# Verifica delle Ipotesi

Stefania Bartoletti

3 Giugno 2020

# Indice

# Verifica delle ipotesi

- Disponiamo di un campione di dati estratti da una distribuzione nota, a meno di uno o più parametri

# Verifica delle ipotesi

- Disponiamo di un campione di dati estratti da una distribuzione nota, a meno di uno o più parametri
- Non siamo interessati a stimare direttamente tali parametri, ma verificare un'ipotesi che dipende da tali parametri

# Verifica delle ipotesi

- Disponiamo di un campione di dati estratti da una distribuzione nota, a meno di uno o più parametri
- Non siamo interessati a stimare direttamente tali parametri, ma verificare un'ipotesi che dipende da tali parametri
- Esempio: Vogliamo confrontare un nuovo farmaco con un placebo e verificare che il nuovo farmaco è effettivamente migliore del placebo dopo averlo somministrato ad un campione di pazienti.

# Definizioni

- $H_0$: l'ipotesi nulla è l'ipotesi di default per il modello che genera i dati

# Definizioni

- $H_0$: l'ipotesi nulla è l'ipotesi di default per il modello che genera i dati
- $H_1$: l'ipotesti alternativa (indicata anche con $H_\mathsf{a}$). Se rifiutiamo l'ipotesi nulla, stiamo di fatto decidendo che l'ipotesi alternativa spiega meglio il campione dei dati.

# Definizioni

- $H_0$: l'ipotesi nulla è l'ipotesi di default per il modello che genera i dati
- $H_1$: l'ipotesti alternativa (indicata anche con $H_\mathrm{a}$). Se rifiutiamo l'ipotesi nulla, stiamo di fatto decidendo che l'ipotesi alternativa spiega meglio il campione dei dati.
- x: la statistica test, che viene calcolata a partire dai dati

# Definizioni

- $H_0$: l'ipotesi nulla è l'ipotesi di default per il modello che genera i dati
- $H_1$: l'ipotesti alternativa (indicata anche con $H_\mathsf{a}$). Se rifiutiamo l'ipotesi nulla, stiamo di fatto decidendo che l'ipotesi alternativa spiega meglio il campione dei dati.
- x: la statistica test, che viene calcolata a partire dai dati
- $f(\mathsf{x}|H_0)$ la distribuzione di probabilità di x assumendo $H_0$

# Definizioni

- $H_0$: l'ipotesi nulla è l'ipotesi di default per il modello che genera i dati
- $H_1$: l'ipotesti alternativa (indicata anche con $H_a$). Se rifiutiamo l'ipotesi nulla, stiamo di fatto decidendo che l'ipotesi alternativa spiega meglio il campione dei dati.
- x: la statistica test, che viene calcolata a partire dai dati
- $f(x|H_0)$ la distribuzione di probabilità di x assumendo $H_0$
- $\mathcal{C}$: la regione critica (rejection region). Se $x \in \mathcal{C}$, rifiutiamo $H_0$

# Nota

- L'ipotesi nulla è solitamente una ipotesi *cauta*. Per essere smentita, ha bisogno di una forte evidenza.

# Nota

- L'ipotesi nulla è solitamente una ipotesi *cauta*. Per essere smentita, ha bisogno di una forte evidenza.
- Pertanto, la regione critica contiene i valori dei dati che sono estremamente incompatibili con $H_0$, ovvero sono sulle code della distribuzione rispetto ai valori più probabili.

# Nota

- L'ipotesi nulla è solitamente una ipotesi *cauta*. Per essere smentita, ha bisogno di una forte evidenza.
- Pertanto, la regione critica contiene i valori dei dati che sono estremamente incompatibili con $H_0$, ovvero sono sulle code della distribuzione rispetto ai valori più probabili.
- Se $x \in \mathcal{C}$, diciamo che i dati non sono compatibili con l'ipotesi nulla.

# Nota

- L'ipotesi nulla è solitamente una ipotesi *cauta*. Per essere smentita, ha bisogno di una forte evidenza.
- Pertanto, la regione critica contiene i valori dei dati che sono estremamente incompatibili con $H_0$, ovvero sono sulle code della distribuzione rispetto ai valori più probabili.
- Se $x \in \mathcal{C}$, diciamo che i dati non sono compatibili con l'ipotesi nulla.
- In ogni caso non diciamo che un'ipotesi è vera, ma solo se i dati sono compatibili con essa.

# Ipotesi semplici e composte

- Una ipotesi si dice semplice se possiamo specificare la corrispondente distribuzione. Per esempio, un parametro assume uno specifico valore.
- Una ipotesi si dice composta se non si può specificare pienamente la sua distribuzione. Ad esempio, il parametro appartiene ad un range di valori.

- La qualità di un test di significatività dipende dal livello di significatività e dalla potenza del test. Idealmente, non vorremmo compiere errori, che sono di due tipi:

- La qualità di un test di significatività dipende dal livello di significatività e dalla potenza del test. Idealmente, non vorremmo compiere errori, che sono di due tipi:
  - Errore di tipo I (prima specie) rifiuto $H_0$ quando $H_0$ è vera

- La qualità di un test di significatività dipende dal livello di significatività e dalla potenza del test. Idealmente, non vorremmo compiere errori, che sono di due tipi:
  - Errore di tipo I (prima specie) rifiuto $H_0$ quando $H_0$ è vera
  - Errore di tipo II (seconda specie) non rifiuto $H_0$ quando $H_1$ è vera

- La qualità di un test di significatività dipende dal livello di significatività e dalla potenza del test. Idealmente, non vorremmo compiere errori, che sono di due tipi:
  - Errore di tipo I (prima specie) rifiuto $H_0$ quando $H_0$ è vera
  - Errore di tipo II (seconda specie) non rifiuto $H_0$ quando $H_1$ è vera
- Definiamo il livello di significatività del test come $\mathbb{P}\{\text{rifiuto } H_0|H_0\} = \mathbb{P}\{\text{Errore di tipo I}\} = \alpha$

- La qualità di un test di significatività dipende dal livello di significatività e dalla potenza del test. Idealmente, non vorremmo compiere errori, che sono di due tipi:
    - Errore di tipo I (prima specie) rifiuto $H_0$ quando $H_0$ è vera
    - Errore di tipo II (seconda specie) non rifiuto $H_0$ quando $H_1$ è vera
- Definiamo il livello di significatività del test come $\mathbb{P}\left\{\text{rifiuto } H_0 | H_0\right\} = \mathbb{P}\left\{\text{Errore di tipo I}\right\} = \alpha$
- Definiamo la di potenza del test come $\mathbb{P}\left\{\text{rifiuto } H_0 | H_1\right\} = 1 - \mathbb{P}\left\{\text{Errore di tipo II}\right\} = 1 - \beta$

# Indice

# Verifica delle ipotesi sulla media di una popolazione normale

Supponiamo che $x_1, x_2, \ldots, x_n$ sia un campione aleatorio proveniente da una popolazione normale di parametri $\mu$ e $\sigma^2$, con varianza nota .

1. Definiamo l'ipotesi nulla $H_0 : \mu = \mu_0$
2. Definiamo l'ipotesi alternativa $H_1 : \mu \neq \mu_0$
3. Scegliamo una statistica test Siccome $\overline{\mathrm{x}}$ è lo stimatore puntuale naturale per $\mu$ sembra ragionevole accettare $H_0$ quando $\overline{x}$ non è troppo lontano da $\mu_0$, guardando quindi a $|\overline{\mathrm{x}} - \mu_0|$

# Verifica delle ipotesi sulla media di una popolazione normale

4. Definiamo un livello di significatività e quindi la regione critica
   Scegliamo $\alpha$ e quindi definiamo la regione critica come

$$\mathcal{C} = \{x_1, x_2, \ldots, x_n : |\overline{x} - \mu_0| > c\}$$
$$\mathbb{P}\left\{|\overline{\mathsf{x}} - \mu_0| > c\,; \mu = \mu_0\right\} = \alpha$$

Chi è $c$?

# Verifica delle ipotesi sulla media di una popolazione normale

4. Definiamo un livello di significatività e quindi la regione critica
   Scegliamo $\alpha$ e quindi definiamo la regione critica come

$$\mathcal{C} = \{x_1, x_2, \ldots, x_n : |\overline{x} - \mu_0| > c\}$$

$$\mathbb{P}\{|\overline{\mathsf{x}} - \mu_0| > c\,; \mu = \mu_0\} = \alpha$$

   Chi è $c$?
   Considerando che $\frac{\overline{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ assumendo $H_0$, allora

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

# Verifica delle ipotesi sulla media di una popolazione normale

4. Definiamo un livello di significatività e quindi la regione critica

   Scegliamo $\alpha$ e quindi definiamo la regione critica come

$$\mathcal{C} = \{x_1, x_2, \ldots, x_n : |\overline{x} - \mu_0| > c\}$$
$$\mathbb{P}\{|\overline{\mathsf{x}} - \mu_0| > c\,; \mu = \mu_0\} = \alpha$$

Chi è $c$?

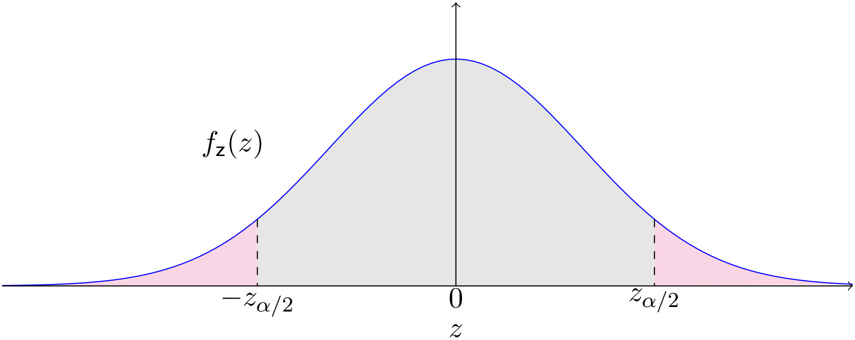
Considerando che $\frac{\overline{\mathsf{x}}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ assumendo $H_0$, allora

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

Quindi, di fatto,

- Rifiuto $H_0$ se $\left|\frac{\overline{\mathsf{x}}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$
- Non rifiuto $H_0$ (accettazione) se $\left|\frac{\overline{\mathsf{x}}-\mu_0}{\sigma/\sqrt{n}}\right| \leqslant z_{\alpha/2}$

# Regione critica




Accetto $H_0$ se

$$\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right| \leqslant z_{\alpha/2}$$

# Verifica delle ipotesi sulla media di una popolazione normale

5. Calcolo la potenza del test
   - Partiamo dalla probabilità di errore di seconda specie, ovvero la probabilità di accettare $H_0$ quando questa non è vera.

# Verifica delle ipotesi sulla media di una popolazione normale

5. Calcolo la potenza del test
    - Partiamo dalla probabilità di errore di seconda specie, ovvero la probabilità di accettare $H_0$ quando questa non è vera.
    - Tale probabilità dipende dal valore di $\mu$ e quindi definiamo

$$\beta(\mu) = \mathbb{P}\left\{\left|\frac{\overline{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}}\right| \leqslant z_{\alpha/2}; \mu\right\}$$

quando $\frac{\mathsf{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

# Verifica delle ipotesi sulla media di una popolazione normale

5. Calcolo la potenza del test
   - Partiamo dalla probabilità di errore di seconda specie, ovvero la probabilità di accettare $H_0$ quando questa non è vera.
   - Tale probabilità dipende dal valore di $\mu$ e quindi definiamo
   $$\beta(\mu) = \mathbb{P}\left\{\left|\frac{\bar{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}}\right| \leqslant z_{\alpha/2}; \mu\right\}$$
   quando $\frac{\mathsf{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
   - La funzione $\beta(\mu)$ è detta curva operativa caratteristica ;

# Verifica delle ipotesi sulla media di una popolazione normale

5. Calcolo la potenza del test
    - Partiamo dalla probabilità di errore di seconda specie, ovvero la probabilità di accettare $H_0$ quando questa non è vera.
    - Tale probabilità dipende dal valore di $\mu$ e quindi definiamo

    $$\beta(\mu) = \mathbb{P}\left\{ \left| \frac{\overline{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \right| \leqslant z_{\alpha/2}; \mu \right\}$$

    quando $\frac{\mathsf{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
    - La funzione $\beta(\mu)$ è detta curva operativa caratteristica ;
    - La funzione $1-\beta(\mu)$ viene detta funzione di potenza del test e rappresenta la probabilità di rifiutare (correttamente) $H_0$ quando $\mu$ è il valore vero.

# Verifica delle ipotesi sulla media di una popolazione normale

5. Calcolo la potenza del test
   - Partiamo dalla probabilità di errore di seconda specie, ovvero la probabilità di accettare $H_0$ quando questa non è vera.
   - Tale probabilità dipende dal valore di $\mu$ e quindi definiamo
   $$\beta(\mu) = \mathbb{P}\left\{ \left| \frac{\bar{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \right| \leqslant z_{\alpha/2}; \mu \right\}$$
   quando $\frac{\mathsf{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
   - La funzione $\beta(\mu)$ è detta curva operativa caratteristica ;
   - La funzione $1-\beta(\mu)$ viene detta funzione di potenza del test e rappresenta la probabilità di rifiutare (correttamente) $H_0$ quando $\mu$ è il valore vero.
   - Si dimostra che
   $$\begin{aligned}\beta(\mu) &= \mathbb{P}\left\{ -z_{\alpha/2} \leqslant \frac{\bar{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \leqslant z_{\alpha/2}; \mu \right\} \\ &= \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) - \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2}\right)\end{aligned}$$

# p-value

- La scelta del livello di significatività varia molto a seconda del problema: sto guardando alla quantità di cacao in una barretta o decidendo sul DNA di un omicidio?)

# p-value

- La scelta del livello di significatività varia molto a seconda del problema: sto guardando alla quantità di cacao in una barretta o decidendo sul DNA di un omicidio?)
- Nella pratica, spesso si specifica il livello di significatività $\alpha$ e il test si basa sul *p-value*.

# p-value

- La scelta del livello di significatività varia molto a seconda del problema: sto guardando alla quantità di cacao in una barretta o decidendo sul DNA di un omicidio?)
- Nella pratica, spesso si specifica il livello di significatività $\alpha$ e il test si basa sul *p-value*.
- Definizione: Il *p-value* è la probabilità, assumendo l'ipotesi nulla, di ottenere un campione dei dati estremo almeno quanto quello osservato.

# p-value

- La scelta del livello di significatività varia molto a seconda del problema: sto guardando alla quantità di cacao in una barretta o decidendo sul DNA di un omicidio?)
- Nella pratica, spesso si specifica il livello di significatività $\alpha$ e il test si basa sul *p-value*.
- Definizione: Il *p-value* è la probabilità, assumendo l'ipotesi nulla, di ottenere un campione dei dati estremo almeno quanto quello osservato.
- Nelle operazioni che si sono susseguite prima, abbiamo prima calcolato la statistica test e poi confrontato tale statistica con un valore che ci siamo ricavati a partire da $\alpha$.

# p-value

- La scelta del livello di significatività varia molto a seconda del problema: sto guardando alla quantità di cacao in una barretta o decidendo sul DNA di un omicidio?)
- Nella pratica, spesso si specifica il livello di significatività $\alpha$ e il test si basa sul *p-value*.
- Definizione: Il *p-value* è la probabilità, assumendo l'ipotesi nulla, di ottenere un campione dei dati estremo almeno quanto quello osservato.
- Nelle operazioni che si sono susseguite prima, abbiamo prima calcolato la statistica test e poi confrontato tale statistica con un valore che ci siamo ricavati a partire da $\alpha$.
- Di fatto potremmo calcolare la statistica test e la probabilità che una normale standard superi il valore ottenuto con il campione osservato. Tale probabilità è il il p-value del test, si confronta con $\alpha$ per decidere se accettare o meno l'ipotesi.

# Esempio

Un segnale di valore $\mu$ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore normale di media nulla e varianza $4$. Il segnale ricevuto da B ha quindi distribuzione $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato per $5$ volte lo stesso segnale: la media campionaria dei segnali ricevuti $\overline{x} = 9.5$. B aveva motivo di supporre che il valore inviato dovesse essere $8$. Si verifichi questa ipotesi.

1. Definiamo l'ipotesi nulla $H_0 : \mu = 8$

# Esempio

Un segnale di valore $\mu$ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore normale di media nulla e varianza $4$. Il segnale ricevuto da B ha quindi distribuzione $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato per $5$ volte lo stesso segnale: la media campionaria dei segnali ricevuti $\overline{x} = 9.5$. B aveva motivo di supporre che il valore inviato dovesse essere $8$. Si verifichi questa ipotesi.

1. Definiamo l'ipotesi nulla $H_0 : \mu = 8$
2. Definiamo l'ipotesi alternativa $H_1 : \mu \neq 8$

# Esempio

Un segnale di valore $\mu$ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore normale di media nulla e varianza $4$. Il segnale ricevuto da B ha quindi distribuzione $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato per $5$ volte lo stesso segnale: la media campionaria dei segnali ricevuti $\overline{x} = 9.5$. B aveva motivo di supporre che il valore inviato dovesse essere $8$. Si verifichi questa ipotesi.

1. Definiamo l'ipotesi nulla $H_0 : \mu = 8$
2. Definiamo l'ipotesi alternativa $H_1 : \mu \neq 8$
3. Scegliamo una statistica test $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{0.5}{2/\sqrt{5}} \simeq 0.559$

# Esempio

Un segnale di valore $\mu$ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore normale di media nulla e varianza $4$. Il segnale ricevuto da B ha quindi distribuzione $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato per $5$ volte lo stesso segnale: la media campionaria dei segnali ricevuti $\overline{x} = 9.5$. B aveva motivo di supporre che il valore inviato dovesse essere $8$. Si verifichi questa ipotesi.

1. Definiamo l'ipotesi nulla $H_0 : \mu = 8$
2. Definiamo l'ipotesi alternativa $H_1 : \mu \neq 8$
3. Scegliamo una statistica test $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{0.5}{2/\sqrt{5}} \simeq 0.559$
4. Calcolo il p-value
   $p = \mathbb{P}\{|\mathsf{z}| > 0.559\} = 2 \times \mathbb{P}\{\mathsf{z} > 0.559\} \simeq 2 \times 0.288 = 0.576$

# Esempio

Un segnale di valore $\mu$ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore normale di media nulla e varianza $4$. Il segnale ricevuto da B ha quindi distribuzione $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato per $5$ volte lo stesso segnale: la media campionaria dei segnali ricevuti $\overline{x} = 9.5$. B aveva motivo di supporre che il valore inviato dovesse essere $8$. Si verifichi questa ipotesi.

1. Definiamo l'ipotesi nulla $H_0 : \mu = 8$
2. Definiamo l'ipotesi alternativa $H_1 : \mu \neq 8$
3. Scegliamo una statistica test $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{0.5}{2/\sqrt{5}} \simeq 0.559$
4. Calcolo il p-value
   $p = \mathbb{P}\{|\mathsf{z}| > 0.559\} = 2 \times \mathbb{P}\{\mathsf{z} > 0.559\} \simeq 2 \times 0.288 = 0.576$

# Esempio

Un segnale di valore $\mu$ trasmesso da una sorgente A, viene raccolto dal ricevente B con un rumore normale di media nulla e varianza $4$. Il segnale ricevuto da B ha quindi distribuzione $\mathcal{N}(\mu, 4)$. Per ridurre il rumore, viene inviato per $5$ volte lo stesso segnale: la media campionaria dei segnali ricevuti $\overline{x} = 9.5$. B aveva motivo di supporre che il valore inviato dovesse essere $8$. Si verifichi questa ipotesi.

1. Definiamo l'ipotesi nulla $H_0 : \mu = 8$
2. Definiamo l'ipotesi alternativa $H_1 : \mu \neq 8$
3. Scegliamo una statistica test $z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{0.5}{2/\sqrt{5}} \simeq 0.559$
4. Calcolo il p-value
   $p = \mathbb{P}\{|\mathsf{z}| > 0.559\} = 2 \times \mathbb{P}\{\mathsf{z} > 0.559\} \simeq 2 \times 0.288 = 0.576$

Sarebbe assurdo eseguire un test con un livello di significatività elevato come $0.576$. Accettiamo $H_0$.

# Caso $\sigma$ non noto

- Quando $\sigma$ non è noto, anche l'ipotesi $H_0 : \mu = \mu_0$ non è un'ipotesi semplice, in quanto non è possibile specificare completamente la distribuzione sotto l'ipotesi nulla.

# Caso $\sigma$ non noto

- Quando $\sigma$ non è noto, anche l'ipotesi $H_0 : \mu = \mu_0$ non è un'ipotesi semplice, in quanto non è possibile specificare completamente la distribuzione sotto l'ipotesi nulla.
- Anche questa volta l'intuito ci dice di rifiutare l'ipotesi nulla quando $\overline{x}$ cade lontano da $\mu_0$, ma quanto lontano?

# Caso $\sigma$ non noto

- Quando $\sigma$ non è noto, anche l'ipotesi $H_0 : \mu = \mu_0$ non è un'ipotesi semplice, in quanto non è possibile specificare completamente la distribuzione sotto l'ipotesi nulla.
- Anche questa volta l'intuito ci dice di rifiutare l'ipotesi nulla quando $\overline{x}$ cade lontano da $\mu_0$, ma quanto lontano?
- Se nel caso con $\sigma$ noto guardavamo a $\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|$, questa volta considereremo $\left|\frac{\overline{x}-\mu_0}{s/\sqrt{n}}\right|$ con

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

- Per ottenere un livello di significatività pari ad $\alpha$, partiamo dalla distribuzione della statistica del test assumendo $H_0$ e imponiamo che la probabilità di rifiutare l'ipotesi nulla sia $\alpha$ (e non più grande)

$$t = \frac{\bar{\mathsf{x}} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

- Per ottenere un livello di significatività pari ad $\alpha$, partiamo dalla distribuzione della statistica del test assumendo $H_0$ e imponiamo che la probabilità di rifiutare l'ipotesi nulla sia $\alpha$ (e non più grande)

$$t = \frac{\bar{\mathsf{x}} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

- Segue che

$$\text{si rifiuta } H_0 \text{ se } \left| \frac{\bar{\mathsf{x}} - \mu_0}{s/\sqrt{n}} \right| > t_{\frac{\alpha}{2}, n-1}$$

$$\text{non si rifiuta } H_0 \text{ se } \left| \frac{\bar{\mathsf{x}} - \mu_0}{s/\sqrt{n}} \right| \leqslant t_{\frac{\alpha}{2}, n-1}$$

# Test unilaterale (a una coda)

- Nel verificare l'ipotesi nulla $\mu = \mu_0$ abbiamo costruito un test che porta ad un rifiuto quando $\bar{x}$ è lontana da $\mu_0$, ovvero. Cosa accade invece quando $H_0 : \mu = \mu_0, \;\; H_1 : \mu > \mu_0$?
- In questo caso valori molto bassi di $\bar{x}$ non ci dovrebbero fare rifiutare l'ipotesi nulla (è più probabile ottenere una $\bar{x}$ piccola quando è vera $H_0$ che non quando  vera $H_1$)

# Test unilaterale (a una coda)

1. Definiamo l'ipotesi nulla $H_0 : \mu = \mu_0$
2. Definiamo l'ipotesi alternativa $H_1 : \mu > \mu_0$
3. Scegliamo una statistica test $\mathsf{z} = \frac{\bar{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$

# Test unilaterale (a una coda)

4. Definiamo un livello di significatività e quindi la regione critica
   Scegliamo $\alpha$ e quindi definiamo la regione critica come

$$\mathcal{C} = \{x_1, x_2, \ldots, x_n : \overline{x} - \mu_0 > c\}$$
$$\mathbb{P}\{\bar{\mathsf{x}} - \mu_0 > c\,; \mu = \mu_0\} = \alpha$$

   Chi è $c$?

# Test unilaterale (a una coda)

4. Definiamo un livello di significatività e quindi la regione critica
   Scegliamo $\alpha$ e quindi definiamo la regione critica come

$$\mathcal{C} = \{x_1, x_2, \ldots, x_n : \overline{x} - \mu_0 > c\}$$

$$\mathbb{P}\{\bar{\mathsf{x}} - \mu_0 > c\,; \mu = \mu_0\} = \alpha$$

   Chi è $c$?
   Considerando che $\frac{\bar{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$ assumendo $H_0$, allora

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

# Test unilaterale (a una coda)

4. Definiamo un livello di significatività e quindi la regione critica
   Scegliamo $\alpha$ e quindi definiamo la regione critica come

$$\mathcal{C} = \{x_1, x_2, \ldots, x_n : \overline{x} - \mu_0 > c\}$$
$$\mathbb{P}\{\bar{\mathsf{x}} - \mu_0 > c\,; \mu = \mu_0\} = \alpha$$

Chi è $c$?

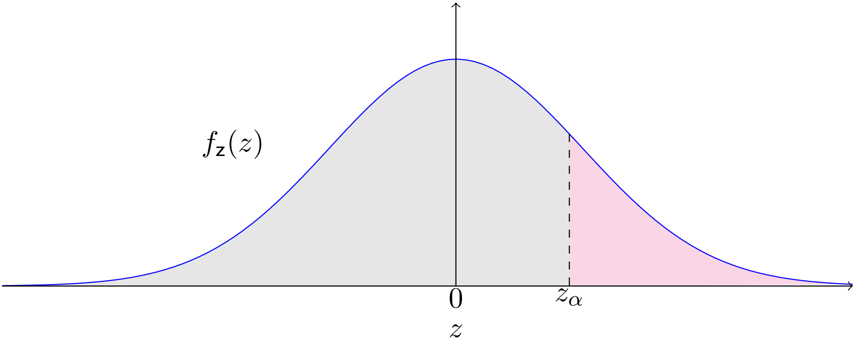
Considerando che $\frac{\bar{\mathsf{x}}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ assumendo $H_0$, allora

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

Quindi, di fatto,

- Rifiuto $H_0$ se $\frac{\bar{\mathsf{x}}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$
- Non rifiuto $H_0$ (accettazione) se $\frac{\bar{\mathsf{x}}-\mu_0}{\sigma/\sqrt{n}} \leqslant z_\alpha$

# Regione critica




Accetto $H_0$ se

$$\frac{\bar{\mathsf{x}} - \mu_0}{\sigma/\sqrt{n}} \leqslant z_\alpha$$

# Indice

- Verifica delle ipotesi
- Verifica delle ipotesi sulla media di una popolazione normale
- Verifica delle ipotesi su una popolazione di Bernoulli

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

Supponiamo che $x_1, x_2, \ldots, x_n$ sia un campione aleatorio proveniente da una popolazione di Bernoulli con parametro $p$

1. Definiamo l'ipotesi nulla $H_0 : p = p_0$

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

Supponiamo che $x_1, x_2, \ldots, x_n$ sia un campione aleatorio proveniente da una popolazione di Bernoulli con parametro $p$

1. Definiamo l'ipotesi nulla $H_0 : p = p_0$
2. Definiamo l'ipotesi alternativa $H_1 : p \neq p_0$

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

Supponiamo che $x_1, x_2, \ldots, x_n$ sia un campione aleatorio proveniente da una popolazione di Bernoulli con parametro $p$

1. Definiamo l'ipotesi nulla $H_0 : p = p_0$
2. Definiamo l'ipotesi alternativa $H_1 : p \neq p_0$
3. Scegliamo una statistica test Siccome $\mathsf{y} = \sum_{i=1}^n \mathsf{x}_n \sim \mathsf{Bin}(n, p)$ e $\overline{x} = \mathsf{y}/n$ è lo stimatore puntuale naturale per $p$ sembra ragionevole accettare $H_0$ quando $\overline{x}$ non è troppo lontano da $p_0$, guardando quindi a $|\mathsf{y} - np_0|$

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

- Quando la numerosità del campione è elevata possiamo ottenere un test approssimato con significatività $\alpha$, utilizzando la distribuzione normale.

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

- Quando la numerosità del campione è elevata possiamo ottenere un test approssimato con significatività $\alpha$, utilizzando la distribuzione normale.
- Quando $n$ è abbastanza grande y è approssimativamente normale, con $\mathbb{E}\{\mathrm{y}\} = np$ e varianza $\mathbb{V}\{\mathrm{y}\} = np(1-p)$

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

- Quando la numerosità del campione è elevata possiamo ottenere un test approssimato con significatività $\alpha$, utilizzando la distribuzione normale.
- Quando $n$ è abbastanza grande y è approssimativamente normale, con $\mathbb{E}\left\{\mathrm{y}\right\} = np$ e varianza $\mathbb{V}\{\mathrm{y}\} = np(1-p)$
- Segue che $\frac{\overline{x}-p}{\sqrt{p(1-p)/n}} \sim \mathcal{N}(0,1)$

# Verifica delle ipotesi su una popolazione di Bernoulli (proporzione)

- Quando la numerosità del campione è elevata possiamo ottenere un test approssimato con significatività $\alpha$, utilizzando la distribuzione normale.
- Quando $n$ è abbastanza grande y è approssimativamente normale, con $\mathbb{E}\{\mathrm{y}\} = np$ e varianza $\mathbb{V}\{\mathrm{y}\} = np(1-p)$
- Segue che $\frac{\overline{x}-p}{\sqrt{p(1-p)/n}} \sim \mathcal{N}(0,1)$
- Si rifiuta l'ipotesi nulla quando $\left|\frac{\overline{\mathrm{x}}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| \geqslant z_{\alpha/2}$

# Esempio

Supponiamo di avere intervistato un campione casuale di $n = 1200$ cittadini e che $k = 631$ di loro abbiano espresso l'intenzione di rieleggere il Sindaco: abbiamo abbastanza evidenza per concludere che il Sindaco dispone di una maggioranza?

1. Definiamo l'ipotesi nulla $H_0 : p \leqslant 0.5$

# Esempio

Supponiamo di avere intervistato un campione casuale di $n = 1200$ cittadini e che $k = 631$ di loro abbiano espresso l'intenzione di rieleggere il Sindaco: abbiamo abbastanza evidenza per concludere che il Sindaco dispone di una maggioranza?

1. Definiamo l'ipotesi nulla $H_0 : p \leqslant 0.5$
2. Definiamo l'ipotesi alternativa $H_1 : p > 0.5$

# Esempio

Supponiamo di avere intervistato un campione casuale di $n = 1200$ cittadini e che $k = 631$ di loro abbiano espresso l'intenzione di rieleggere il Sindaco: abbiamo abbastanza evidenza per concludere che il Sindaco dispone di una maggioranza?

1. Definiamo l'ipotesi nulla $H_0 : p \leqslant 0.5$
2. Definiamo l'ipotesi alternativa $H_1 : p > 0.5$
3. Scegliamo una statistica test $\overline{x} = k/n = 0.52$; la statistica test osservata è $z = \frac{\overline{x} - p_0}{\sqrt{p_0(1-p_0)/n}} = 1.79$

# Esempio

Supponiamo di avere intervistato un campione casuale di $n = 1200$ cittadini e che $k = 631$ di loro abbiano espresso l'intenzione di rieleggere il Sindaco: abbiamo abbastanza evidenza per concludere che il Sindaco dispone di una maggioranza?

1. Definiamo l'ipotesi nulla $H_0 : p \leqslant 0.5$
2. Definiamo l'ipotesi alternativa $H_1 : p > 0.5$
3. Scegliamo una statistica test $\overline{x} = k/n = 0.52$; la statistica test osservata è $z = \frac{\overline{x}-p_0}{\sqrt{p_0(1-p_0)/n}} = 1.79$
4. Calcolo il p-value $p = \mathbb{P}\{z > 1.79\} = 0.037$

# Esempio

Supponiamo di avere intervistato un campione casuale di $n = 1200$ cittadini e che $k = 631$ di loro abbiano espresso l'intenzione di rieleggere il Sindaco: abbiamo abbastanza evidenza per concludere che il Sindaco dispone di una maggioranza?

1. Definiamo l'ipotesi nulla $H_0 : p \leqslant 0.5$
2. Definiamo l'ipotesi alternativa $H_1 : p > 0.5$
3. Scegliamo una statistica test $\overline{x} = k/n = 0.52$; la statistica test osservata è $z = \frac{\overline{x} - p_0}{\sqrt{p_0(1-p_0)/n}} = 1.79$
4. Calcolo il p-value $p = \mathbb{P}\{z > 1.79\} = 0.037$

# Esempio

Supponiamo di avere intervistato un campione casuale di $n = 1200$ cittadini e che $k = 631$ di loro abbiano espresso l'intenzione di rieleggere il Sindaco: abbiamo abbastanza evidenza per concludere che il Sindaco dispone di una maggioranza?

1. Definiamo l'ipotesi nulla $H_0 : p \leqslant 0.5$
2. Definiamo l'ipotesi alternativa $H_1 : p > 0.5$
3. Scegliamo una statistica test $\overline{x} = k/n = 0.52$; la statistica test osservata è $z = \frac{\overline{x}-p_0}{\sqrt{p_0(1-p_0)/n}} = 1.79$
4. Calcolo il p-value $p = \mathbb{P}\{z > 1.79\} = 0.037$

Rifiutiamo $H_0$ con $\alpha = 0.05$ e $p = 0.037$

# Relazione tra Verifica delle ipotesi e intervalli di confidenza

Abbiamo dimostrato nelle lezioni precedenti che un intervallo di confidenza bilaterale ad un livello $1-\alpha$ per la media di una distribuzione normale con varianza nota $\sigma^2$, è dato da

$$\mu \in \left(\bar{\mathsf{x}} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{\mathsf{x}} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Intendendo che

$$\mathbb{P}\left\{\bar{\mathsf{x}} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{\mathsf{x}} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\right\} = 1-\alpha$$

# Relazione tra Verifica delle ipotesi e intervalli di confidenza

Eseguendo una verifica sulle ipotesi bilaterali $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ con un livello di significatività $\alpha$, quello che facciamo è accettare l'ipotesi nulla quando quando

$$\mu_0 \in \left( \bar{\mathsf{x}} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{\mathsf{x}} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

# Relazione tra Verifica delle ipotesi e intervalli di confidenza

Eseguendo una verifica sulle ipotesi bilaterali $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ con un livello di significatività $\alpha$, quello che facciamo è accettare l'ipotesi nulla quando quando

$$\mu_0 \in \left( \bar{\mathsf{x}} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{\mathsf{x}} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

E ritroviamo che

$$\mathbb{P} \left\{ \bar{\mathsf{x}} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leqslant \mu_0 \leqslant \bar{\mathsf{x}} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \right\} = 1 - \alpha$$

# Relazione tra Verifica delle ipotesi e intervalli di confidenza

Eseguendo una verifica sulle ipotesi bilaterali $H_0 : \mu = \mu_0$ contro $H_1 : \mu \neq \mu_0$ con un livello di significatività $\alpha$, quello che facciamo è accettare l'ipotesi nulla quando quando

$$\mu_0 \in \left(\bar{\mathsf{x}} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{\mathsf{x}} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

E ritroviamo che

$$\mathbb{P}\left\{\bar{\mathsf{x}} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leqslant \mu_0 \leqslant \bar{\mathsf{x}} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\right\} = 1 - \alpha$$

Abbiamo definito gli intervalli di confidenza per la stima del parametro, ma di fatto possiamo costruire un test delle ipotesi partendo dall'intervallo di confidenza con livello $1 - \alpha$. Accettiamo l'ipotesi nulla se il parametro rientra all'interno dell'intervallo.

# Schema per la verifica di ipotesi