Formulario Test e ipotesi

Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza nota.
Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza NON nota.
Errore di prima specie
Errore di seconda specie
Regione critica del test
Osservazione:
Livello di significatività
Confronto tra due regioni critiche
P-value o P-dei-dati
P-value nel caso di una variabile normale standard
P-value compreso tra quali valori
Verifica delle ipotesi sulla media di una popolazione normale con varianza nota.
a) Test bilaterali
a.1) Test bilaterali → funzione potenza (decisione corretta)
b) Test unilaterale destro
c) test unilaterale sinistro
Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza nota.
Verifica delle ipotesi sulla media di una popolazione normale con varianza NON nota: il t-test
a) t - Test bilaterale
b) t - test unilaterale destro
b) t - test unilaterale sinistro
Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza NON nota.
Verifica delle ipotesi sulla differenza delle medie di due campioni normali (numerosità molto grandi con varianza nota e con varianza NON nota)
a) se le varianza σ_x^2 e σ_y^2 sono note:
a) se le varianza σ_x^2 e σ_y^2 NON sono note:

Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza nota.

Se $X_1, \dots, X_n$ è un campione estratto da una popolazione $\mathcal{N}(\mu, \sigma^2)$ con σ^2 **nota** e $\vec{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, allora otteniamo questa tabella:

H_0	H_1	Statistica del test, Z	Si rifiuta H_0 con livello di signficativà α se :	p-value (o p-dei-dati) se $Z = z$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\left\lvert\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right\rvert > z_{\frac{\alpha}{2}}$	$2P(Z \geq \lvert z \rvert)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$	$P(Z \geq z)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$	$P(Z \leq z)$

dove

$$Z \sim \mathcal{N}(0,1)$$

$$Z = \frac{\vec{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$z_{\frac{\alpha}{2}} = \Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

$$z_{\alpha} = \Phi(z_{\alpha}) = 1 - \alpha$$

📌 Un'ipotesi viene rifiutata se nel rispettivo intervallo di confidenza, il valor della media relativa all'ipotesi, non è compresa nell'intervallo!

▼ Esempio

2. Sapendo che la varianza della FE è pari a 2.5^2, costruire un intervallo di confidenza bilatero di livello 95% per la FE media.

   Soluzione: $IC_{0.95}(\mu) = [\bar{X} \pm z_{0.975} \cdot \frac{\sigma}{\sqrt{n}}] = [56.847 \pm 1.96 \cdot \frac{2.5}{\sqrt{10}}] = [55.297; 58.396]$

3. È possibile affermare al 5% che la valvola ristabilisca una FE diversa da quella target di quelle attualmente in commercio, pari a 55? E all'1%?

   Soluzione: Nel primo caso, non è necessario eseguire il test, ma si può sfruttare il precedente risultato dell'intervallo di confidenza. Dal momento che il valore target $\mu_0 = 55$ non è contenuto nell'intervallo, posso affermare che rifiuto H_0 a livello $\alpha = 0.05$. Nel caso di $\alpha = 0.01$, invece, eseguo la procedura, ottenendo una regione critica $R_{0.01} = (\bar{X} < 55 - 2.57 \cdot 2.5/\sqrt{(10)} = 52.968 \cup \bar{X} > 55 + 2.57 \cdot 2.5/\sqrt{(10)} = 57.032$. Dal momento che 56.847 risulta compreso tra i due limiti, non ho evidenza sufficiente per rifiutare H_0, come si evince anche dall'intervallo di confidenza $IC(\mu)_{0.95} = [54.81; 58.88]$.

In questo caso il valore 55 non è nell'intervallo di confidenza $[55.297; 58.396]$ → quindi non possiamo accettarlo. Se facciamo i calcoli questo ci viene confermato.

Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza NON nota.

Se $X_1, \ldots, X_n$ è un campione con distribuzione $\mathcal{N}(\mu, \sigma^2)$ con σ^2 **NON nota** e $\vec{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, allora otteniamo questa tabella:

H_0	H_1	Statistica del test, T_{n-1}	Si rifiuta H_0 con livello di signficatività α se :	p-value (o p-dei-dati) se $T_{n-1} = t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\left\lvert\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right\rvert > t_{\frac{\alpha}{2},n-1}$	$2P(T_{n-1} \geq \lvert t\rvert)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > t_{\alpha,n-1}$	$P(T_{n-1} \geq t)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} < -t_{\alpha,n-1}$	$P(T_{n-1} \leq t)$

dove

$$T_{n-1} \sim t_{n-1}, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$T_{n-1} = \frac{\vec{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$t_{\frac{\alpha}{2},n-1} = \phi_{T_{n-1}}(t_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

$$t_{\alpha,n-1} = \phi_{T_{n-1}}(t_{\alpha}) = 1 - \alpha$$

Supponiamo di avere una variabile aleatoria X che ha una **funzione di densità**:

$$f(x, \theta) \qquad \theta \in \Omega$$

$\theta \in$ insieme Ω

- Esempio:
  - $f(x,\theta) = \theta e^{-\theta x}\mathbf{1}_{[0;\infty)}(x); \theta = \lambda$ (variabile aleatoria esponenziale)
  - $f(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\theta)^2}{2\sigma^2}}; \theta = \mu$ (variabile aleatoria normale)
  - $f(x,\theta) = \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-\mu)^2}{2\theta}}; \theta = \sigma^2$ (variabile aleatoria normale)

Supponiamo di pensare che

$$\theta \in \Omega_0 \vee \theta \in \Omega_1$$
$$\text{dove } \Omega = \Omega_0 \cup \Omega_1 \text{ con } \Omega_0 \cap \Omega_1 = \emptyset$$

Etichettiamo queste ipotesi come:

$$H_0 : \theta \in \Omega_0 \underline{\text{ contro }} H_1 : \theta \in \Omega_1$$

L'obiettivo è decidere le ipotesi H_0 e H_1

$$\begin{cases} H_0 & \text{Si dice ipotesi nulla} \\ H_1 & \text{Si dice ipotesi alternativa} \end{cases}$$

> H_0 è vero $\Leftrightarrow \theta \in \Omega_0$

> H_1 è vero $\Leftrightarrow \theta \in \Omega_1$

L'**ipotesi nulla H_0** rappresenta **nessun cambiamento** o nessuna differenza rispetto al passato, mentre l'**ipotesi alternativa H_1** rappresenta il **cambiamento** o la differenza.

Errore di prima specie

Un errore di prima specie si verifica quando l'ipotesi H_0 viene rifiutata quando è vera. L'**errore di prima specie** è spesso considerato il **peggiore** dei due errori.

- Se accettiamo $H_1(\theta \in \Omega_1)$ quando H_0 è vera (rifiutiamo H_0 quando è vera ($\theta \in \Omega_0$)) → **Errore di prima specie.**

Errore di seconda specie

Un errore di seconda specie si verifica quando l'ipotesi H_1 viene rifiutata quando è vera (cioè l'ipotesi H_0 viene accettata quando è falsa).

- Se accettiamo H_0 quando H_1 è vera (rifiutiamo H_1 quando è vera) → **Errore di seconda specie.**

Decisione	H_0 è vera	H_1 è vera
Rifiuta H_0	Errore di prima specie	Decisione corretta
Accetta H_0	Decisione corretta	Errore di seconda specie

H_0 è vera → $\theta \in \Omega_0$

H_1 è vera → $\theta \in \Omega_1$

Regione critica del test

Sia $X_1, X_2, \ldots, X_n$ un **campione** di una popolazione che ha densità $f(x, \theta)$, $\theta \in \Omega = \Omega_0 \cup \Omega_1$.

Consideriamo l'ipotesi nulla H_0 contro l'ipotesi alternativa H_1

$$H_0 : \theta \in \Omega_0 \text{ contro } H_1 : \theta \in \Omega_1 \qquad \Omega = \Omega_0 \cup \Omega_1$$

Un **test** di H_0 contro H_1 si basa su una regione $C \subset \mathbb{R}^n$ detta **regione critica del test**. La sua regola decisionale corrispondente (test) è:

- **Rifiuta** H_0 (accetta H_1) se $(X_1, \ldots, X_n) \in C$
- **Accetta** H_0 (rifiuta H_1) se $(X_1, \ldots, X_n) \notin C$

Osservazione:

1. Siccome la densità congiunta di $X_1, \ldots, X_n$ dipende da θ, allora $P_\theta\Big((X_1, \ldots, X_n) \in C\Big)$ dipende da θ.
2. Se $\theta \in \Omega_0$ (H_0 è vera), in questo caso si ha:

$$P_\theta[(X_1, \ldots, X_n) \in C] \rightarrow P_\theta(\text{errore di prima specie})$$

   che indica la probabilità di rifiutare H_0, quando questa è vera.

   - Dove P_θ indica che $\theta \in \Omega_0(H_0 \text{ è vera})$.
   - $[(X_1, \ldots, X_n) \in C]$ → significa: "rifiuta H_0"

3. Se $\theta \in \Omega_1$ (H_1 è vera), in questo caso si ha:

$$P_\theta[(X_1, \dots, X_n) \notin C] \rightarrow P_\theta(\text{errore di seconda specie}) = 1 - P_\theta[(X_1, \dots, X_n) \in C]$$

che indica la probabilità di rifiutare H_1, quando questa è vera.

- Dove P_θ indica che $\theta \in \Omega_1 (H_1$ è vera).
- $[(X_1, \dots, X_n) \notin C]$ → significa: "rifiuta H_1"

4. La funzione

$$\gamma_g(\theta) = P_\theta[(X_1, \dots, X_n) \in C]; \quad \theta \in \Omega_1 \rightarrow H_1 \text{ è vera}$$

$[(X_1, \dots, X_n) \in C]$ → significa: "accetta H_1"

si dice **funzione potenza del test** e indica la probabilità di accettare H_1 quando questa è vera (decisione corretta).

Livello di significatività

L'obiettivo, ovviamente, è **selezionare una regione critica** tra tutte le possibili regioni critiche che *riduca al minimo le probabilità di questi errori*. In generale, *non è possibile minimizzare entrambi gli errori* contemporaneamente. Tuttavia, l'**errore di prima specie** è spesso considerato il **peggiore** dei due errori. Si procede quindi selezionando le regioni critiche che **delimitano la probabilità di errore** di **prima specie** e **tra queste regioni critiche** si cerca di selezionarne una che **minimizzi** la **probabilità di errore di seconda specie**.

Il livello di significatività α di un test con regione critica C, è definito da:

$$\alpha = \max_{\theta \in \Omega_0} P_\theta[(X_1, \dots, X_n) \in C] = \max_{\theta \in \Omega_0} P_\theta(\text{errore di prima specie})$$

Un test con livello di significatività α, deve avere una **probabilità di errore di prima specie minore o uguale ad α**. Il livello di significatività α del test viene normalmente fissato in anticipo, con valori tipici dell'ordine $0.1, 0.05 \dots$

Ora su regioni critiche con livello di significatività α, vogliamo considerare le regioni critiche che hanno minori probabilità di errore di seconda specie. Esaminiamo anche il complemento dell'errore di seconda specie, ovvero il rifiuto di H_0 quando H_1 è vera, che è una decisione corretta. Desideriamo massimizzare la probabilità di quest'ultima decisione. Cioè, per $\theta \in \Omega_1$ (H_1 è vera), vogliamo massimizzare

$$1 - P_\theta[\text{Errore di seconda specie}] = P_\theta[(X_1, \dots, X_n) \in C]$$

Quindi la funzione $\gamma_C(\theta) = P_\theta[(X_1, \dots, X_n) \in C], \theta \in \Omega_1$ si dice **funzione di potenza del test**. Per un valore di $\theta \in \Omega_1$ fissato, la potenza del test $\gamma_C(\theta)$ è la probabilità di rifiutare H_0 quando $\theta \in \Omega_1$ è il valore vero (questa decisione è corretta).

Confronto tra due regioni critiche

Date due regioni critiche C_1 e C_2 che hanno lo stesso livello di significatività α, C_1 è migliore di C_2 se

$$\gamma_{C_1}(\theta) \geq \gamma_{C_2}(\theta) \qquad \forall \theta \in \Omega_1$$

P-value o P-dei-dati

É la significatività massima α per cui accettiamo H_0

Il *p-value* indica quanto probabile è che una **statistica del test S_n**, sia uguale o "più estrema" di quella effettivamente osservata, che si denota con s_n.

Il *p-value* indica quanto è grande il mio errore e comunque accettare H_0. → Più è alto il *p-value* più ci concediamo valori di errore.

Per "più estremo" si intende (in base al tipo di test che si intende effettuare):

- *p-value* $:= P(S_n \geq s_n | H_0 \text{ è vera})$ per un test unilaterale destro;
- *p-value* $:= P(S_n \leq s_n | H_0 \text{ è vera})$ per un test unilaterale sinistro;
- *p-value* $:= P(|S_n| \geq |s_n|) | H_0 \text{ è vera})$ per un test bilaterale.

La statistica del test S_n è un valore che calcoliamo dai dati a nostra disposizione. Più il *p-value* ***è piccolo***, più è grande la significatività poiché il **risultato ci dice che l'ipotesi H_0 considerata non spiega adeguatamente i dati osservati, cioè è poco credibile**. Quindi un altro ragionamento che possiamo adottare per capire se accettare o rifiutare l'ipotesi nulla è quello di confrontare il *p-value* con il livello di significatività del test α **ipotizzato inizialmente**. Infatti si ha che:

1. Se il *p-value* $> \alpha$ l'evidenza empirica non è sufficientemente contraria all'ipotesi nulla che quindi **non può essere rifiutata** → Non rifiutiamo H_0
2. Se il *p-value* $\leq \alpha$ l'evidenza empirica è fortemente contraria all'ipotesi nulla che quindi **va rifiutata** → rifiutiamo H_0

P-value nel caso di una variabile normale standard

- Se $H_1 : \mu > \mu_0$ *p-value*$= P(Z \geq z)$ → **test unilaterale destro.**
- Se $H_1 : \mu < \mu_0$ *p-value*$= P(Z \leq z)$ → **test unilaterale sinistro.**
- Se $H_1 : \mu \neq \mu_0$ *p-value*$= P(|Z| \geq |z|)$ → **test bilaterale**.

dove $Z = \frac{\vec{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$ e $z = \frac{\vec{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ è il valore assunto da Z

In pratica spesso non si fissa in anticipo il livello di significatività, ma si osservano i dati e si ricava il ***p-value*** corrispondente. Se il ***p-value*** **risulta molto maggiore** di quanto siamo disposti ad accettare come **probabilità di un errore di prima specie**, conviene **accettare l'ipotesi nulla**; se invece è **molto piccolo**, possiamo **rifiutarla**.

▼ Esempio *p-value*

Esempio. Si vuole verificare l'ipotesi nulla $H_0 : \mu = 175$ contro l'ipotesi alternativa $H_1 : \mu > 175$. Si estrae un campione di 10 elementi la cui media campionaria è $\bar{x} = 181,5$ e la cui varianza è $\sigma^2 = 100$. Il test è condotto ad un livello di significatività $\alpha = 0,05$.
Costruiamo la statistica del test a partire dello stimatore naturale del valore atteso, ovvero la media campionaria $\overline{X}$. La statistica del test (tenendo conto che H_0 è vera, cioè $\mu = 175$ e $E(\overline{X}) = 175$) è data da

$$S_{10} = Z = \frac{\overline{X} - 175}{\sigma/\sqrt{10}} = \frac{\overline{X} - 175}{\sqrt{10}} \sim \mathcal{N}(0,1)$$

Il valore osservato della statistica del test è

$$s_n = z = \frac{181,5 - 175}{\sqrt{10}} = \frac{6,5}{\sqrt{10}} = 2,055$$

Si ottiene quindi

$$p - \text{value} = P(Z \geq 2,055) = 1 - \Phi(2.055) = 1 - 0.98 = 0.02 < \alpha = 0.05$$

Allora l'ipotesi nulla H_0 può essere rifiutata.

Siccome $H_1 : \mu > 175 \qquad \mu_0 = 175$, allora di effettua un test **unilaterale destro**.

P-value compreso tra quali valori

> Il *p-value* è compreso tra i valori di α. Se si chiede di effettuare un test al 5% e all' 1%, quindi $\alpha = 0.05$ e $\alpha = 0,01$ → il *p-value* del test sarà **compreso** tra l'1 e il 5%.

Verifica delle ipotesi sulla media di una popolazione normale con varianza nota.

a) Test bilaterali

Supponiamo che $X_1, \ldots, X_n$ sia un campione aleatorio proveniente da una popolazione normale di media μ **incognita** e varianza σ^2 **nota**. Fissata una costante μ_0, vogliamo verificare l'ipotesi nulla contro l'ipotesi alternativa:

$$H_0 : \mu = \mu_0 \ \underline{\text{contro}} \ H_1 : \mu \neq \mu_0$$

Siccome $\vec{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$ è lo stimatore puntuale naturale per μ. La regione critica ragionevole del test sarà del tipo:

$$C := \{(X_1, X_2, \ldots, X_n) : |\vec{X} - \mu_0| > c\}$$

Per un'opportuna scelta della costante c. Ricordiamo che l'ipotesi H_0 può essere rifiutata quando $\vec{x}$ non è vicina a μ → $\vec{x} - \mu_0 \neq 0 \to |\vec{x}| - \mu_0 > 0 \to |\vec{x} - \mu_0| > c$ per alcuni $c > 0$.

- H_0 è rifiutata se $(X_1, \ldots, X_n) \in C \Leftrightarrow |\vec{X} - \mu_0| > c$
- H_0 è accettata se $(X_1, \ldots, X_n) \notin C \Leftrightarrow |\vec{X} - \mu_0| \leq c$

$$z_{\frac{\alpha}{2}} = \frac{c\sqrt{n}}{\sigma}$$
$$\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

$$c = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$P\left(\left| \frac{\vec{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > z_{\frac{\alpha}{2}} \right) = \alpha$$

In conclusione:

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{|\vec{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} > z_{\frac{\alpha}{2}} \\ . \\ \text{Accettare } H_0 \text{ se } \frac{|\vec{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}} \end{cases}$$

Dove $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$

Essendo $z = \frac{\vec{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ il valore assunto da Z. allora

$$p - value \ = P(|Z| \geq z) = \ 2P(Z \geq |z|) = 2(1 - \phi(|z|))$$

▼ Esempio

Esempio. Un manager di una compagnia telefonica ritiene che la bolletta mensile per il cellulare dei loro clienti sia cambiata, e che in media non sia più pari a 40 euro al mese. Supponiamo di sapere che $\sigma = 10$ euro e che la media campionaria di un campione di 64 clienti è $\bar{x} = 41.54$ euro. La compagnia desidera verificare

$$H_0 : \mu = 40 \ \text{ contro } \ H_1 : \mu \neq 40$$

con livello di significatività $\alpha = 0.05$.
Abbiamo $\Phi(z_{0.025}) = 1 - 0.025 = 0.975$. Si ottiene quindi $z_{0.025} = 1.96$.
Perciò il test ad un livello di significatività $\alpha = 0.05$ dovrà rifiutare H_0 se $\frac{|\overline{X}-40|}{\sigma/\sqrt{n}} > 1.96$ e accettare H_0 se $\frac{|\overline{X}-40|}{\sigma/\sqrt{n}} \leq 1.96$. Siccome

$$z = \frac{|\bar{x} - 40|}{\sigma/\sqrt{n}} = \frac{41.54 - 40}{10/\sqrt{64}} = 1.24 < 1.96,$$

allora l'ipotesi H_0 va accettata. Inoltre si ha
p-value=2(1-$\Phi(z)) = 2(1 - \Phi(1.24)) = 0.21 > \alpha = 0.05$

---

a.1) Test bilaterali → funzione potenza (decisione corretta)

Ora fissiamo il livello di significatività del test α e $\mu \neq \mu_0$ (cioè H_1 è vera e $E(\vec{X}) = \mu$), la probabilità di rifiutare H_0 quando μ è il valore vero (**decisione corretta** → $\theta \in \Omega_1$), data da

$$\begin{aligned} \gamma_c(\mu) &= P\Big((X_1, \ldots, X_n) \in C\Big) \\ &= 1 - \left[\Phi\left(\frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} + z_{\frac{\alpha}{2}}\right) - \Phi\left(\frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} - z_{\frac{\alpha}{2}}\right)\right] \end{aligned}$$

Dove $\frac{\mu_0 - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$

$\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \longrightarrow$ dalla tabella troviamo il valore di $z_{\frac{\alpha}{2}}$

▼ Dimostrazione

$$
\begin{aligned}
\gamma_C(\mu) &= P_\mu\Big((X_1, X_2, \ldots, X_n) \in C)\Big) \\
&= P_\mu(|\overline{X} - \mu_0| > c) \\
&= 1 - P_\mu(|\overline{X} - \mu_0| \leq c) \\
&= 1 - P_\mu\Big(-c \leq \overline{X} - \mu + (\mu - \mu_0) \leq c\Big) \\
&= 1 - P_\mu\Big((\mu_0 - \mu) - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq (\mu_0 - \mu) + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\Big) \\
&= 1 - P_\mu\Big(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\frac{\alpha}{2}}\Big) \\
&= 1 - \Big[\Phi\Big(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\frac{\alpha}{2}}\Big) - \Phi\Big(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\frac{\alpha}{2}}\Big)\Big]
\end{aligned}
$$

---

b) Test unilaterale destro

Supponiamo che $X_1, \ldots, X_n$ sia un campione aleatorio proveniente da una popolazione normale di media μ **incognita** e varianza σ^2 **nota**.

Fissata una costante μ_0, vogliamo verificare l'ipotesi nulla contro l'ipotesi alternativa:

$$H_0 : \mu = \mu_0 \ \underline{\text{contro}} \ H_1 : \mu > \mu_0$$

In questo caso:

$$C := \{(X_1, \ldots, X_n) : \vec{X} - \mu_0 > c\}$$

Fissiamo un livello di significatività α. Perciò

$$\alpha = P_{\mu_0}[(X_1, \ldots, X_n) \in C] = P_{\mu_0}(\vec{X} - \mu_0 > c) \quad E(\vec{X}) = \mu_0$$

- La **statistica** Z, sotto l'ipotesi $H_0 : (E(\vec{X}) = \mu_0)$ → H_0 è vera, è:

$$Z = \frac{\vec{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

- La **statistica** Z, sotto l'ipotesi $H_1 : (E(\vec{X}) = \mu)$ → H_1 è vera, è:

$$Z = \frac{\vec{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

Sia z_α tale che $P(Z > z_\alpha) = \alpha$, cioè $\Phi(z_\alpha) = 1 - \alpha$. Allora

$$c = z_\alpha \frac{\sigma}{\sqrt{n}}$$

In conclusione, ad un livello di **significatività α:**

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{\vec{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha \\ . \\ \text{Accettare } H_0 \text{ se } \frac{\vec{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \le z_\alpha \end{cases}$$

Dove il valore di $z_\alpha \to \Phi(z_\alpha) = 1 - \alpha$ è dato dalla tabella.

$$p - value = P(Z \ge z) = 1 - \Phi(z)$$

dove $z = \frac{\vec{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ è il valore assunto di Z.

La regola che abbiamo utilizzato per verificare l'ipotesi $H_0 : \mu = \mu_0$ contro l'ipotesi $H_1 : \mu > \mu_0$ vale anche per verificare ad un livello di significatività α, l'ipotesi unilaterale

$$H_0 : \mu \le \mu_0 \ \underline{\text{contro}} \ H_1 : \mu > \mu_0$$

---

c) test unilaterale sinistro

Vogliamo verificare l'ipotesi nulla H_0 contro l'ipotesi alternativa H_1:

$$H_0 : \mu = \mu_0 \ \underline{\text{contro}} \ H_1 : \mu < \mu_0$$

otteniamo ad un livello di significatività α:

$$c = -z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{\vec{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_\alpha \\ . \\ \text{Accettare } H_0 \text{ se } \frac{\vec{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \ge -z_\alpha \end{cases}$$

Dove il valore di $z_\alpha \to \Phi(z_\alpha) = 1 - \alpha$ è dato dalla tabella.

$$p - value = P(Z \le z) = \Phi(z)$$

dove $z = \frac{\vec{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ è il valore assunto di Z.

Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza nota.

Se $X_1, \dots, X_n$ è un campione estratto da una popolazione $\mathcal{N}(\mu, \sigma^2)$ con σ^2 **nota** e $\vec{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, allora otteniamo questa tabella:

H_0	H_1	Statistica del test, Z	Si rifiuta H_0 con livello di signficativà α se :	p-value (o p-dei-dati) se $Z = z$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\left\lvert\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right\rvert > z_{\frac{\alpha}{2}}$	$2P(Z \geq \lvert z\rvert)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$	$P(Z \geq z)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$	$P(Z \leq z)$

dove

$$Z \sim \mathcal{N}(0,1)$$

Verifica delle ipotesi sulla media di una popolazione normale con varianza NON nota: il t-test

Supponiamo che $X_1, \ldots, X_n$ sia un campione aleatorio proveniente da una popolazione normale di media μ **incognita** e varianza σ^2 **non nota**. Fissata una costante μ_0, vogliamo verificare l'ipotesi nulla $H_0 : \mu = \mu_0$ contro l'ipotesi $H_1 : \mu \neq \mu_0$. Siccome la deviazione standard campionaria

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

approssima bene la deviazione standard σ della popolazione, allora la **statistica del test sotto l'ipotesi H_0 è**:

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

dove t_{n-1} è una variabile aleatoria t di Student con $n-1$ gradi di libertà.

a) t - Test bilaterale

Siccome σ non è più conosciuta, la **sostituiamo** con il suo stimatore, la **deviazione standard campionaria** S

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{\vec{X}-\mu_0}{\frac{S}{\sqrt{n}}} > t_{\frac{\alpha}{2},n-1} \\ . \\ \text{Accettare } H_0 \text{ se } \frac{\vec{X}-\mu_0}{\frac{S}{\sqrt{n}}} \leq t_{\frac{\alpha}{2},n-1} \end{cases}$$

dove

$$P\left(\left|\frac{\vec{X} - \mu_0}{\frac{S}{\sqrt{n}}}\right| > t_{\frac{\alpha}{2},n-1}\right) = \alpha$$

Cioè $\phi_{t_{n-1}}(t_{\frac{\alpha}{2},n-1}) = 1 - \frac{\alpha}{2}$, dove $\phi_{t_{n-1}}$ è la funzione di ripartizione della distribuzione t di Student con $n-1$ gradi di libertà.

▼ Esempio

Esempio. Si vuole verificare se il costo medio di una camera di hotel per 2 nel parco di Yellowstone sia pari a 168\$. Per un campione di 25 hotel è stato determinato che $\bar{x} = 172.50\$$ e $s = 15.40\$$. Verificare l'ipotesi ad un livello = 0.05. Assumiamo che la popolazione abbia distribuzione normale. Le ipotesi del test sono

$$H_0 : \mu = 168 \quad \text{contro} \;\; H_1 : \mu \neq 168$$

Abbiamo

$$t_{0.025,24} = 2.064, \; \frac{\bar{x} - 168}{s/\sqrt{n}} = \frac{172.5 - 168}{15.40/\sqrt{25}} = 1.46 < t_{0.025,24}$$

Allora concludiamo che non possiamo rifiutare H_0.

$$\phi_{t_{25-1}}(t_{\frac{0.05}{2},25-1}) = 1 - \frac{0.05}{2} \rightarrow \phi_{t_{24}}(t_{0.025,24}) = 0.975 \rightarrow \text{Dalla tabella t di Student} \rightarrow t_{0.025,24} = 2.064$$

---

b) t - test unilaterale destro

$$H_0 : \mu = \mu_0 (o\; H_0 : \mu \leq \mu_0), \; \underline{\text{contro}} \; H_1 : \mu > \mu_0$$

ad un livello di significatività α:

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{\vec{X} - \mu_0}{\frac{S}{\sqrt{n}}} > t_{\alpha,n-1} \\ . \\ \text{Accettare } H_0 \text{ se } \frac{\vec{X} - \mu_0}{\frac{S}{\sqrt{n}}} \leq t_{\alpha,n-1} \end{cases}$$

dove $\phi_{t_{n-1}}(t_{\alpha,n-1}) = 1 - \alpha$

---

b) t - test unilaterale sinistro

$$H_0 : \mu = \mu_0 (o\; H_0 : \mu \geq \mu_0), \; \underline{\text{contro}} \; H_1 : \mu < \mu_0$$

ad un livello di significatività α:

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{\vec{X} - \mu_0}{\frac{S}{\sqrt{n}}} < -t_{\alpha,n-1} \\ . \\ \text{Accettare } H_0 \text{ se } \frac{\vec{X} - \mu_0}{\frac{S}{\sqrt{n}}} \geq -t_{\alpha,n-1} \end{cases}$$

dove $\phi_{t_{n-1}}(t_{\alpha,n-1}) = 1 - \alpha$

Tabella riassuntiva verifica delle ipotesi sulla media di una popolazione normale con varianza NON nota.

Se $X_1, \dots, X_n$ è un campione con distribuzione $\mathcal{N}(\mu, \sigma^2)$ con σ^2 **NON nota** e $\vec{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, allora otteniamo questa tabella:

H_0	H_1	Statistica del test, T_{n-1}	Si rifiuta H_0 con livello di signficatività α se :	p-value (o p-dei-dati) se $T_{n-1} = t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\left\lvert\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right\rvert > t_{\frac{\alpha}{2},n-1}$	$2P(T_{n-1} \geq \lvert t\rvert)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > t_{\alpha,n-1}$	$P(T_{n-1} \geq t)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} < -t_{\alpha,n-1}$	$P(T_{n-1} \leq t)$

dove

$$T_{n-1} \sim t_{n-1}, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Verifica delle ipotesi sulla differenza delle medie di due campioni normali (numerosità molto grandi con varianza nota e con varianza NON nota)

In molte applicazioni è utile **confrontare se due popolazioni**, da cui sono estratti due campioni distinti, possono essere ritenute uguali oppure se tra esse si può riscontrare una differenza significativa. Questi problemi possono essere risolti effettuando un **test** sulla **differenza dei valori medi**.

Consideriamo $X_1, \dots, X_n$ → campione normale con media μ_x e varianza σ_x^2.

Consideriamo $Y_1, \dots, Y_m$ → campione normale con media μ_y e varianza σ_y^2.

$$\vec{X} \sim \mathcal{N}\Big(\mu_x, \frac{\sigma_x^2}{n}\Big)$$

$$\vec{Y} \sim \mathcal{N}\Big(\mu_y, \frac{\sigma_y^2}{m}\Big) \rightarrow -\vec{Y} \sim \mathcal{N}\Big(-\mu_y, \frac{\sigma_y^2}{m}\Big)$$

perché $Y = \alpha x + \beta \sim \mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$.

$\vec{X}$ e $-\vec{Y}$ sono **indipendenti**, quindi $\vec{X} - \vec{Y}$ è una variabile aleatoria normale:

$$E(\vec{X} - \vec{Y}) = \mu_x - \mu_y$$
$$Var(\vec{X} - \vec{Y}) = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$$
$$\vec{X} - \vec{Y} \sim \mathcal{N}(\mu_x - \mu_y, \tfrac{\sigma_x^2}{n} + \tfrac{\sigma_y^2}{m})$$

Allora

$$\frac{\vec{X} - \vec{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \mathcal{N}(0,1)$$

Se $\mu_x = \mu_y$, allora

$$\frac{\vec{X} - \vec{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \mathcal{N}(0,1)$$

Vogliamo verificare:

$$H_0 : \mu_x = \mu_y \ \underline{\text{contro}} \ H_1 : \mu_x \neq \mu_y$$

Siccome $\vec{X}$ è uno stimatore di μ_x e $\vec{Y}$ è uno stimatore di μ_y, segue che $\vec{X} - \vec{Y}$ può essere usato per stimare $\mu_x - \mu_y$. Poiché $H_0 : \mu_x = \mu_y$, la forma del test è:

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } |\vec{X} - \vec{Y}| > c \\ . \\ \text{Accettare } H_0 \text{ se } |\vec{X} - \vec{Y}| \leq c \end{cases}$$

per un opportuno valore di c.

a) se le varianza σ_x^2 e σ_y^2 sono note:

$$\begin{aligned} \alpha &= P_{H_0}(|\overline{X} - \overline{Y}| > c) \\ &= P_{H_0}\left(\left|\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right| > z_{\frac{\alpha}{2}}\right), \quad \text{sotto } H_0 : \mu_x = \mu_y \\ &= P\left(\left|\frac{\overline{X} - \overline{Y}}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}}\right| > z_{\frac{\alpha}{2}}\right), \quad \text{con } z_{\frac{\alpha}{2}} = \frac{c}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \end{aligned}$$

Si deduce che un test ad un livello di significatività α

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{|\vec{X} - \vec{Y}|}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} > z_{\frac{\alpha}{2}} \\ . \\ \text{Accettare } H_0 \text{ se } \frac{|\vec{X} - \vec{Y}|}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \leq z_{\frac{\alpha}{2}} \end{cases}$$

Dove $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ → dalla tabella di **distribuzione normale**.

$$c = z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

---

a) se le varianza σ_x^2 e σ_y^2 NON sono note:

La statistica del test è:

$$Z = \frac{\vec{X} - \vec{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

Dove

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \vec{X})^2$$
$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{m}(Y_i - \vec{Y})^2$$

Questa statistica ha una distribuzione complicata. Tuttavia nel caso in cui n e m sono entrambi **dei numeri molto grandi**, si può dimostrare che la distribuzione di Z e approssimativamente $\mathcal{N}(0,1)$. In questo caso per verificare approssimativamente ad un livello di significatività α

$$H_0 : \mu_x = \mu_y \ \underline{\text{contro}} \ H_1 : \mu_x \neq \mu_y$$

$$\begin{cases} \text{Rifiutare } H_0 \text{ se } \frac{|\vec{X}-\vec{Y}|}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} > z_{\frac{\alpha}{2}} \\ . \\ \text{Accettare } H_0 \text{ se } \frac{|\vec{X}-\vec{Y}|}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} \leq z_{\frac{\alpha}{2}} \end{cases}$$

Dove $\Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ → dalla tabella di **distribuzione normale**.
