} \cdot SE = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{ accettiamo } H_0$$

Diversamente **rifiutiamo l'ipotesi nulla**, quindi accettiamo quella alternativa

$$|\bar{x}_C - \mu_0| > z_{1-\frac{\alpha}{2}} \cdot SE = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{ rifiutiamo } H_0$$



VARIANZA NOTA

$$x_{1,2} = \mu_0 \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$P(x_1 < \bar{x}_c < x_2) = 0,99$$

VARIANZA IGNOTA

$$x_{1,2} = \mu_0 \pm t_{1-\frac{\alpha}{2};\, n-1} \frac{\hat{s}}{\sqrt{n}}$$

TEST BILATERALE CON VARIANZA IGNOTA

Quando la varianza della popolazione σ^2 è ignota usiamo la **varianza campionaria corretta** del campione $\boldsymbol{s_C^2}$

$$s_C^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x_i}{n-1} - \bar{x}^2 \cdot \frac{n}{n-1}$$
Inoltre usiamo la distribuzione **t-student con $n-1$ gradi di libertà**

Nel test bilaterale la **zona di accettazione** dell'ipotesi nulla in questo caso è:

$$I_{H_0} = (x_1, x_2) = \left(\mu_0 - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s_C}{\sqrt{n}}\ , \mu_0 + z_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s_C}{\sqrt{n}}\right)$$
Accettiamo l'ipotesi nulla se la media campionaria $\bar{x}_C$ appartiene a questo intervallo
$$\bar{x}_C \in (x_1, x_2) = \left(\mu_0 - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s_C}{\sqrt{n}}\ , \mu_0 + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s_C}{\sqrt{n}}\right)$$
In altre parole la distanza (valore assoluto) tra la media campionaria $\bar{x}_C$ e la media ipotizzata μ_o non deve eccedere t volte l'errore standard ($SE = \frac{\sigma}{\sqrt{n}}$)

$$|\bar{x}_C - \mu_0| < t_{1-\frac{\alpha}{2}, n-1} \cdot SE = t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{ accettiamo } H_0$$
Diversamente **rifiutiamo l'ipotesi nulla**, quindi accettiamo quella alternativa

$$|\bar{x}_C - \mu_0| > t_{1-\frac{\alpha}{2}, n-1} \cdot SE = t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{ rifiutiamo } H_0$$

STATISTICA TEST E IL P-VALUE

La statistica test insieme al p-value costituiscono gli elementi più importanti di un test di ipotesi.
Nel caso del test sulla media la **statistica-test** indica il dato campionario standardizzato.
Essa è pari alla differenza tra la media campionaria e quella ipotizzata divisa per l'errore standard

$$z - test = \frac{\bar{x}_C - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad o \quad t - test = \frac{\bar{x}_C - \mu_0}{\frac{s_C}{\sqrt{n}}}$$
Possiamo usare la statistica test per determinare l'esito di un test sulla base di un livello alfa.
In particolare se:

$$t - test \in \left(-t_{1-\frac{\alpha}{2}, n-1}\ , \mu_0 + t_{1-\frac{\alpha}{2}, n-1}\right)$$

il p-value viene valutato sulla base della statistica test.
Esso indica il maggior valore di alfa per cui accettiamo l'ipotesi nulla.

Nel caso di test bilaterale è il doppio dell'area che si trova a destra del valore assoluto della statistica test compresa tra la t-student e l'asse t

$$p - value = 2 \cdot \int_{|t-test|}^{+\infty} f(t, n-1)\, dt$$
Dove $f(t, n-1)$ indica la funzione t-student con $n-1$ gradi di libertà.

Nel caso di test monolaterale sinistro è l'area a sinistra della statistica test

$$p - value = \int_{-\infty}^{t-test} f(t, n-1)\, dt$$

Nel caso di test monolaterale destro è l'area a destra della statistica test

$$p - value = \int_{t-test}^{+\infty} f(t, n-1)\, dt$$

Quando il p-value è minore di α rifiutiamo l'ipotesi nulla

$$p - value < \alpha \ \rightarrow \text{rifiutiamo } H_0$$

In generale minore è il valore del p-value minore è la probabilità di accettare l'ipotesi nulla. Maggiori dunque sono le probabilità che vi sia evidenza empirica della significatività dell'ipotesi alternativa.

TEST MONOLATERALI

Per i test bilaterali riportiamo le zone di accettazione dell'ipotesi nulla con la t-student.
(Vale sempre il discorso di varianza nota e ignota

Nel **test monolaterale sinistro**:

$H_0: \mu = \mu_0 \qquad H_1: \mu < \mu_0$

$$\text{se } \bar{x}_C > \mu_0 - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s_C}{\sqrt{n}} \rightarrow \quad \text{accettiamo } H_0$$

Nel **test monolaterale destro**:

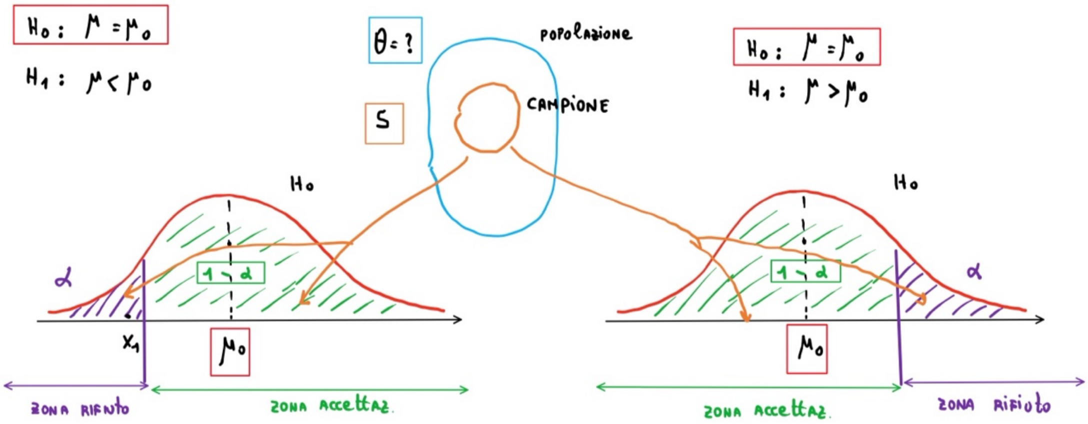
$H_0: \mu = \mu_0 \qquad H_1: \mu < \mu_0$

$$\text{se } \bar{x}_C < \mu_0 + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s_C}{\sqrt{n}} \rightarrow \quad \text{accettiamo } H_0$$

TEST BIDIREZIONALE

SINISTRO

DESTRO



VARIANZA NOTA

$$x_1 = \mu_0 - z_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$$

VARIANZA IGNOTA

$$x_1 = \mu_0 - t_{1-\alpha;\, n-1} \frac{\hat{S}}{\sqrt{n}}$$

VARIANZA NOTA

$$x_2 = \mu_0 + z_{1-\alpha} \cdot \frac{S}{\sqrt{n}}$$

VARIANZA IGNOTA

$$x_2 = \mu_0 + t_{1-\alpha;\, n-1} \frac{\hat{S}}{\sqrt{n}}$$

TEST SU UNA PROPORZIONE

Il test di ipotesi su una proporzione si presenta nei seguenti modo.

TEST BILATERALE

Nel test bilaterale:

$$H_0\colon\ \pi = \pi_0 \qquad\qquad H_1\colon\ \pi \neq \pi_0$$

Accettiamo l'ipotesi nulla se la proporzione campionaria p_c si trova nella zona di accettazione

$$p_c \in (p_1, p_2) = \left(\pi_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0\,(1-\pi_0)}{n}}, \qquad \pi_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0\,(1-\pi_0)}{n}} \right)$$

Dove l'errore standard è: $SE = \sqrt{\frac{\pi_0\,(1-\pi_0)}{n}}$

NB: l'errore standard è calcolato in riferimento alla proporzione ipotizzata e non a quella campionaria come avviene nell'intervallo di confidenza!

STATISTICA TEST E P-VALUE

La statistica test è data dal rapporto tra la differenza tra la proporzione campionaria e quella ipotizzata, e l'errore standard (SE)

$$z - test = \frac{p_C - \pi_0}{SE} = \frac{p_C - \pi_0}{\sqrt{\frac{\pi_0\ (1 - \pi_0)}{n}}}$$

In base alla statistica test possiamo accettare o meno l'ipotesi nulla:

Accettiamo l'ipotesi nulla se

$$z - test \in \left(-z_{1-\frac{\alpha}{2}} \cdot, \qquad z_{1-\frac{\alpha}{2}}\right)$$

Il **p-value** è la doppia area a destra dello z-test compresa tra la normale standardizzata e l'asse z

$$p - value = 2 \cdot \int_{|z-test|}^{+\infty} f(z)\, dz$$

Rifiutiamo il test quando il p-value è inferiore ad α

TEST MONOLATERALI SINISTRO E DESTRO SULLA PROPORZIONE

Nei monolaterali sinistro e destro l'ipotesi alternativa risulta:

$$H_1:\ p_c < \pi \qquad\qquad H_1:\ p_c > \pi$$

Accettiamo l'ipotesi nulla rispettivamente se

$p_c < \pi_0 - z_{1-\alpha} \cdot \sqrt{\frac{\pi_0\ (1-\pi_0)}{n}}$ (test monolaterale sinistro)

$p_c > \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0\ (1-\pi_0)}{n}}$ (test monolaterale destro)

Oppure sulla base della statistica test se:

$t - test < -z_{1-\alpha}$ (test monolaterale sinistro)

$t - test > +z_{1-\alpha}$ (test monolaterale destro)

I p-value ammontano a

$$p - value = \int_{-\infty}^{t-test} f(z)\, dz \quad = 1 - \oint_z (t - test) \ \ \text{(monolaterale sx)}$$

$$p - value = \int_{z-test}^{+\infty} f(z)\, dz \qquad \text{(monolaterale dx)}$$

TEST DI IPOTESI SULLA VARIANZA

Nel test di ipotesi su una varianza si contrappongono due ipotesi
L'ipotesi nulla H_0 che sostiene che un valore della varianza σ^2 è quello ipotizzato σ_0^2

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

A seconda che il test sia bilaterale, monolaterale destro o monolaterale sinistro l'ipotesi alternativa H_1 risulta:
H_1: $\sigma^2 \neq \sigma_0^2$ $\quad$ H_1: $\sigma^2 < \sigma_0^2$ $\quad$ H_1: $\sigma^2 > \sigma_0^2$

Il test è posto su un certo **valore di significatività $\boldsymbol{\alpha}$** cui è associata la regione di rifiuto dell'ipotesi nulla

ACCETTAZIONE DEL TEST

Definiamo s_C^2 la **varianza campionaria** corretta:

$$s_C^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum x_i}{n-1} - \bar{x}^2 \cdot \frac{n}{n-1}$$

Per stabilire l'esito del test usiamo la **funzione chi-quadrato (χ^2) con n-1 gradi di libertà**

Accettiamo il **test bilaterale** se

$$s_C^2 \in (\sigma_1^2, \sigma_2^2) = \left(\sigma_0^2 \cdot \frac{\chi_{\frac{\alpha}{2},n-1}}{n-1};\ \sigma_0^2 \cdot \frac{\chi_{1-\frac{\alpha}{2},n-1}}{n-1}\right)$$

Mentre nei casi **mono-lateri sinistro e destro** rispettivamente se:

$$s_C^2 > \sigma_0^2 \cdot \frac{\chi_{\frac{\alpha}{2},n-1}}{n-1}\ (\text{sx}) \qquad s_C^2 < \sigma_0^2 \cdot \frac{\chi_{1-\frac{\alpha}{2},n-1}}{n-1}\ (\text{dx})$$

La **statistica-test** per il test di ipotesi sulla varianza è:

$$\chi^2 - test = \frac{s_C^2}{\sigma_0^2} \cdot (n-1)$$

Usando il **chi-quadro-test** possiamo stabilire l'esito del test: accettiamo H_0 se:

$$\chi^2 - test \in \left(\chi_{\frac{\alpha}{2},n-1}, \chi_{1-\frac{\alpha}{2},n-1}\right)\ (\text{bilat})$$

$$\chi^2 - test > \chi_{\frac{\alpha}{2},n-1}\ (\text{sx}) \quad \chi^2 - test < \chi_{1-\frac{\alpha}{2},n-1}\ (\text{dx})$$

Il **p-value** risulta l'area esterna al $\chi^2 - test$ (doppia nel caso bilaterale)

TEST DI IPOTESI SUL BETA DI REGRESSIONE (COEFFICIENTE ANGOLARE)

Nel test di ipotesi sul beta di regressione β_1 (coefficiente angolare) si contrappongono due ipotesi L'ipotesi nulla H_0 che sostiene che un valore della varianza σ^2 è quello ipotizzato σ_0^2

H_0: $\beta_1 = \beta_{1_0}$

A seconda che il test sia bilaterale, monolaterale destro o monolaterale sinistro l'ipotesi alternativa H_1 risulta:
H_1: $\beta_1 \neq \beta_{1_0}$ $\quad$ H_1: $\beta_1 < \beta_{1_0}$ $\quad$ H_1: $\beta_1 > \beta_{1_0}$

Il test è posto su un certo **valore di significatività α** cui è associata la regione di rifiuto dell'ipotesi nulla

ACCETTAZIONE DEL TEST

Per stabilire l'esito del test utilizziamo la funzione t-student con $n - 2$ gradi di libertà.

Definiamo B_1 il beta tratto dai dati campionari

$$B_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum x_i y_i - n\,\bar{x}\,\bar{y}}{\sum x_i^2 - n\,\bar{x}^2}$$

Dove: S_{XY} è la codevianza tra le x e le y, $\quad S_{XX}$ è la devianza delle x

Accettiamo il **test bilaterale** se

$$B_1 \in (\beta_{11}, \beta_{12}) = \left(\beta_{1_0} - t_{1-\frac{\alpha}{2},n-2} \cdot SE;\ \beta_{1_0} - t_{1-\frac{\alpha}{2},n-2} \cdot SE\right)$$

In particolare l'errore standard sul beta1 viene calcolato nel seguente modo:

$$SE = \sqrt{\frac{s_\varepsilon^2}{S_{XX}}} = \sqrt{\frac{S_{YY} - B_1 \cdot S_{XY}}{S_{XX}}}$$

Dove s_ε^2 è la varianza degli errori di regressione, $\quad S_{YY}$ è la devianza delle y

Mentre nei casi **mono-lateri sinistro e destro** rispettivamente se:

$$B_1 > \beta_{1_0} - t_{1-\frac{\alpha}{2},n-2} \cdot SE \text{ (sx)} \qquad s_C^2 < \beta_{1_0} - t_{1-\frac{\alpha}{2},n-2} \cdot SE \text{ (dx)}$$

La **statistica-test** per il test di ipotesi sul beta angolare è:

$$t - test = \frac{B_1 - \beta_{1_0}}{SE}$$

Usando il **chi-quadro-test** possiamo stabilire l'esito del test: accettiamo H_0 se:

$$t - test \in \left(-t_{1-\frac{\alpha}{2},n-2}, +t_{1-\frac{\alpha}{2},n-2}\right) \text{ (bilat)}$$

$$\chi^2 - test > -t_{1-\frac{\alpha}{2},n-2} \text{ (sx)} \quad \chi^2 - test < t_{1-\frac{\alpha}{2},n-2} \text{ (dx)}$$

Il **p-value** risulta l'area esterna al $t - test$ (doppia nel caso bilaterale)

CONFRONTO TRA MEDIE

Un interessantissimo caso di applicazione del test di ipotesi riguarda il confronto tra medie.
In particolare il test di ipotesi si basa sul fatto che le medie delle due popolazioni di riferimento per i campioni siano la stessa.
In altre parole la differenza tra le medie delle popolazioni è pari a zero:

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

Mentre come ipotesi alternative abbiamo al solito:

$$H_1: \mu_2 - \mu_1 = 0 \ \ (\text{bil.}) \qquad H_1: \mu_2 - \mu_1 < 0 \ \ (\text{monol. dx}) \qquad H_1: \mu_2 - \mu_1 > 0 \ \ (\text{monol. sx})$$

VARIANZA NOTA O IGNOTA
La teoria generale sul test delle medie rimane invariata.
Quando la varianza delle popolazioni di riferimento risultano note procediamo con l'uso della normale standardizzata.
Diversamente quando le varianze risultano ignote si procede con la distribuzione t-student.
Circa i gradi di libertà vedremo le da utilizzare e le vedremo nelle diverse casistiche.

CAMPIONI DIPENDENTI E INDIPENDENTI

La prima grande distinzione che facciamo all'interno di questa teoria è se questo test riguarda due campioni che sono indipendenti o dipendenti (appaiati)

CAMPIONI INDIPENDENTI

Una prima distinzione che facciamo per i campioni dipendenti è se conosciamo o non conosciamo le varianze delle popolazioni di riferimento.

VARIANZE POPOLAZIONI NOTE

Se le conosciamo usiamo la normale standardizzata.
Qui possiamo avere due situazioni: le varianze sono uguali oppure diverse.
Se le **varianze sono uguali** usiamo il seguente errore standard:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \rightarrow \ SE = \sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Se le **varianze sono diverse** tra di loro abbiamo una formula più generale per il calcolo di SE:

$$\sigma_1^2 \neq \sigma_2^2 = \sigma^2 \rightarrow \ SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

VARIANZE POPOLAZIONI IGNOTE
Quando le varianze delle popolazioni di riferimento sono ignote bisogna procedere al **test F sulla varianza** per capire se possiamo considerarle uguali oppure no.

Al tal fine si utilizzano le due varianze campionarie corrette dei campioni s_1^2 e s_2^2

Quando l'esito del test F ci dice co possiamo considerare le **varianze uguali** allora calcoliamo la varianza media:

$$\overline{\sigma^2} = \frac{s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1)}{(n_1 + 1) + (n_2 + 1)}$$

L'errore standard che usiamo per il test è:

$$SE = \bar{\sigma} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Per la t-student usiamo $n_1 + n_2 - 2$ gradi di libertà.

Se l'esito del test F sulla varianza ci dice che dobbiamo considerare le varianze diverse allora l'errore standard diventa:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

In questo caso i gradi di libertà che utilizziamo per la distribuzione t-student seguono la seguente formula

$$\text{g. d. l.} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$