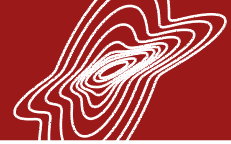


METODI STATISTICI PER LA BIOINGEGNERIA

A.A. 2024-2025

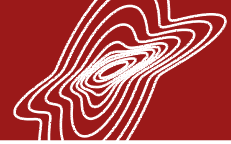
Prof. Alessandra Bertoldo

Ing. Mattia De Francisci, Ing. Claudia Tarricone



Passi fondamentali di un test di ipotesi

1. Verifica degli assunti di base
2. Formulazione dell'ipotesi statistica
3. Costruzione della statistica
4. Determinazione della distribuzione della statistica
5. Definizione della regola di decisione



1. Verifica degli assunti di base

2. Formulazione dell'ipotesi statistica

3. Costruzione della statistica

4. Determinazione della distribuzione della statistica

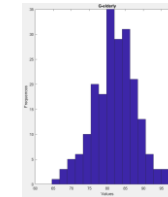
5. Definizione della regola di decisione

Dati due campioni (dimensione N_1 e N_2 ,
media m_1 e m_2 e standard deviation s_1 e s_2) si assume:

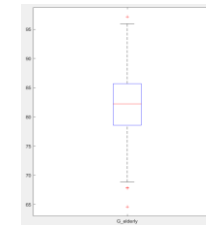
1. Campioni indipendenti
2. Estratti da popolazioni **gaussiane**
3. Varianza equivalente

Ispezione Visiva

▪ Istogramma
(*matlab: histogram*)

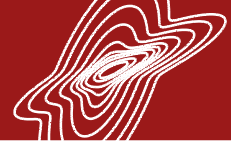


▪ Boxplot
(*matlab: boxplot*)



Valutazione quantitativa

- Skewness ≈ 0
(*matlab: skewness*)
- Media \approx mediana
- Kurtosis ≈ 3
(*matlab: kurtosis*)
- Test gaussianità (*matlab: lillietest*)



1. Verifica degli assunti di base

2. Formulazione dell'ipotesi statistica



3. Costruzione della statistica

4. Determinazione della distribuzione della statistica

5. Definizione della regola di decisione

Ipotesi nulla
(ipotesi sottoposta a statistica)

$$H_0: m_1 = m_2$$

Ipotesi alternativa

$$H_1: m_1 \neq m_2$$

Il test è a una o due code?



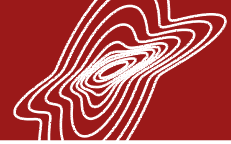
1. Verifica degli assunti di base
2. Formulazione dell'ipotesi statistica
3. Costruzione della statistica
4. Determinazione della distribuzione della statistica
5. Definizione della regola di decisione



Statistica t:

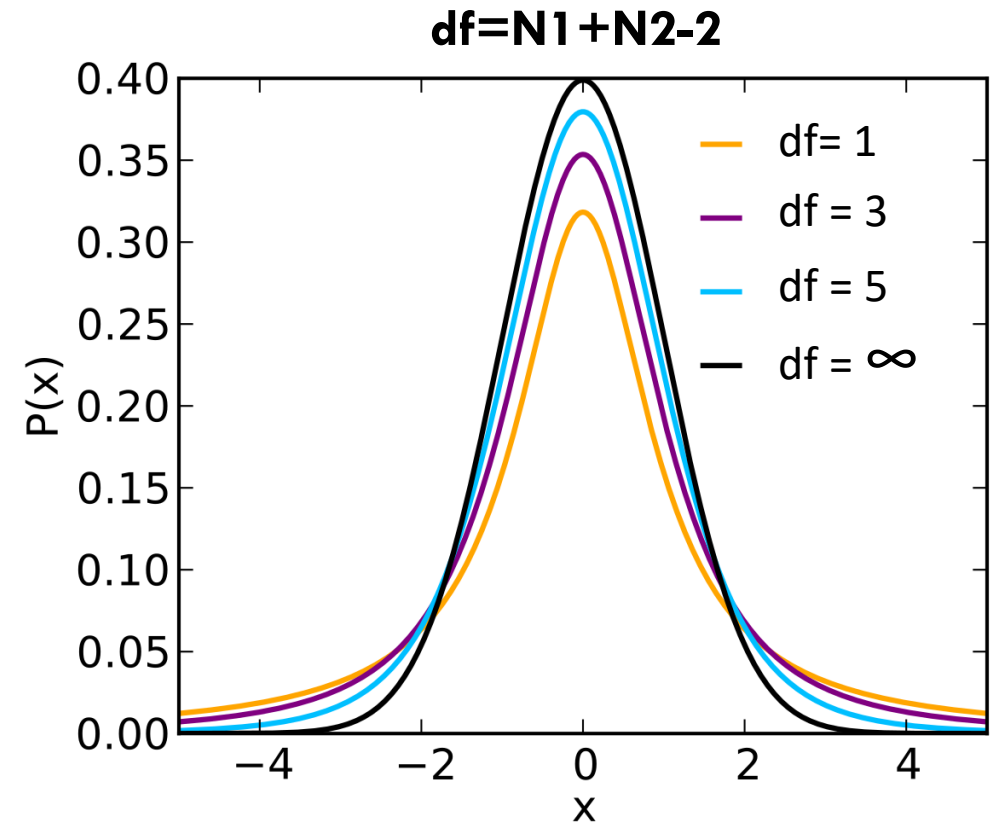
$$t = \frac{m_1 - m_2}{s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

dove $s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$



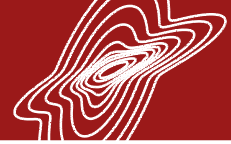
1. Verifica degli assunti di base
2. Formulazione dell'ipotesi statistica
3. Costruzione della statistica
4. Determinazione della distribuzione della statistica
5. Definizione della regola di decisione

Se H_0 è verificata la statistica t ha una distribuzione nota, detta **distribuzione di Student**, la cui forma dipende unicamente dai gradi di libertà



matlab: **tpdf**

per il calcolo della distribuzione di Student



1. Verifica degli assunti di base

2. Formulazione dell'ipotesi statistica

3. Costruzione della statistica

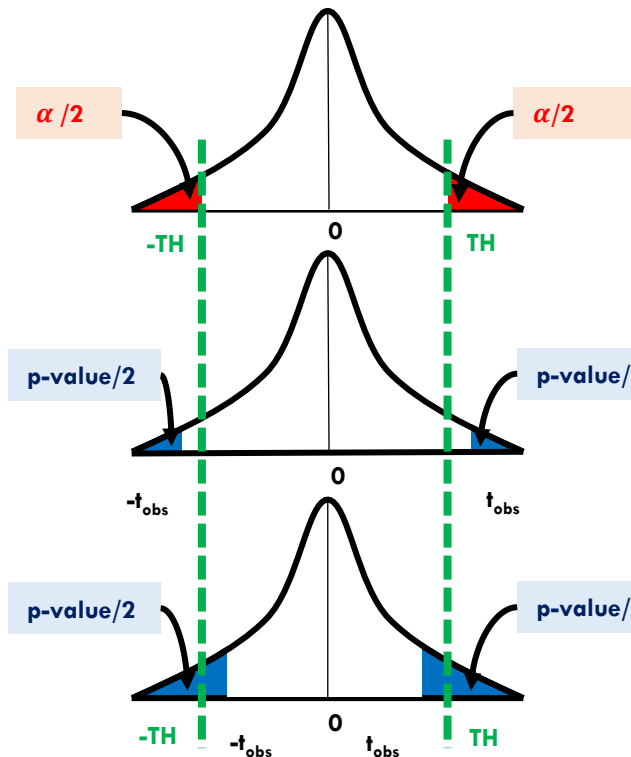
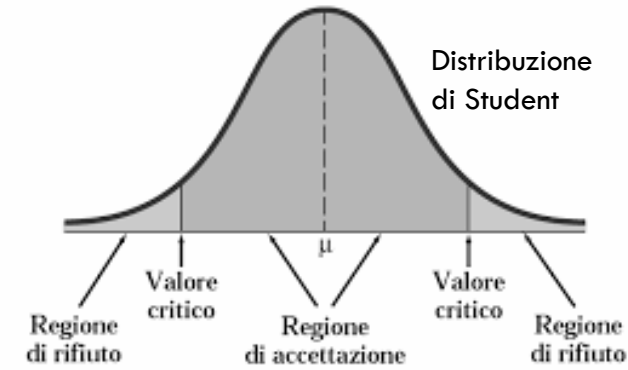
4. Determinazione della distribuzione della statistica

5. Definizione della regola di decisione



Fissato un **livello di significatività α**
(es. $\alpha = 0.05$)

1. Accetto H_0 se t cade all'interno dell'intervallo di confidenza
2. Rifiuto H_0 se t cade fuori dell'intervallo di confidenza (valori poco probabili)

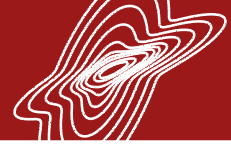


matlab: **trapz**

per il calcolo dell'area sotto la curva

Se **p-value** $< \alpha$ $\rightarrow H_0$ rifiutata

Se **p-value** $> \alpha$ $\rightarrow H_0$ accettata



N.B. il test di student è implementato in matlab a seconda dei casi di studio:

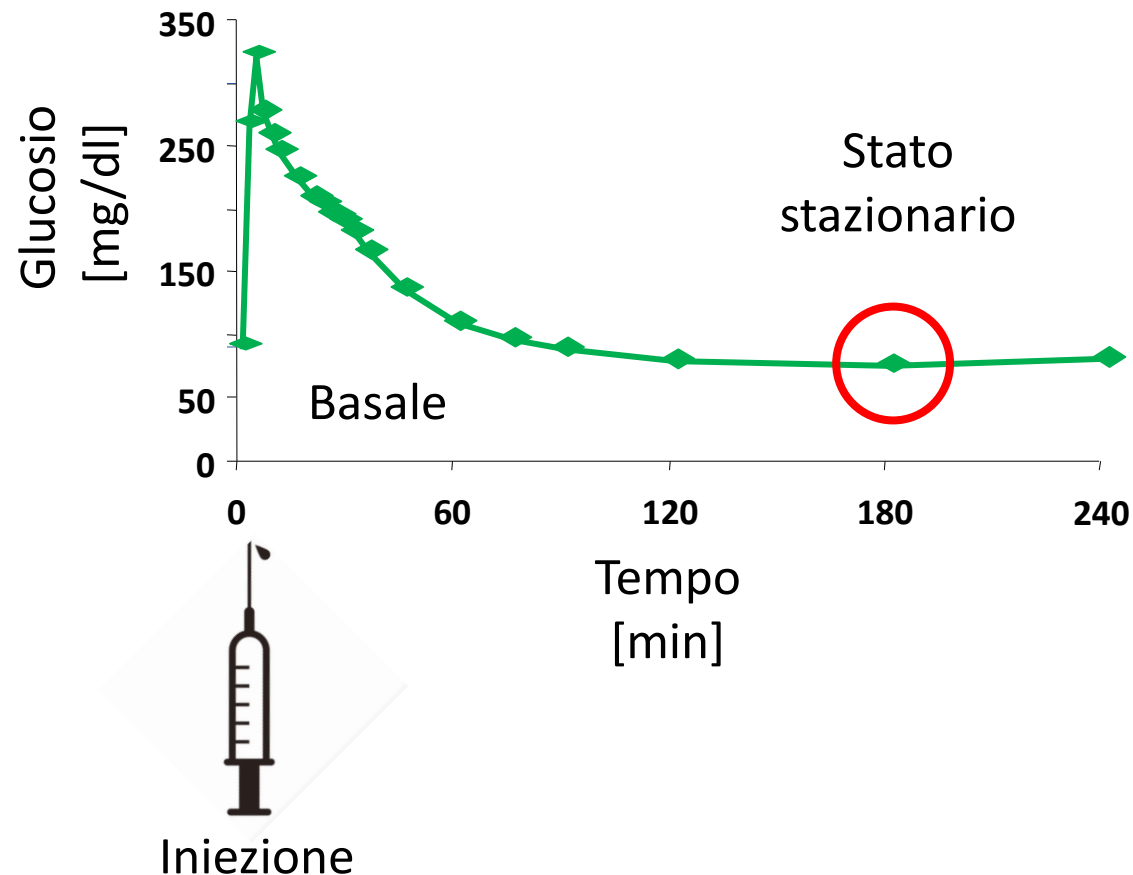
Funzione MATLAB	Tipo	H0
ttest(x)	One-sample t-test	Dati in x appartengono ad una distribuzione normale, con media ZERO e varianza non nota
ttest(x,y)	Paired-sample t-test	Dati in x-y appartengono ad una distribuzione gaussiana con media ZERO e varianze non nota
ttest2(x,y)	Two-sample t-test	Dati in x e y appartengono ad una distribuzione normale con media equivalente e uguale varianza ma non nota

N.B. ttest(x,y) vs ttest2(x,y) nel primo caso abbiamo due misure ripetute (o appaiate, paired) dello stesso campione, mentre nel secondo caso abbiamo due campioni diversi



ESERCIZIO – Prima parte (punti da A a C) da effettuare a casa (codice a disposizione)

I dati a disposizione per il laboratorio sono dati provenienti da un centro americano di un test per la determinazione della sensibilità insulinica: **IntraVenous Glucose Tolerance Test (IVGTT)**



Abbiamo a disposizione i dati di concentrazione di Glucosio corrispondente a un prelievo **allo stato stazionario**

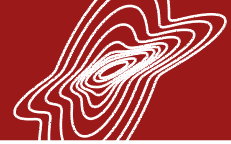
Per due popolazioni:

- Giovani
- Anziani

I dati sono contenuti nel file

`data_IVGTT_SS.mat` sotto forma di matrice double 2D e con le etichette corrispondenti in forma di array di celle.

OBIETTIVO: implementare il test di **student** per verificare se il valore di glicemia **differisce statisticamente** tra le due popolazioni perchè influenzato dall'età.



- **DATI**

Cosa trovate nel file `data_IVGTT_SS.mat` nella cartella `data`:

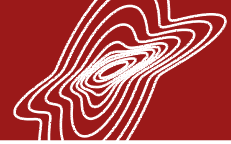
data: matrice 2D double di dimensioni 204x2

Ogni riga corrisponde ad un diverso soggetto

Ogni colonna corrisponde ad una diverso gruppo (e.g., young, elderly)

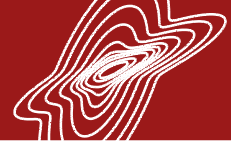
labels: array di celle di dimensioni 1x2

Ogni elemento corrisponde ad una diversa etichetta dei gruppi di soggetti a disposizione.



A. Visualizzazione dei dati:

- Caricare in due **vettori colonna** (nomi di esempio: **Gss_elderly** e **Gss_young**) i dati relativi alle due variabili.
- Eliminare eventuali valori **negativi**, **NaN** o **inf** perché non fisiologici (**any**, **isnan**, **isinf**).
- Calcolare **media**, **standard deviation**, **mediana** e **moda** (**mean**, **median**, **std**, **mode**) delle distribuzioni delle due variabili e visualizzare i valori nella command window (**disp**).



B. Verifica dell'assunzione di gaussianità per le variabili oggetto del test di Student:

1. Controllare la distribuzione delle due variabili con i **boxplot** (**boxplot**) ed **istogrammi con la normalizzazione «pdf»** (**histogram**) (fare in una stessa finestra grafica (**subplot**) gli istogrammi delle due variabili affiancati)
2. Utilizzare **ksdensity** per stimare la **funzione di densità di probabilità** (ddp) della variabile **Gss_elderly**
3. Utilizzando la funzione **normpdf**, calcolare la ddp gaussiana con media e deviazione standard pari alla media e deviazione standard campionarie di **Gss_elderly**
4. Sovrapporre all'istogramma (comando **hold on**) la ddp stimata e la gaussiana calcolate ai punti precedenti. Le due distribuzioni sono uguali?
5. Ripetere i punti 2-4 per la variabile **Gss_young**
6. Calcolare **skewness** (**skewness**) e **kurtosis** (**kurtosis**), e visualizzare i valori nella command window (**disp**)
7. Verificare l'assunzione di gaussianità con il Lilliefors test (**lillietest**)

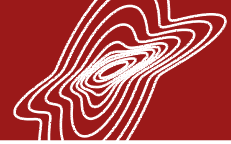


C. Calcolare la statistica t_{obs} utilizzando la seguente formula (si vedano slide del vademecum):

Statistica t:

$$t = \frac{m_1 - m_2}{s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

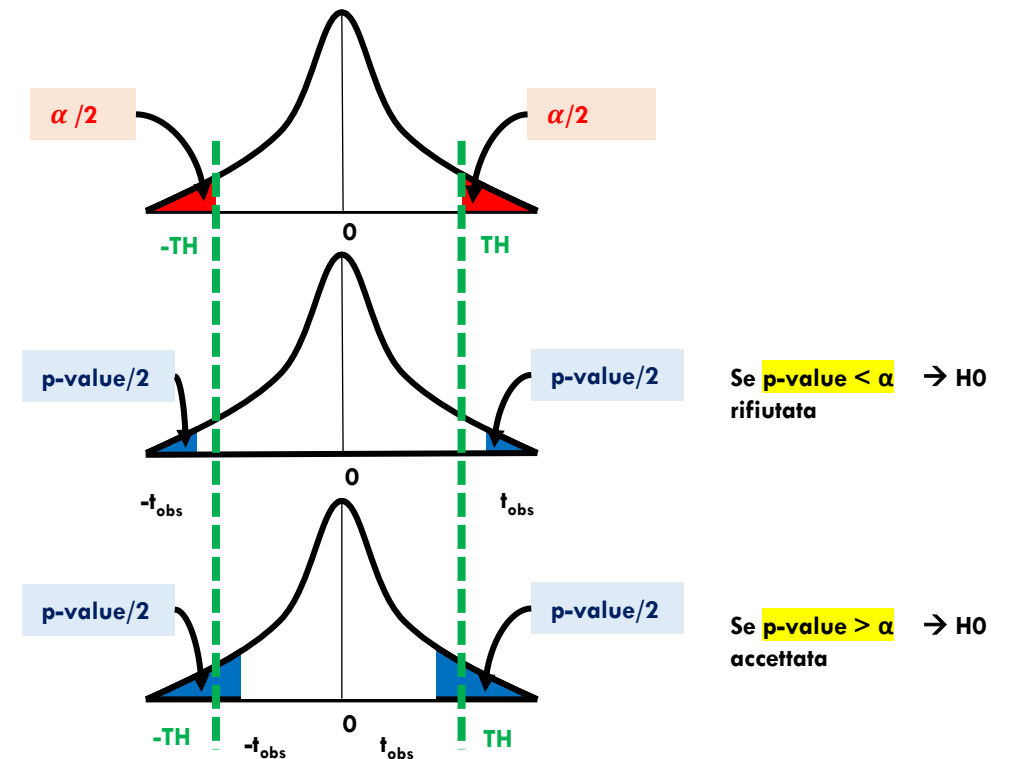
dove $s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$

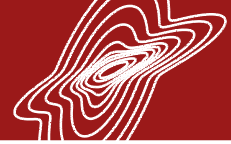


D. Calcolare la **distribuzione di Student** nel range « $x = [-10:0.01:10]$ » usando la funzione (**tpdf**)

E. Calcolare il **pvalue** usando la funzione **tcdf** e verificare se l'ipotesi nulla (H_0) è accettata o rifiutata considerando un livello di significatività $\alpha = 5\%$. (valore atteso: **p-value = 0.2944**, pari a 2 volte il valore dell'integrale)

(BONUS) ricalcolare il p-value usando la funzione **trapz** per approssimare numericamente l'integrale dalla distribuzione di Student





- F. Calcolare il valore critico (TH) corrispondente ad α significativo: testare in un ciclo for i valori del range x a partire dal valore 1 e trovare il valore per cui l'area delle due code è immediatamente inferiore ad α .

- G. Confrontando la t_{obs} con il valore critico (TH), verificare se l'ipotesi nulla (H_0) è accettata. (BONUS: visualizzare i valori di t_{obs} e TH nel plot della distribuzione di Student)

- H. Confrontare i risultati calcolati con le formule inserite manualmente con i parametri `h`, `p`, `stats` restituiti dalla function matlab `ttest2`.