METODI STATISTICI PER LA BIOINGEGNERIA (B)

PARTE 5: VERIFICA DELLE IPOTESI MEDIANTE TEST STATISTICI PARAMETRICI (SECONDA PARTE)

A.A. 2025-2026

Prof. Martina Vettoretti

ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- Test per verificare se due popolazioni normali hanno la stessa varianza

VERIFICA DI IPOTESI SULLA MEDIA DI UNA POPOLAZIONE NORMALE CON VARIANZA INCOGNITA

- Campione aleatorio $X_1, X_2, \dots, X_n$ estratto da una popolazione normale con media μ e varianza σ^2 entrambe incognite (caso molto più comune!).

- Vogliamo verificare il sistema di ipotesi bilaterale:
  - $H_0: \mu = \mu_0$
  - $H_1: \mu \neq \mu_0$

- Il test statistico che possiamo utilizzare in questa situazione si chiama **t test a un campione bilaterale.**

- In Matlab è implementato dalla funzione ttest.

LA STATISTICA DEL T TEST A UN CAMPIONE

- Per il test statistico con σ^2 noto (z test) abbiamo utilizzato come statistica del test la variabile aleatoria:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- In questo caso σ^2 è incognito. Possiamo però stimare questo valore a partire dai dati utilizzando la deviazione standard campionaria, S, come stimatore:

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

- La statistica del test risulta quindi la variabile aleatoria T così definita:

$$T := \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

T TEST BILATERALE: LA REGIONE CRITICA (1/2)

- Si può dimostrare che quando H_0 è vera ($\mu = \mu_0$), la variabile aleatoria T ha distribuzione **t di Student con n-1 gradi di libertà**.

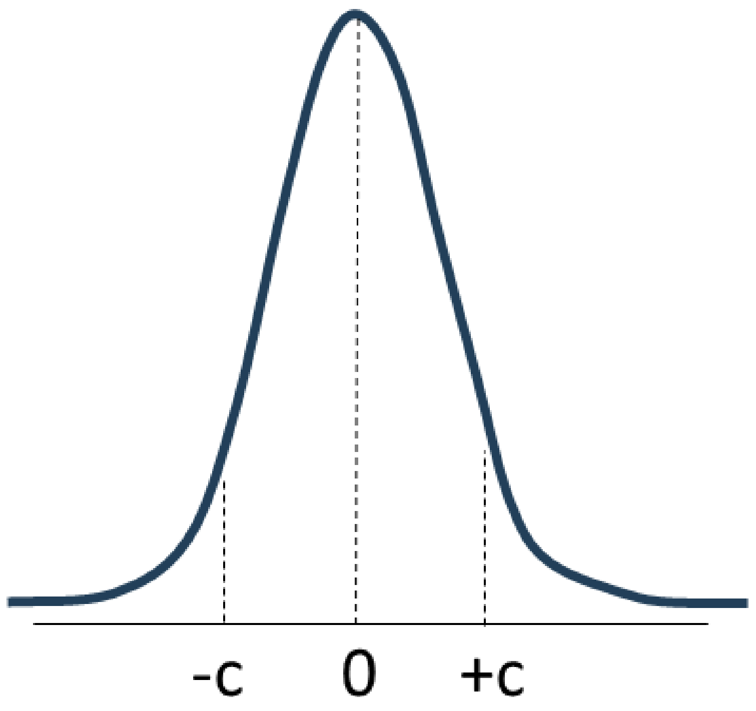
$$\text{Se } \mu = \mu_0 \rightarrow T \sim t_{n-1}$$

- **Definizione della regione critica:** rifiutiamo H_0 quando il valore osservato per T è in valore assoluto più grande di una certa quantità c

$\rightarrow$ μ è sufficientemente distante da μ_0

$$R := \{T : |T| > c\}$$

Ddp di una t di Student



T TEST BILATERALE: LA REGIONE CRITICA (2/2)

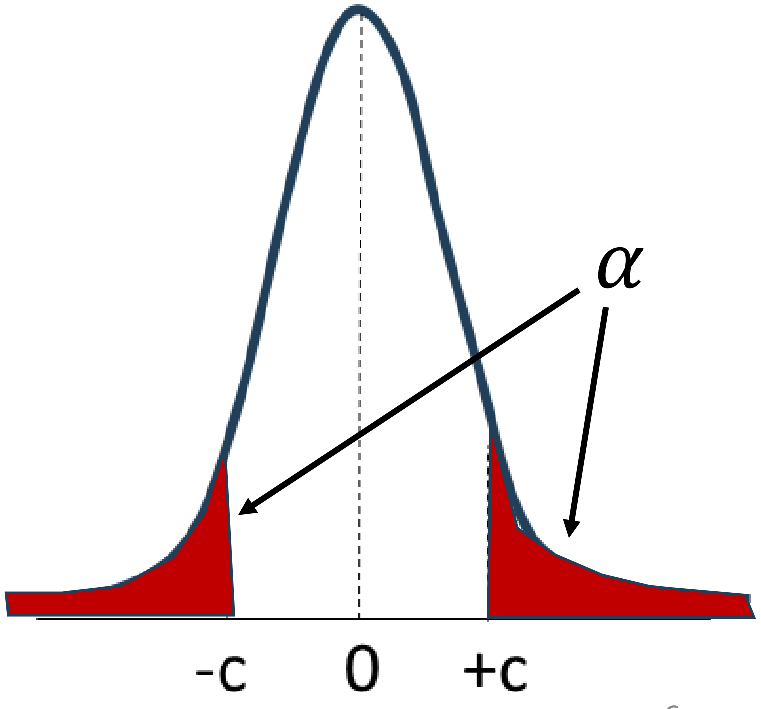
- **Livello di significatività scelto pari α** → Per garantire che la probabilità di rifiutare H_0 quando H_0 è vera (errore di prima specie) sia α, dobbiamo imporre che:
  - $P(|T| > c) = \alpha$
  - $P(T > c) + P(T \leq -c) = 2 \cdot P(T > c) = \alpha$
  - $P(T > c) = \alpha/2$

Per la simmetria della t di Student

c è il quantile della t di Student con n-1 gradi di libertà di probabilità α/2:

$$c = t_{\alpha/2,n-1}$$

Ddp di una t di Student



T TEST BILATERALE: REGOLA DECISIONALE (1/2)

- Se $|T| > t_{\frac{\alpha}{2},n-1}$ → rifiutiamo H_0 (quindi accettiamo H_1)
  - Possiamo dire che μ è significativamente diverso da μ_0

- Se $|T| \leq t_{\frac{\alpha}{2},n-1}$ → non possiamo rifiutare H_0 (quindi «accettiamo» H_0)
  - Non possiamo dire nulla su μ

T TEST BILATERALE: REGOLA DECISIONALE (2/2)

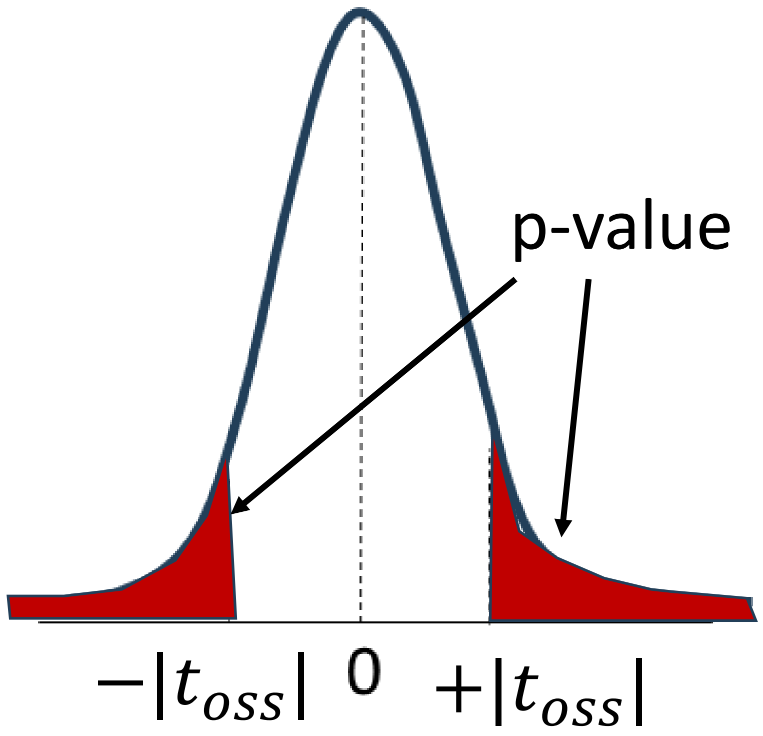
Regola basata sul p-value:

$$t_{oss} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- $\bar{x}$ e s: media e deviazione standard campionarie calcolate a partire dalle osservazioni del campione $x_1, x_2, \dots, x_n$.

$$p - value = P(|T| > |t_{oss}|) = 2 \cdot P(T > |t_{oss}|)$$

- Se p-value $< \alpha$ → rifiutiamo H_0
- Se p-value $\geq \alpha$ → non possiamo rifiutare H_0



ESERCIZIO 1

La quantità di principio attivo contenuta nelle pillole vendute da una ditta farmaceutica può avere una certa variabilità, comunque il suo valore medio dichiarato è 20 mg. Per convalidare questa affermazione, si misura la quantità di principio attivo in un campione di 25 pillole, trovando:

- media campionaria di 19.7 mg
- deviazione standard campionaria di 1.3 mg.

Che conclusioni possiamo trarre dai dati? Con un livello di significatività del 5%, possiamo dire che i dati dell'esperimento dimostrano che l'affermazione della ditta farmaceutica non è vera?

Nota: assumere che i dati provengano da una distribuzione normale.

SOLUZIONE

- Assumiamo che il campione sia normale con media μ e varianza σ^2 incognite.
- Applichiamo il t test a un campione bilaterale per verificare il sistema di ipotesi:
  - $H_0 : \mu = 20\, mg$
  - $H_1 : \mu \neq 20\, mg$
- Valore osservato per la statistica del test:

$$t_{oss} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{19.7 - 20.0}{1.3/\sqrt{25}} = -1.1538$$

- Con α=5%, $t_{\alpha/2,n-1} = t_{0.025,24} = 2.064$ → (In Matlab: tinv(1-alpha/2,n-1))
- $|t_{oss}| \leq t_{\alpha/2,n-1}$ → non possiamo rifiutare H_0

  → possiamo quindi confutare l'affermazione fatta dall'azienda farmaceutica?

T TEST A UN CAMPIONE UNILATERALE O A UNA CODA (1/2)

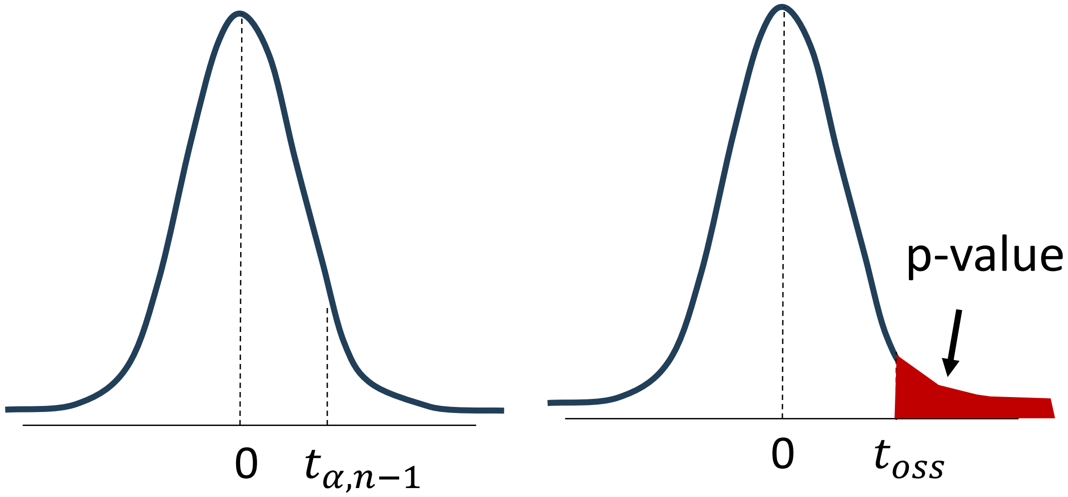
- Campione aleatorio $X_1, X_2, \dots, X_n$ estratto da una popolazione normale con media e varianza entrambe incognite.
- Vogliamo verificare il sistema di ipotesi unilaterale:
  - $H_0: \mu = \mu_0$ (o $H_0: \mu \leq \mu_0$)
  - $H_1: \mu > \mu_0$

  con un livello di significatività α.

$$\boxed{t_{oss} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}}$$

→ **t test a un campione unilaterale** o a **una coda**

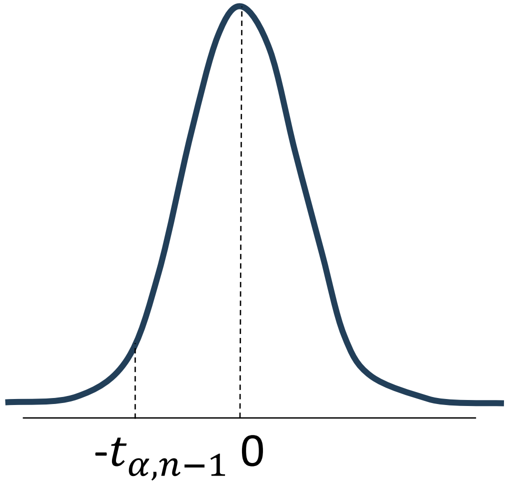
- Regola decisionale del test:
  - Se $t_{oss} > t_{\alpha,n-1} \rightarrow$ rifiutiamo H_0
  - Se $t_{oss} \leq t_{\alpha,n-1} \rightarrow$ non possiamo rifiutare H_0



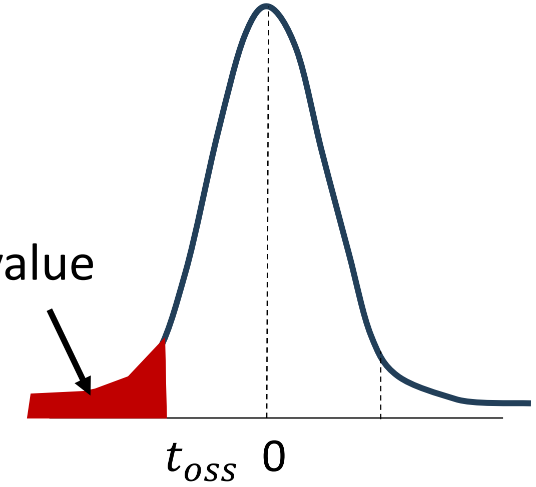
T TEST A UN CAMPIONE UNILATERALE O A UNA CODA (2/2)

- Analogamente, se il sistema di ipotesi è:
  - $H_0: \mu = \mu_0$ (o $H_0: \mu \geq \mu_0$)
  - $H_1: \mu < \mu_0$

$$\boxed{t_{oss} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}}$$



- La regola decisionale del test diventa:
  - Se $t_{oss} < -t_{\alpha,n-1}$ → rifiutiamo H_0
  - Se $t_{oss} \geq -t_{\alpha,n-1}$ → non possiamo rifiutare H_0



ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- Test per verificare se due popolazioni normali hanno la stessa varianza

VERIFICARE SE DUE POPOLAZIONI NORMALI HANNO LA STESSA MEDIA – CASO CON VARIANZE NOTE

- Supponiamo di disporre di due campioni aleatori indipendenti provenienti da due popolazioni normali con medie incognite e varianze note.

$$X_1, X_2, \dots, X_n \quad X_i \sim N(\mu_X, \sigma_X^2)$$

$$Y_1, Y_2, \dots, Y_m \quad Y_i \sim N(\mu_Y, \sigma_Y^2)$$

μ_X, μ_Y → parametri incogniti

σ_X^2, σ_Y^2 → parametri noti

- Ipotizziamo di voler verificare il sistema di ipotesi bilaterale:
  - $H_0: \mu_X = \mu_Y$
  - $H_1: \mu_X \neq \mu_Y$

➡ **z test bilaterale a due campioni**

LA STATISTICA DEL TEST

- Come statistica del test ci serve una quantità che rappresenti quanto sono distanti tra loro μ_X e μ_Y. Pensando di stimare μ_X con $\bar{X}$ e μ_Y con $\bar{Y}$, potremmo considerare la variabile aleatoria:

$$V = \bar{X} - \bar{Y}, \quad V \sim N\left(\mu_X - \mu_y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$$

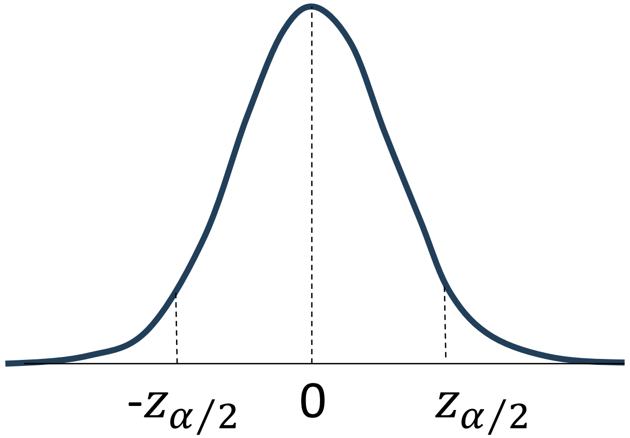
- Quando vale H_0, $\mu_X - \mu_y = 0 \rightarrow V \sim N(0, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m})$.
- Standardizzando V otteniamo la statistica del test che nell'ipotesi:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

$\boxed{\text{Se vale } H_0 \rightarrow Z \sim N(0,1)}$

REGOLA DECISIONALE

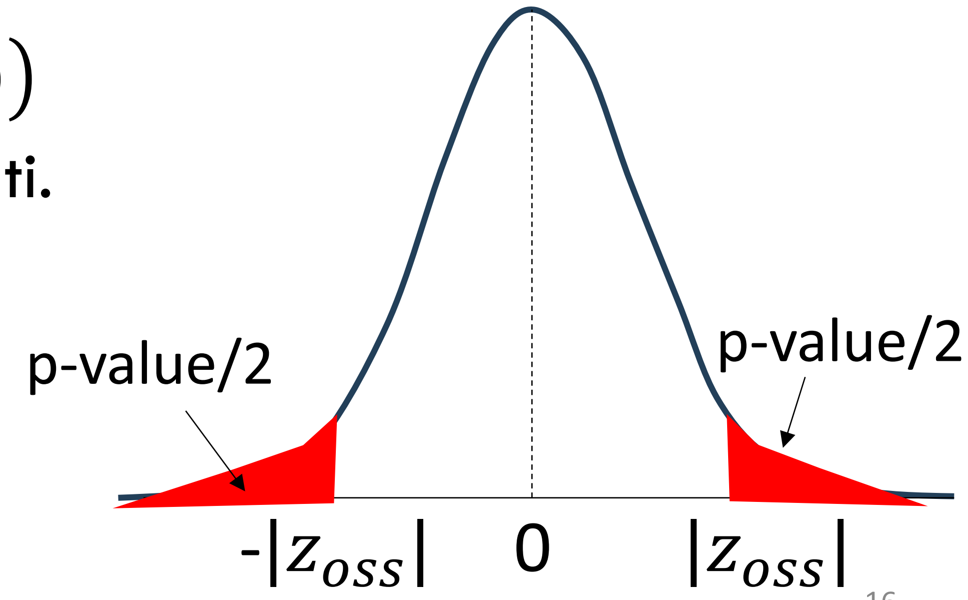
- Regola decisionale per livello di significatività α:
  - Se $|Z| > z_{\alpha/2}$ → rifiutiamo H_0 → μ_X e μ_Y significativamente diverse
  - Se $|Z| \leq z_{\alpha/2}$ → non possiamo rifiutare H_0 → non possiamo dire che μ_X e μ_Y siano significativamente diverse (ma neanche uguali!)



- $p - value = P(|Z| > |z_{oss}|) = 2 \cdot \left(1 - \Phi(|z_{oss}|)\right)$

dove z_{oss} è il valore di Z calcolato a partire dai dati.

- Regola decisionale basata sul p-value:
  - Se $p - value < \alpha$ → rifiutiamo H_0
  - Se $p - value \geq \alpha$ → non possiamo rifiutare H_0



ESERCIZIO 2

Uno scienziato che si occupa di inquinamento ambientale vuole verificare se due campioni di soluzioni in suo possesso hanno lo stesso pH. Per stabilire se questo sia vero, vengono fatte 10 misurazioni indipendenti per ciascuna soluzione. Il metodo di misura utilizzato garantisce che i valori misurati hanno distribuzione normale con deviazione standard pari a 0.05. I dati raccolti sono:

- Soluzione A: {6.24, 6.31, 6.28, 6.30, 6.25, 6.26, 6.24, 6.29, 6.22, 6.28}
- Soluzione B: {6.27, 6.25, 6.33, 6.27, 6.24, 6.31, 6.28, 6.29, 6.34, 6.27}

Possiamo dire che c'è una differenza significativa nel pH delle due soluzioni al 5% di significatività?

SOLUZIONE

- Assumiamo che i dati dei due campioni provengano da distribuzioni normali con medie μ_X e μ_Y incognite e varianze note $\sigma_X^2 = \sigma_Y^2 = 0.05^2$.
- Applichiamo lo z test bilaterale a due campioni per verificare il sistema di ipotesi:
  - $H_0: \mu_X = \mu_Y$
  - $H_1: \mu_X \neq \mu_Y$
- Calcoliamo il valore della statistica del test:

$$z_{oss} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{6.267 - 6.285}{\sqrt{2 \cdot 0.05^2/10}} = -0.805$$

- Con α=5%, $z_{\alpha/2} = z_{0.025} \cong 1.96$
- Poiché $|z_{oss}| < z_{\alpha/2}$, non possiamo rifiutare H_0 → non possiamo affermare che i pH delle due soluzioni siano significativamente diversi.

TEST UNILATERALE PER CONFRONTARE LA MEDIA DI DUE POPOLAZIONI NORMALI CON VARIANZA NOTA

Il test appena visto esiste anche nella sua versione unilaterale (o a una coda).

- Sistema di ipotesi:
  - $H_0: \mu_X = \mu_Y$ (o $\mu_X \leq \mu_Y$)
  - $H_1: \mu_X > \mu_Y$

$$\boxed{Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}}$$

- Regola decisionale:
  - se $Z > z_\alpha$ → rifiutiamo H_0
  - se $Z \leq z_\alpha$ → non possiamo rifiutare H_0

ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- Test per verificare se due popolazioni normali hanno la stessa varianza

VERIFICARE SE DUE POPOLAZIONI NORMALI HANNO LA STESSA MEDIA – CASO CON VARIANZE INCOGNITE MA UGUALI

- Supponiamo di disporre di due campioni aleatori indipendenti provenienti da due popolazioni normali con medie incognite e varianze incognite ma uguali →**popolazioni omoschedastiche** (hanno la stessa varianza).

$$X_1, X_2, \dots, X_n \quad X_i \sim N(\mu_X, \sigma^2)$$
$$Y_1, Y_2, \dots, Y_m \quad Y_i \sim N(\mu_Y, \sigma^2)$$

μ_X, μ_Y, σ^2 → parametri incogniti

- Vogliamo verificare il sistema di ipotesi bilaterale:
  - $H_0: \mu_X = \mu_Y$
  - $H_1: \mu_X \neq \mu_Y$

➡ **t test a due campioni bilaterale (campioni indipendenti e omoschedastici)**

- In Matlab è implementato dalla funzione ttest2.

STATISTICA DEL TEST

- Statistica del test:

$$T := \frac{\bar{X} - \bar{Y}}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- S_p^2 è detto stimatore «pooled» di σ^2:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

→ Media pesata di S_X^2 e S_Y^2, entrambi stimatori di σ^2.

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad S_Y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \bar{Y})^2$$

REGOLA DECISIONALE DEL TEST

Quando H_0 è vera ($\mu_X = \mu_Y$), T è distribuita come una t di Student con n+m-2 gradi di libertà.

$$\text{Se } \mu_X = \mu_Y \rightarrow T \sim t_{n+m-2}$$

Regola decisionale:

- Se $|T| > t_{\frac{\alpha}{2},n+m-2} \rightarrow$ rifiutiamo H_0 $\rightarrow$ μ_X e μ_Y sono significativamente diverse
- Se $|T| \leq t_{\frac{\alpha}{2},n+m-2} \rightarrow$ non possiamo rifiutare H_0 $\rightarrow$ non concludiamo nulla

$$p - value = P(|T| > |t_{oss}|) = 2 \cdot P(T > |t_{oss}|)$$

t_{oss} è il valore osservato per la statistica del test (calcolato sulla base dei dati).

- Se p-value $< \alpha$ $\rightarrow$ rifiutiamo H_0
- Se p-value $\geq \alpha$ $\rightarrow$ non possiamo rifiutare H_0

VERSIONE UNILATERALE DEL TEST

- Assunzioni:

  $X_1, X_2, \dots, X_n \quad X_i \sim N(\mu_X, \sigma^2)$

  $Y_1, Y_2, \dots, Y_m \quad Y_i \sim N(\mu_Y, \sigma^2)$

$\mu_X, \mu_Y, \sigma^2 \rightarrow$ parametri incogniti

$$T := \frac{\bar{X} - \bar{Y}}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- Sistema di ipotesi:
  - $H_0 : \mu_X \leq \mu_Y$
  - $H_1 : \mu_X > \mu_Y$
- Regola decisionale:
  - Se $\mathrm{T} > t_{\alpha, n+m-2} \rightarrow$ rifiutiamo $H_0 \rightarrow \mu_X$ è significativamente maggiore di μ_Y
  - Se $\mathrm{T} \leq t_{\alpha, n+m-2} \rightarrow$ non possiamo rifiutare $H_0 \rightarrow$ non concludiamo nulla
- Regola decisionale basata sul p-value:

$$p - value = P(T > t_{oss})$$

  - Se p-value $< \alpha \rightarrow$ rifiutiamo H_0
  - Se p-value $\geq \alpha \rightarrow$ non possiamo rifiutare H_0

ESERCIZIO 3

Un gruppo di 22 volontari presso un centro di ricerca medica viene esposto a vari tipi di virus influenzali e tenuto sotto controllo medico. Ad un campione casuale di 10 volontari viene somministrato un integratore con 4 grammi di vitamina C al giorno. Agli altri 12 volontari viene somministrato un placebo non distinguibile dall'integratore. Per ciascun soggetto viene registrata la durata della malattia in giorni. I dati raccolti sono:

- Trattati con vitamina C: {5.5, 6.0, 7.0, 6.0, 7.5, 6.0, 7.5, 5.5, 7.0, 6.5}
- Trattati con placebo: {6.5, 6.0, 8.5, 7.0, 6.5, 8.0, 7.5, 6.5, 7.5, 6.0, 8.5, 7.0}

Si può concludere che l'assunzione di 4 grammi di vitamina C al giorno abbia accorciato il decorso medio della malattia, con un livello di significatività al 5%?

SOLUZIONE

- Assumiamo che i dati dei due campioni provengano da distribuzioni normali con medie μ_c e μ_p incognite e varianze incognite ma uguali $\sigma_c^2 = \sigma_p^2 = \sigma^2$.
- Applichiamo il t test appena visto per verificare il sistema di ipotesi:
  - $H_0: \mu_c = \mu_p$
  - $H_1: \mu_p > \mu_c$
- Valore osservato per la statistica del test:

$$t_{oss} = \frac{\overline{X_p} - \overline{X_c}}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{7.125 - 6.450}{\sqrt{0.689} \cdot \sqrt{\frac{1}{10} + \frac{1}{12}}} = 1.90$$

Con α=5%, $t_{\alpha,n+m-2} = t_{0.05,20} \cong 1.725$

- Poiché $t_{oss} > t_{\alpha,n+m-2}$, possiamo rifiutare H_0 → la somministrazione di vitamina C ha ridotto significativamente la durata della malattie rispetto al gruppo placebo.

ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- Test per verificare se due popolazioni normali hanno la stessa varianza

VERIFICARE SE DUE POPOLAZIONI NORMALI HANNO LA STESSA MEDIA – CASO CON VARIANZE INCOGNITE E DIVERSE

- Due campioni aleatori indipendenti provenienti da due popolazioni normali con medie incognite e varianze incognite e diverse → **popolazioni eteroschedastiche** (hanno varianza diversa).

$$X_1, X_2, \dots, X_n \quad X_i \sim N(\mu_X, \sigma_X^2)$$
$$Y_1, Y_2, \dots, Y_m \quad Y_i \sim N(\mu_Y, \sigma_Y^2)$$

$\mu_X, \mu_Y, \sigma_X^2 \neq \sigma_Y^2$
parametri incogniti

- Sistema di ipotesi bilaterale:
  - $H_0: \mu_X = \mu_Y$
  - $H_1: \mu_X \neq \mu_Y$
- Problema statistico che non ha una soluzione: non è ancora stato trovato un test statistico che permette di verificare questa ipotesi per popolazioni normali eteroschedastiche con livello di significatività esattamente pari ad α → problema di Behrens-Fisher
- Esistono dei test che ne danno una soluzione approssimata → **t test di Welch**

T TEST DI WELCH

- Statistica del test:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

- Distribuzione della statistica del test quando vale H_0:

$$T \sim t_\nu$$

$$\nu \approx \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}$$

- Regola decisionale:
  - Se $|T| > t_{\frac{\alpha}{2},\nu} \rightarrow$ rifiutiamo H_0
  - Se $|T| \leq t_{\frac{\alpha}{2},\nu} \rightarrow$ non possiamo rifiutare H_0

ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- Test per verificare se due popolazioni normali hanno la stessa varianza

IL CASO IN CUI ABBIAMO COPPIE DI DATI

- Supponiamo di disporre di un insieme di n coppie di dati:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

- Stiamo osservando due campioni di dati, **X** e **Y**, dipendenti tra loro, per cui ad ogni elemento del campione **X**, corrisponde un elemento del campione **Y**.

$$\boldsymbol{X} = (X_1, X_2, \dots, X_n)$$

$$\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)$$

- Per esempio i due campioni potrebbero rappresentare gli esiti di due misurazioni effettuate sulle stesse unità statistiche.

ESEMPIO

- Vogliamo valutare l'efficacia di un programma di attività fisica sulla riduzione dell'indice di massa corporea (BMI) in soggetti obesi. Reclutiamo 50 soggetti obesi e misuriamo il loro BMI al reclutamento e dopo il completamento del programma di attività fisica. Otteniamo due campioni di dati appaiati: il primo campione, **X**, contiene le misure di BMI al reclutamento; il secondo campione, **Y**, contiene le misure di BMI raccolte per gli stessi soggetti del campione **X** al completamento del programma di attività fisica.
- Ipotizziamo di voler verificare se il programma di attività fisica ha avuto un impatto significativo sul BMI. Che test possiamo usare?
- I test che abbiamo visto finora ipotizzano che i due campioni siano indipendenti, non sono quindi appropriati quando si analizzano coppie di dati.

➡ **t test per dati appaiati (paired t test)**

IL T TEST PER DATI APPAIATI (1/3)

- Supponiamo che i due campioni di dati appaiati siano:

$$\boldsymbol{X} = (X_1, X_2, \dots, X_n)$$

$$\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)$$

- Si definiscono n variabili aleatorie $W_i, i = 1, \dots, n$ pari alla differenza tra gli elementi di X e i rispettivi elementi di Y:

$$W_i = X_i - Y_i$$

- Il test assume che le variabili W_i siano <u>normali</u> di media μ_W e varianza σ_W^2 entrambe incognite.

IL T TEST PER DATI APPAIATI (2/3)

- Sistema di ipotesi bilaterale:
  - $H_0: \mu_W = 0 \rightarrow \mu_X = \mu_Y$
  - $H_1: \mu_W \neq 0 \rightarrow \mu_X \neq \mu_Y$

Applicato al nostro esempio:
- se vale H_0 → il programma di attività fisica non ha avuto un impatto sul BMI medio.
- se vale H_1 → il programma di attività fisica ha portato ad una variazione significativa del BMI medio.

- Statistica del test:

$$T = \frac{\bar{W}}{S_W}\sqrt{n}$$

$$\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i\,, \qquad S_W = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(W_i - \bar{W}\,)^2}$$

IL T TEST PER DATI APPAIATI (3/3)

- Se vale l'ipotesi nulla, ovvero $\mu_W = 0$, la statistica del test ha distribuzione t di Student con n-1 gradi di libertà.

$$T \sim t_{n-1}$$

- Regola decisionale:
  - Se $|T| > t_{\frac{\alpha}{2},n-1}$ → rifiutiamo H_0
  - Se $|T| \leq t_{\frac{\alpha}{2},n-1}$ → non possiamo rifiutare H_0
- In Matlab il t test per dati appaiati si può implementare con la funzione ttest.

ESERCIZIO 4 (DA PROVARE IN AUTONOMIA)

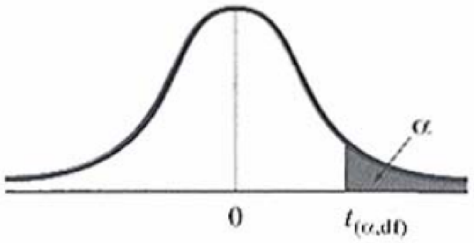
A 10 donne incinte è stata somministrata una iniezione di ossitocina per stimolare il travaglio. Le pressioni sanguigne sistoliche immediatamente prima e dopo la somministrazione sono state:

Paziente	1	2	3	4	5	6	7	8	9	10
Prima	134	122	132	130	128	140	118	127	125	142
Dopo	140	130	135	126	134	138	124	126	132	144

Con un livello di significatività al 5%, possiamo dire che l'ossitocina ha alterato in modo significativo la pressione sistolica?

TABELLA DEI QUANTILI DELLA T DI STUDENT

Tavola della distribuzione T di Student



Gradi di libertà	Area nella coda di destra								
	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	3.078	6.314	12.706	15.894	31.821	63.656	127.321	318.289	636.578
2	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.214	12.924
4	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.894	6.869
6	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850

ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- Test per verificare se due popolazioni normali hanno la stessa varianza

VERIFICA DI IPOTESI SULLA VARIANZA DI UNA POPOLAZIONE NORMALE

- Sia $X_1, X_2, \dots, X_n$ un campione aleatorio con distribuzione <u>normale</u> avente media μ e varianza σ^2 entrambe incognite.
- Sistema di ipotesi bilaterale:
  - $H_0: \sigma^2 = \sigma_0^2$
  - $H_1: \sigma^2 \neq \sigma_0^2$
- Statistica del test: $Y = \frac{S^2}{\sigma_0^2}(n-1)$, dove S^2 è la varianza campionaria.
- Quando vale H_0, Y è una chi-quadro con n-1 gradi di libertà:

$$Y \sim \chi^2_{n-1}$$

- Regola decisionale:
  - Se $\mathrm{Y} < \chi^2_{1-\frac{\alpha}{2},n-1}$ o $\mathrm{Y} > \chi^2_{\frac{\alpha}{2},n-1}$ → rifiutiamo H_0
  - Se $\chi^2_{1-\frac{\alpha}{2},n-1} \leq \mathrm{Y} \leq \chi^2_{\frac{\alpha}{2},n-1}$ → non possiamo rifiutare H_0

VERIFICA DI IPOTESI SULLA VARIANZA DI UNA POPOLAZIONE NORMALE – TEST UNILATERALE

- Domanda: come formuleresti una versione unilaterale del test per la verifica di ipotesi sulla varianza di una popolazione normale?

ALCUNI TEST STATISTICI PARAMETRICI

- Test per verificare ipotesi sulla media di una popolazione normale
  - Caso in cui la varianza è nota
  - Caso in cui la varianza è incognita: il t test
- Test per verificare se due popolazioni normali hanno la stessa media
  - Caso in cui le varianze sono note
  - Caso in cui le varianze sono incognite ma uguali
  - Caso in cui le varianze sono incognite e diverse
  - Caso in cui le due popolazioni sono dipendenti
- Test per verificare ipotesi sulla varianza di una popolazione normale
- $\boxed{\text{Test per verificare se due popolazioni normali hanno la stessa varianza}}$

VERIFICARE SE DUE POPOLAZIONI NORMALI HANNO LA STESSA VARIANZA (1/2)

- Supponiamo di disporre di due campioni aleatori indipendenti provenienti da due popolazioni normali con medie incognite e varianze incognite.

$$X_1, X_2, \dots, X_n \quad X_i \sim N(\mu_X, \sigma_X^2)$$
$$Y_1, Y_2, \dots, Y_m \quad Y_i \sim N(\mu_Y, \sigma_Y^2)$$

- Sistema di ipotesi bilaterale:
  - $H_0: \sigma_X^2 = \sigma_Y^2$
  - $H_1: \sigma_X^2 \neq \sigma_Y^2$
- Statistica del test:

$$F = \frac{S_X^2}{S_Y^2}$$

- S_X^2: varianza campionaria del campione X
- S_Y^2: varianza campionaria del campione Y

VERIFICARE SE DUE POPOLAZIONI NORMALI HANNO LA STESSA VARIANZA (2/2)

- Quando vale H_0, F ha distribuzione F di Fisher con n-1 ed m-1 gradi di libertà:

$$F \sim F_{n-1,m-1}$$

- Regola decisionale:
  - Se $\mathrm{F} < F_{1-\frac{\alpha}{2},n-1,m-1}$ o $F > F_{\frac{\alpha}{2},n-1,m-1}$ → rifiutiamo H_0
  - Se $F_{1-\frac{\alpha}{2},n-1,m-1} \leq F \leq F_{\frac{\alpha}{2},n-1,m-1}$ → non possiamo rifiutare H_0

- Domanda: come formuleresti una versione unilaterale del test?