Test statistici

In qualsiasi studio clinico è prevista la raccolta di dati, che vengono utilizzati per descrivere la popolazione in studio e per rispondere a uno o più quesiti di ricerca. Di solito, **l'obiettivo di uno studio è valutare se esista un'associazione fra alcune caratteristiche o se intervenendo**, ad esempio con un trattamento, sia possibile modificare specifici parametri o modificare la storia naturale di una patologia.

Ad esempio, si potrebbero raccogliere dati per indagare se esiste una correlazione fra livello di obesità e livello di controllo metabolico in una popolazione di soggetti affetti da diabete tipo 2, o per valutare se un intervento educativo rivolto a promuovere stili di vita più salutari sia in grado di ridurre i livelli di emoglobina glicosilata in soggetti con diabete tipo 1.

Per rispondere a questi quesiti, utilizziamo i test statistici, che ci permetteranno di accettare o di rifiutare (confutare) una ipotesi.

- **Fasi della verifica di ipotesi statistica**
  - formulazione del **sistema di ipotesi**: è costituito dall'ipotesi da verificare, detta ipotesi nulla (Ho), e dall'ipotesi alternativa (H1), generalmente la negazione logica della prima
  - scegliere la **statistica test**: una quantità calcolata sui dati osservati, che sintetizza l'informazione portata dal campione ai fini dell'inferenza
  - esplicitare le **assunzioni**: ipotesi ausiliarie che non vengono sottoposte a verifica, ma si rendono necessarie per lo sviluppo formale del metodo: di solito riguardano la distribuzione della variabile dipendente (ovvero del processo di misura che genera i dati osservati)
  - determinare la **distribuzione campionaria** della statistica test: immaginando di ripetere il test infinite volte (principio del campionamento ripetuto), la statistica test assumerà valori diversi, descrivendo una propria distribuzione
  - prefissare il **livello di significatività** del test: il test statistico può portare a rifiutare una ipotesi vera, ma questo deve avvenire "raramente"; il livello di significatività stabilisce con quale probabilità il test potrà condurre ad una decisione sbagliata
  - ⇨ determinare la **regione di rifiuto** per l'ipotesi Ho: è il punto di arrivo del metodo, ci permette di decidere se accettare o rifiutare l'ipotesi

- L'ipotesi da verificare (o meglio da falsificare) viene detta **Ipotesi Nulla** e indicata con **Ho**, mentre l'ipotesi alternativa viene indicata con H1
  - Ho viene detta "ipotesi *nulla*" perché si preferisce formulare come Ho l'ipotesi che descrive una situazione di riferimento, o che rappresenta un valore base, rispetto alla quale evidenziare una differenza o un effetto
  - Ho deve essere una ipotesi *puntuale* (cioè una affermazione ben precisa, un valore determinato): si tratta di un requisito per lo sviluppo formale del metodo
  - mentre H1 può essere una ipotesi *complessa* (un insieme di valori alternativi)

Formulazione dell'ipotesi nulla e dell'ipotesi alternativa:

Ipotesi nulla (H0): Il risultato è ottenuto per effetto del caso

Ipotesi alternativa (H1): Il risultato è ottenuto a causa della relazione esistente tra la variabile indipendente e la variabile dipendente (o della relazione esistente tra le diverse variabili considerate nel disegno sperimentale)

- **Il Sistema di Ipotesi**
- Ho e H1 costituiscono il **sistema di ipotesi**
- Il sistema di ipotesi si dice di tipo **bilaterale** (o bidirezionale, o a due code) quando H1 è una ipotesi complessa (cioè descrive più valori) e comprende sia i valori minori che quelli maggiori rispetto al valore puntuale previsto da Ho:

$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu \neq \mu_o \end{cases}$$

dove μ_o è un valore determinato (es. 74 kg)

- Il sistema di ipotesi è invece **unilaterale** (o unidirezionale, o ad una coda) nei seguenti casi:

$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu > \mu_o \end{cases} \quad \text{oppure} \quad \begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu < \mu_o \end{cases}$$

- Esempio
  Vogliamo verificare se c'è una differenza di altezza (o qualunque altra variabile quantitativa) tra due popolazioni. Come possiamo formulare l'ipotesi ?
- Le due popolazioni, A e B, avranno ovviamente due diverse distribuzioni per la variabile altezza:
  - diciamo di accontentarci di confrontare le medie delle due distribuzioni, supponendo che presentino uguali variabilità e andamento cioè forma (o molto simili), ad esempio Normale
    (queste sono *assunzioni* : in particolare la normalità della distribuzione, che possiamo ritenere in questo caso plausibile, per la variabile altezza)
  - allora abbiamo individuato come formalizzare l'ipotesi in termini statistici: può essere espressa come confronto tra le medie delle due distribuzioni
- Ricordiamo che Ho deve essere una ipotesi puntuale, cioè una affermazione secca sulla situazione prevista e non un range di possibilità: per questo dobbiamo formulare come ipotesi nulla Ho che le due medie siano uguali,
  e come H_1 che siano diverse:

$$\begin{cases} H_0 : \ \mu_A = \mu_B \\ H_1 : \ \mu_A \neq \mu_B \end{cases}$$

Test Statistici su 2 gruppi

- Definizione di ipotesi nulla H_0: i due gruppi sono uguali

- Definizione di ipotesi alternativa H_1: i due gruppi sono diversi

Test Statistici su 2 gruppi

Il risultato di un test di ipotesi è una decisione: rifiuto H0 (decido che c'è differenza tra i due gruppi) o la assumo vera (decido che non c'è differenza tra i due gruppi)

DECISIONE

REALTA'	**H0 vera** Gruppo 1 = Gruppo 2	**H0 rifiutata** Gruppo 1 ≠ Gruppo 2
H0 vera Gruppo 1 = Gruppo 2	VERI	FALSI POSITIVI
H0 falsa Gruppo 1 ≠ Gruppo 2	FALSI NEGATIVI	VERI

Quando prendo una decisione, ci sono 4 possibili situazioni in cui mi posso trovare

DECISIONE

REALTA'	**H0 vera** Gruppo 1 = Gruppo 2	**H0 rifiutata** Gruppo 1 ≠ Gruppo 2
H0 vera Gruppo 1 = Gruppo 2	OK	**FP** Errore di tipo I
H0 falsa Gruppo 1 ≠ Gruppo 2	**FN** Errore di tipo II	OK

- La probabilità che un test mi faccia fare un errore di tipo FP è detta FP rate (errore di tipo I)
- La probabilità che un test mi faccia fare un errore di tipo FN è detta FN rate (errore di tipo II)

Un errore di tipo I è il rifiuto di un'ipotesi nulla vera

Un errore di tipo II è il non- rifiuto di un'ipotesi falsa nulla

- Cerco un test che minimizzi gli errori
- In genere scelgo il test che minimizza il FN rate

- Gli statistici ci hanno messo a disposizione tutta una serie di test con le relative ipotesi, in cui il test costituisce la scelta migliore.
- Diverse ipotesi $\Rightarrow$ test diverso $\Rightarrow$ risultato diverso
- Bisogna fare attenzione alla veridicità delle ipotesi che si assumono

- **La Statistica Test (o Test statistico)**
- La **statistica test** (o semplicemente **Test)** è una quantità che viene calcolata a partire dai dati osservati, in grado di riassumere l'informazione campionaria rilevante ai fini dell'inferenza, cioè della valutazione della verosimiglianza dell'ipotesi
- La statistica test da utilizzare varia a seconda del problema, cioè del sistema di ipotesi (e delle assunzioni ausiliarie): la scelta è in realtà più semplice di quanto si possa temere, almeno nei problemi standard che si incontrano più frequentemente
- I principali problemi di verifica di ipotesi hanno infatti una soluzione nota, già sviluppata, ovvero un Test pronto all'uso: il lavoro diventa quello di cercare di ricondurre il problema reale ad una di queste situazioni standard
- Esempio:

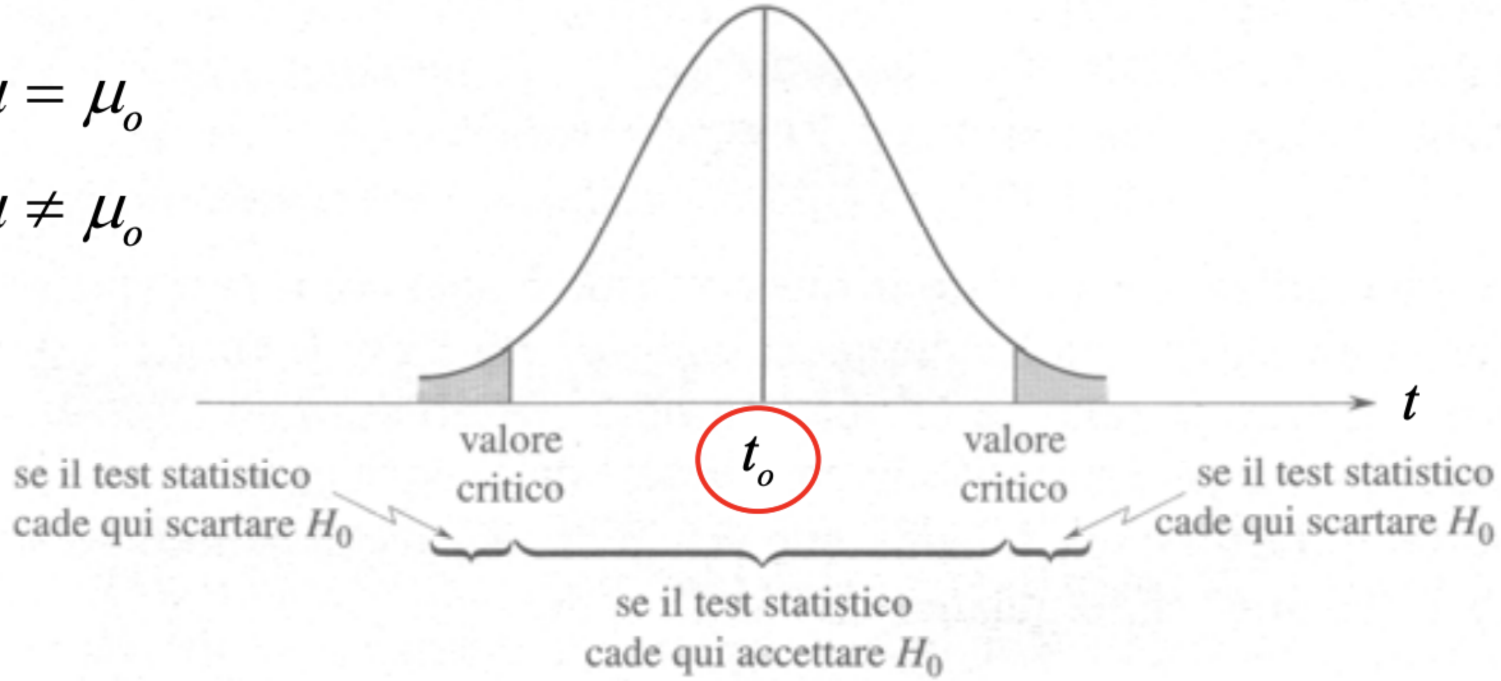
$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu \neq \mu_o \end{cases} \rightarrow \boxed{t = \frac{\bar{x} - \mu_o}{s / \sqrt{n}}}$$

test T di Student

- La statistica test osservata (a posteriori) non è altro che un numero, calcolato sui dati campionari: esattamente come la media o la deviazione standard campionarie (spesso si basa proprio su tali statistiche campionarie)

- **La Regione di rifiuto**
- La **Regione di rifiuto** è l'insieme dei valori che la statistica test *non dovrebbe* assumere, se è vera l'ipotesi nulla, se non per effetto del caso e con una probabilità molto bassa

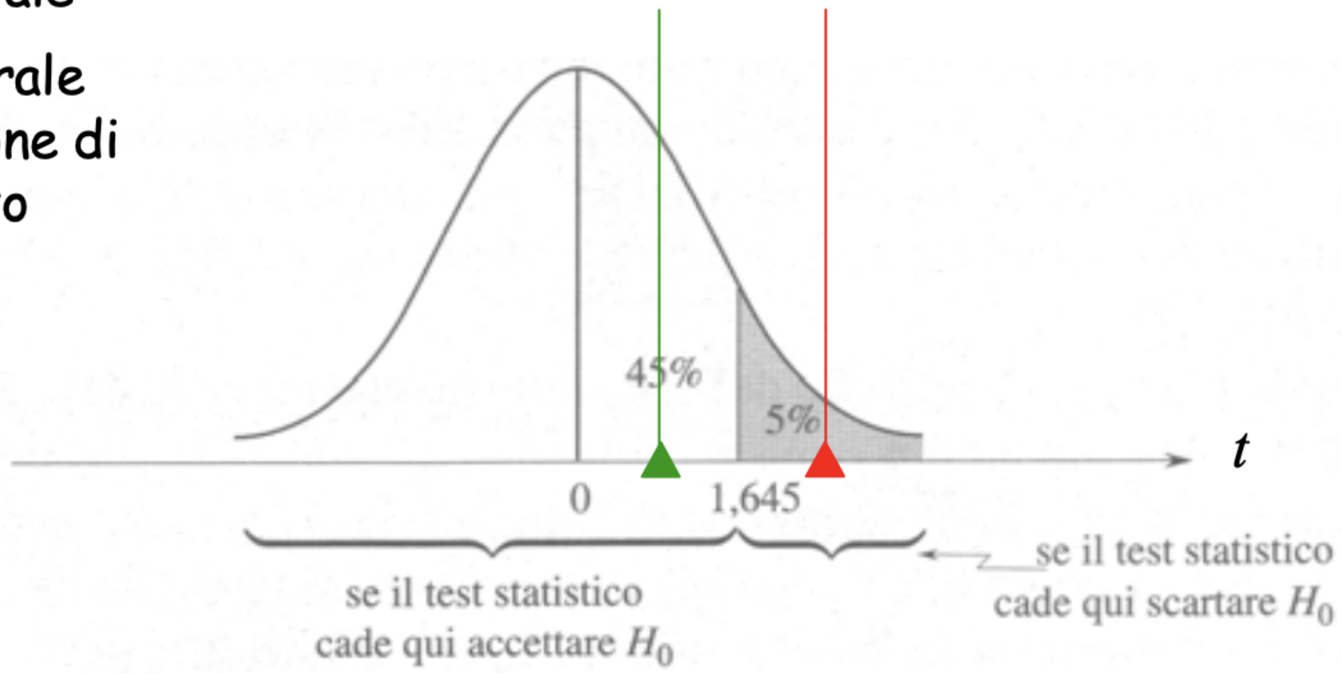
$$\begin{cases} H_0 : \mu = \mu_o \\ H_1 : \mu \neq \mu_o \end{cases}$$



- Se il valore assunto dalla statistica test cade nella regione di rifiuto, il risultato campionario risulta **significativamente** lontano dall'atteso
- Quando il valore osservato del Test cade nella regione di rifiuto, questo conduce al rifiuto dell'ipotesi nulla, perché si valuta il risultato empirico troppo lontano e quindi in disaccordo con quanto previsto dall'ipotesi

- Nella maggior parte delle applicazioni pratiche, la regione di rifiuto consisterà in un intervallo, o nell'unione di due intervalli, a seconda che il test sia unilaterale o invece bilaterale
- Nel caso di un test unilaterale (es. coda a destra) la regione di rifiuto sarà tutta da un lato

$$\begin{cases} H_0 : \mu = \mu_o \\ H_1 : \mu > \mu_o \end{cases}$$



- Arrivare a determinare la regione di rifiuto significa anche avere la regola di accettazione/rifiuto dell'ipotesi Ho: tutto il metodo consiste effettivamente nella determinazione della regione di rifiuto per il Test
- La determinazione della regione di rifiuto richiede che si conosca la *distribuzione* della statistica Test

NOTA 1:

GRADI DI LIBERTA’: per molti test statistici il numero di gradi di libertà è uguale al numero di dati meno 1.

Esempio: se abbiamo a disposizione in data set di 20 campioni con una certa media, allora i gradi di libertà sono 19 poiché conoscendo 19 campioni e il valore della media posso ricavare il 20° campione

Test parametrici e test non parametrici

Test parametrico ➔ basato sui parametri media e deviazione standard, da usare solo nel caso in cui la variabile di interesse sia continua e normalmente distribuita.

In tutti gli altri casi sono da preferire i test non parametrici. Tali test si basano sui ranghi delle osservazioni, non sul loro reale valore. In altre parole, le osservazioni vengono messe in ordine crescente, e a ognuna si attribuisce un numero corrispondente alla posizione che quell'osservazione occupa nella graduatoria (rango). I test statistici non parametrici vengono quindi basati sul confronto fra le somme dei ranghi.

Test parametrici con il test t di Student e l’ANOVA sono basati su alcune assunzioni…

1. Variabili continue o almeno misurate in un intervallo (es. non conosco il valore assoluto, ma posso quantificare le differenze fra due valori)

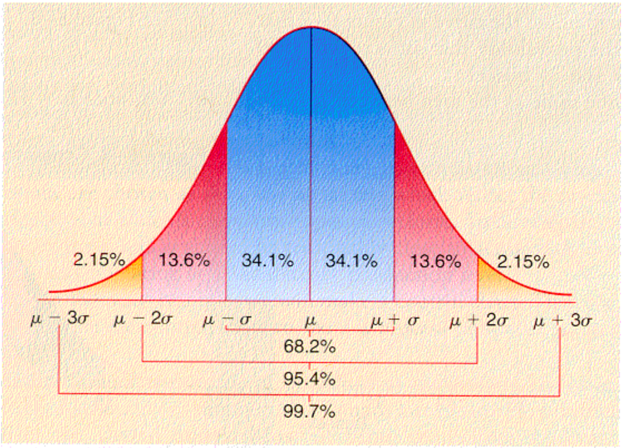
2. Indipendenza fra media e varianza (l’errore di misura deve essere indipendente dal valore misurato)

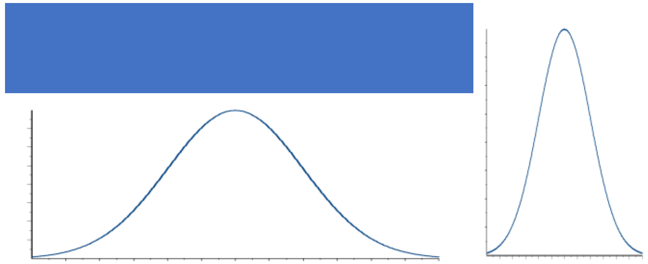
3. Variabili distribuite in modo (approssimativamente) normale

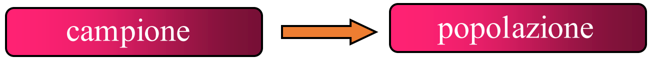
4. Omogeneità delle varianze

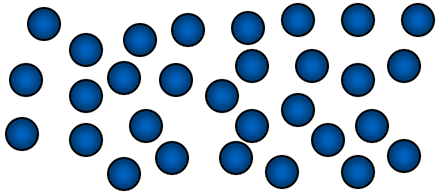
5. I risultati ottenuti con l’analisi di campioni si applicano alle popolazioni

6. Dimensione campione > 10 (meglio se ≥30)



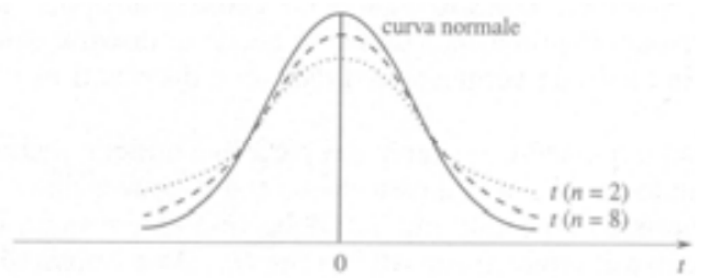






- **La distribuzione campionaria della statistica Test**
- A posteriori, dato un campione, la statistica test *osservata* è un numero.
  A priori, se immaginiamo di ripetere infinite volte il campionamento, la statistica test assume valori sempre diversi, descrivendo una distribuzione tipica di quel test
- Ogni test ha cioè una propria distribuzione specifica, con una forma caratteristica (tanto che in molti casi prende il nome dal test stesso), che deve essere determinata per poter procedere alla determinazione della regione di rifiuto
- La statistica test è una trasformazione delle n variabili indipendenti X(i) che descrivono le singole osservazioni: quindi la sua distribuzione dipende da quella delle variabili elementari X(i) di cui si compone, e dal loro numero (n), oltre ovviamente che dalla sua espressione analitica
- Esempio: il test T di Student

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



- Ricavare la distribuzione della statistica test è un problema matematico che richiede nozioni avanzate di calcolo delle probabilità
- Per i problemi standard, questo lavoro è *già* stato fatto e quindi conosciamo le distribuzioni dei test che si usano più frequentemente; oltre alla Normale, le distribuzioni che ricorrono maggiormente sono:
  T di Student, F di Snedecor, Chi-Quadrato

- **Test sulla media di una popolazione**
- Uno dei problemi più semplici è quello dell'ipotesi sulla media di una popolazione, che si può presentare con due varianti:
  - la varianza della popolazione è nota
  - la varianza della popolazione è ignota
- **Primo caso: Varianza nota**
- Quando la varianza è nota, per testare una ipotesi sulla media si usa il test z :

$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu \neq \mu_o \end{cases} \rightarrow \boxed{z = \frac{\bar{x} - \mu_o}{\sigma / \sqrt{n}}} \sim N(0,1) \,|\, H_o \text{ è vera}$$

- che si distribuisce normalmente, quando Ho è vera, infatti:

$$x_i \sim N(\mu, \sigma) \quad i.i.d. \, \forall i \quad \Rightarrow \quad \bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

$$\text{Quando } H_0 \text{ è vera}: \ \mu = \mu_o \quad \text{e se inoltre } \sigma \text{ è nota}: \quad z = \frac{\bar{x} - \mu_o}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$x_i \sim N(\bar{x}, \sigma^2) \quad \Longrightarrow \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \Longrightarrow \quad Var[\bar{x}] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$\boxed{\text{Se } \bar{x} = \mu_0}$$

$$E[z] = E\left[\frac{\bar{x}}{\sigma/\sqrt{n}} - \frac{\mu_0}{\sigma/\sqrt{n}}\right] = E\left[\frac{0}{\sigma\frac{1}{\sqrt{n}}}\right] = 0$$

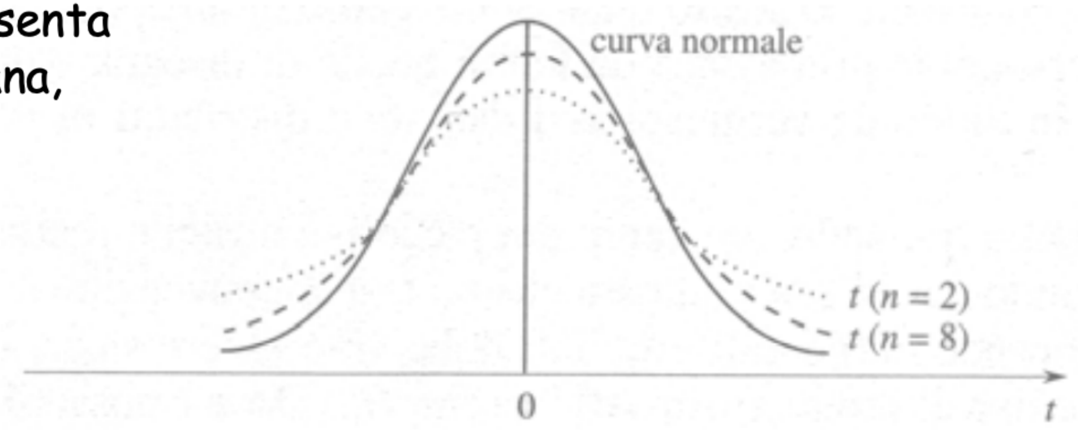
$$Var[z] = Var\left[\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right] = Var\left[\frac{\bar{x}}{\sigma/\sqrt{n}}\right] = \frac{n}{\sigma^2} Var[\bar{x}] = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1$$

$$\hat{\sigma}^2 = s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

- **Secondo caso: Varianza ignota**
- Quando la varianza è ignota, si perviene al **test t di Student** :

$$\begin{cases} H_0 : \mu = \mu_o \\ H_1 : \mu \neq \mu_o \end{cases} \rightarrow \boxed{t = \frac{\bar{x} - \mu_o}{s / \sqrt{n}}} \sim t_{n-1} \mid H_o \text{ è vera}$$

- La statistica test t di Student presenta una distribuzione a forma di campana, simile alla distribuzione Normale



- **La Distribuzione t di Student**
- La forma della curva t di Student è caratterizzata da un unico parametro, detto *gradi di libertà*
- Al crescere del valore di questo parametro (cioè al crescere dei gradi di libertà), la forma della distribuzione t si avvicina sempre più a quella della Normale:
  - per n > 30 si può utilizzare direttamente la tavola della Normale
  - per n <= 30 esistono e si devono usare le tavole specifiche per la t di Student
- Perché valga questo risultato è necessario che siano vere alcune assunzioni: le X(i) devono essere i.i.d. cioè indipendenti e identicamente distribuite in modo normale

VERIFICA DI IPOTESI

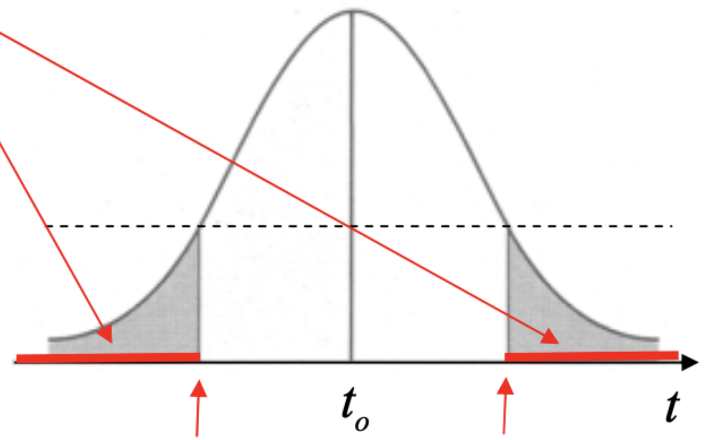
- **Il livello di Significatività del test**
- Un criterio per determinare la regione di rifiuto consiste nel prefissare il rischio (la probabilità) di rifiutare erroneamente l'ipotesi Ho nel caso che sia vera
- La probabilità di commettere tale errore, detto **errore di I tipo**, viene chiamata **livello di significatività *del test*** e indicata con α :

$$\alpha = P\{rifiutare\ H_o \mid H_o\ vera\}$$

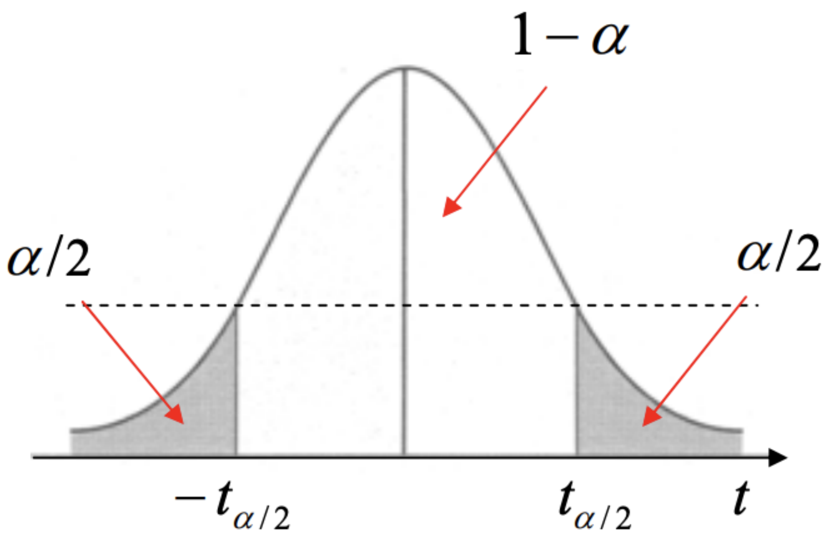
- Se, ad esempio, si fissa α =0,05 significa che si accetta il rischio di sbagliare conclusione, nel senso di rifiutare una ipotesi nulla vera, 5 volte su 100
- Il livello di significatività prefissato determina la "dimensione" della regione di rifiuto (e di quella di accettazione):

$$\alpha = P\{\, t \in [regione\ di\ rifiuto]\,\}$$

- maggiore è il valore di α tollerato e maggiore sarà la dimensione della regione di rifiuto
- al contrario, minore è il valore prefissato di α e maggiore sarà la dimensione della regione di accettazione: quindi il test risulterà più *conservativo* nei confronti dell'ipotesi Ho



- Nel caso di test bilaterale, o a due code:



$$\begin{cases} H_0 : \mu = \mu_o \\ H_1 : \mu \neq \mu_o \end{cases}$$

si individua una regione di rifiuto del tipo

$$(-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, +\infty) \ \textit{tale che}$$

$$\alpha = P\{t \leq -t_{\alpha/2} \cup t \geq t_{\alpha/2}\}$$

$$\textit{ovvero una regione di accettazione}: \ (-t_{\alpha/2}, \ t_{\alpha/2}) \ \textit{per la quale}:$$

$$P\{-t_{\alpha/2} < t < t_{\alpha/2}\} = 1-\alpha$$

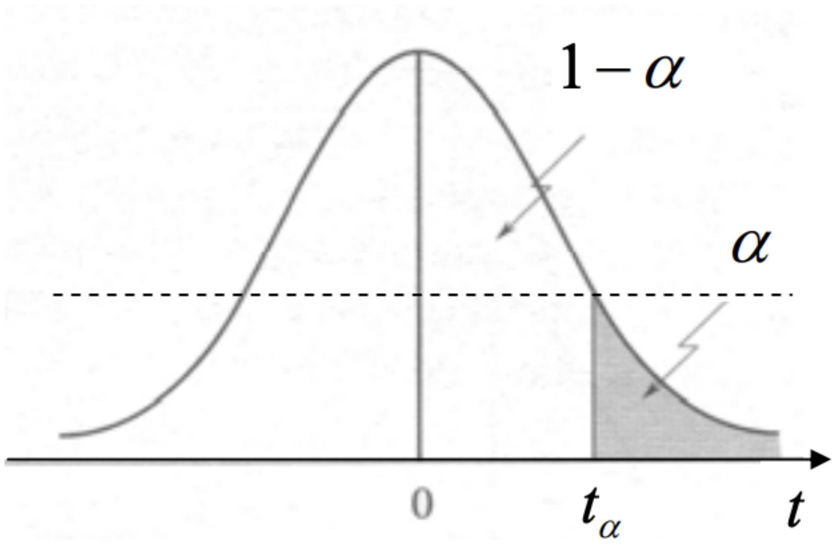
dove $-t_{\alpha/2}$ e $t_{\alpha/2}$ sono i valori critici a sinistra e a destra della statistica test ottenuti in corrispondenza del livello di significatività α prefissato, equiripartito (α/2) sulle due code della distribuzione della statistica test

- Se il valore osservato della statistica test cade in tale regione

  => l'ipotesi nulla viene rifiutata

- Nel caso di test unilaterale, o ad una coda:

$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu > \mu_o \end{cases}$$



si individua una regione di rifiuto del tipo

$$[t_\alpha \,, +\infty) \ \ tale\ che \ \ P\{ \ t \geq t_\alpha \} = \alpha$$

$$ovvero\ una\ regione\ di\ accettazione\ per\ cui\ sia \quad P\{ t < t_\alpha \} = 1-\alpha$$

t_α è il valore critico della statistica test, ottenuto in corrispondenza del livello di significatività α prefissato

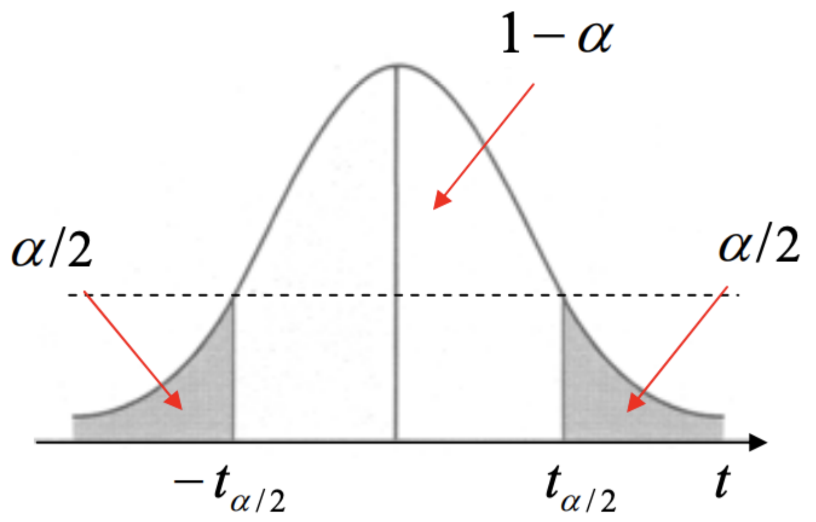
- Se il valore campionario osservato di t è maggiore del valore critico, l'ipotesi nulla viene rifiutata, altrimenti viene accettata

- Test bilaterale (a due code):

$$\alpha = P\{ t \le -t_{\alpha/2} , t \ge t_{\alpha/2} \}$$

$$P\{ -t_{\alpha/2} < t < t_{\alpha/2} \} = 1-\alpha$$

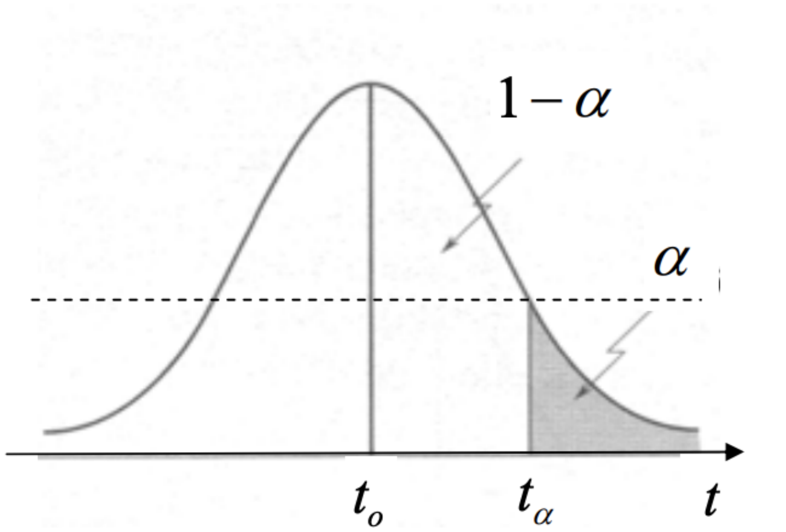
l'area α viene divisa a metà e così si determinano i due valori critici per il test: $-t_{\alpha/2}$ e $t_{\alpha/2}$



- Nel caso di test unilaterale (a una coda) l'area di dimensione α sarà tutta da una parte (es. a destra):

$$\alpha = P\{ t \ge t_{\alpha} \}$$

$$P\{ t < t_{\alpha} \} = 1-\alpha$$



- Comunemente, i valori utilizzati per α sono 0,05 o 0,01 : 0,01 è più conservativo nei confronti di Ho, che viene rifiutata solo di fronte ad un risultato empirico più netto
- Queste soglie sono arbitrarie e puramente orientative: i numeri indicati non hanno altra proprietà che quella di essere numeri tondi, ma sono quelli universalmente utilizzati

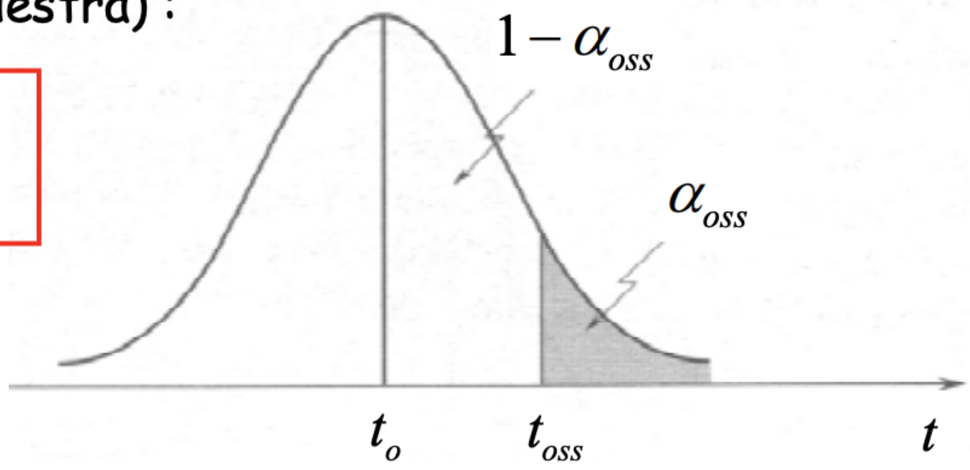
- **Significatività Osservata (o p-value)**
- Il livello di significatività **osservato** (α **osservato**) è la probabilità di commettere un errore di I tipo (rifiutare Ho quando è vera) in base al risultato campionario osservato:

$$\alpha_{oss} = P\{rifiutare\ H_o \mid risultato\ campionario\ osservato\}$$

- In pratica, l' α osservato è la probabilità che la statistica test possa produrre un valore ancora più lontano dall'atteso di quello osservato, nel caso che sia vera Ho. Ad es. nel caso di un test ad una coda (a destra) :

$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu > \mu_o \end{cases}$$

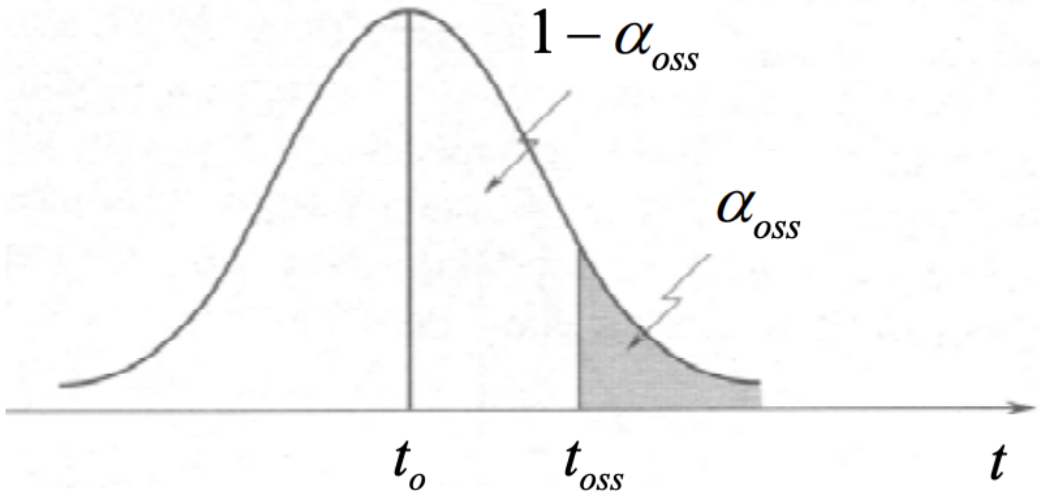
$$\boxed{\alpha_{oss} = P\{\, t \geq t_{oss} \,\}}$$



- L' α osservato è una misura della *verosimiglianza* dell'ipotesi nulla, in base al risultato osservato nel campione:
  - maggiore è l' α osservato, e più l'ipotesi nulla è verosimile e quindi risulta "confermata" dalle osservazioni, cioè compatibile con il risultato empirico
  - viceversa minore è l' α osservato, e più è improbabile che Ho sia vera: il risultato osservato risulta troppo (si dice *significativamente*) diverso dall'atteso, cioè da quanto previsto da Ho

- In pratica, se $\alpha_{oss} < \alpha_{prefissato}$

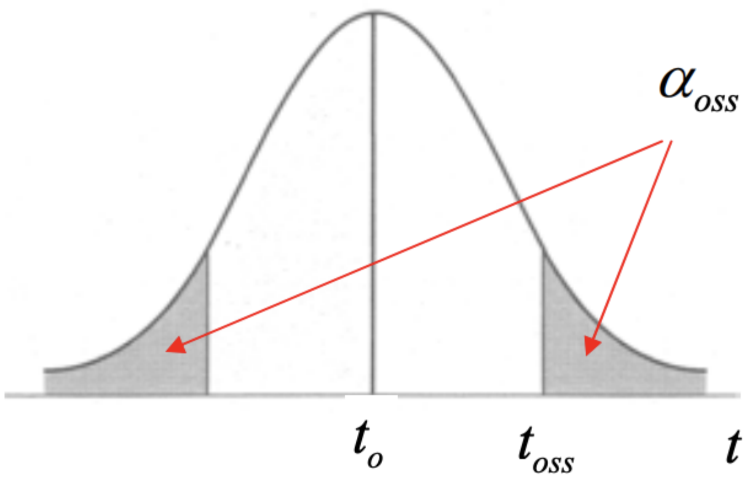
  => si rifiuta Ho



- L' α osservato permette di trarre immediatamente la conclusione sull'accettazione o il rifiuto dell'ipotesi, indicandoci anche *quanto* l'ipotesi è verosimile (e quindi confermata) sulla base dei dati osservati
- Per questa ragione tutti i programmi software utilizzano in realtà questo secondo metodo, invece di calcolare i valori critici per il test, e forniscono quindi l' α osservato, spesso indicato come **p-value**
- Esempio
  Se risulta α osservato = 0,045 che conclusione traiamo sull'ipotesi ?
  Al 5% di significatività, Ho viene rifiutata: ma la sua verosimiglianza è molto vicina alla soglia del 5% (e ancora compatibile con la soglia inferiore del 1%): quindi capiamo che avremmo bisogno di una maggiore evidenza empirica (un campione più numeroso) per essere più sicuri

- Nel caso di test bilaterale (a due code):

$$\begin{cases} H_0 : \ \mu = \mu_o \\ H_1 : \ \mu \neq \mu_o \end{cases}$$



$$\alpha_{oss} = P\{ t \leq -t_{oss} \ , \ t \geq t_{oss} \}$$

se la distribuzione del test è simmetrica :

$$\alpha_{oss} = 2\, P\{ t \geq t_{oss} \}$$

- La regola di accettazione/rifiuto con il metodo del p-value resta sempre la stessa:

$$\alpha_{oss} < \alpha_{prefissato}$$

dunque se risulta p-value < 0,05 => si rifiuta Ho

- Osservazione: il rifiuto di Ho, che rappresenta generalmente l'ipotesi di riferimento ovvero di "indifferenza", significa ritenere *significativa* una differenza tra gruppi, ovvero l'effetto di un fattore studiato sulla risposta osservata.

- ***Accettazione/rifiuto dell'ipotesi***
- Ricapitolando, dopo aver scelto il test appropriato al problema, esistono dunque due metodi equivalenti di procedere
- Primo metodo: classico (o dei valori critici del test)
  - prefissare il valore arbitrario di α (es. 0,05 oppure 0,01), al di sotto del quale si vuole decidere per il rifiuto di Ho
  - determinare di conseguenza, conoscendo la distribuzione della statistica test (sotto Ho), il valore critico (test ad una coda) o i due valori critici (test a due code) che individuano le regioni di accettazione e di rifiuto
  - verificare in quale regione cade il valore osservato della statistica test
- Secondo metodo: **p-value** (α osservato)
  - calcolare il livello di significatività α osservato (p-value)
  - si rifiuta l'ipotesi nulla se l' α osservato è minore di una soglia prefissata
  - con questo metodo abbiamo anche una indicazione di *quanto* l'ipotesi nulla è confermata (o invece falsificata) dai dati osservati: quanto più α osservato è piccolo, tanto più il risultato osservato risulta significativamente diverso da quello atteso

Test statistici

1. Analisi della Varianza

Vogliamo testare l'ipotesi che date due popolazioni (due gruppi), la loro varianza è uguale.

1) l'ipotesi nulla è quindi:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

2) Le ipotesi aggiuntive è che i dati ottenuti campionando le due popolazioni siano realizzazioni di due v.a. Normali con $pop1 \sim N(\mu_1, \sigma_1^2)$ e $pop2 \sim N(\mu_2, \sigma_2^2)$

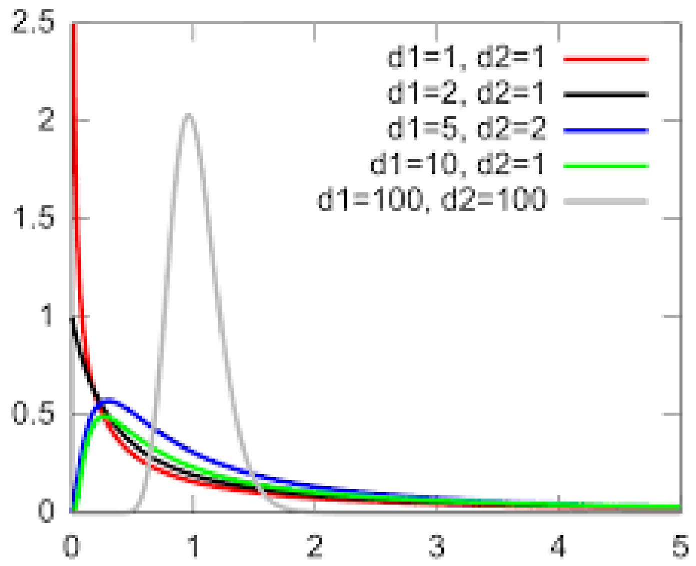
3) Dobbiamo trovare una statistica test la cui distribuzione teorica è nota quando è vera l'ipotesi nulla

$$F = \frac{varianza\ campionaria\ della\ popolazione\ 1}{varianza\ campionaria\ della\ popolazione\ 2}$$

sotto l'ipotesi H0 ossia la varianze sono uguali, F è una v.a. con densità di probabilità F (o di Fisher-Snedecor)

$$F \sim \mathcal{F}(n-1, m-1)$$

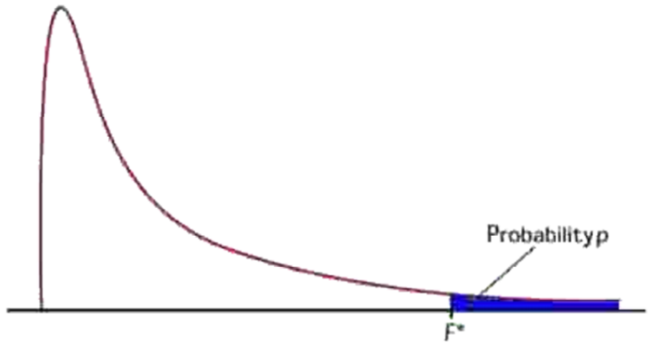
Definita quindi dalla conoscenza del numero di campioni della popolazione 1 (n) e del numero di campioni della popolazione 2 (m)



- E' continua
- Varia tra zero e infinito
- Dipende dai gradi di libertà del numeratore e quelli del denominatore
- E' circa centrata sul valore 1

Quindi:

1. Calcolo la varianza dei due campioni
2. Determino il valore di $F_{calcolato}$
3. Decido il livello di significatività (alpha) ad esempio 0.05
4. Determino il valore di $F_{critico}$
5. 5. Se $F_{calcolato} > F_{critico}$ rifiuto H0



$S_{1,2}$= varianza campionaria

Esempio

- Da due campioni si ottengono i seguenti dati:

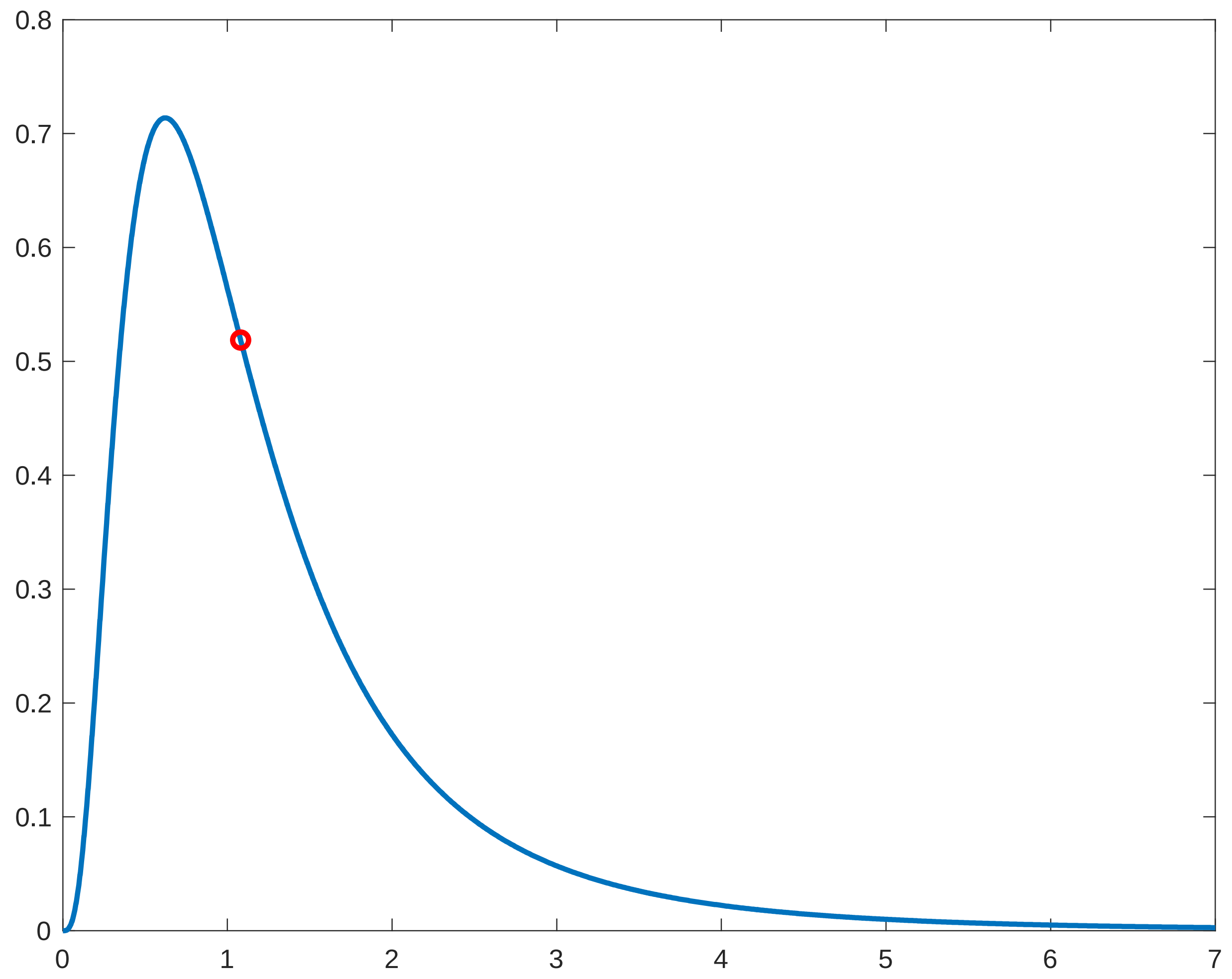
$s_1^2 = 1.5278$; $n_1 = 9$
$s_2^2 = 1.6556$; $n_2 = 10$

- Domanda: le varianze delle popolazioni da cui sono stati estratti i due campioni sono significativamente diverse?

- Il test di questa ipotesi è semplice: definisco la regione di accettazione per la statistica F con l'α prescelto, faccio il rapporto tra le varianze (che diventa l'F_{calc}) e vedo se tale rapporto cade nella regione di accettazione (circa centrata sul valore 1, cioè l'ipotesi nulla che $s_1^2 = s_2^2$ o s_1^2 / s_2^2).

- Praticamente, visto che la distribuzione F è asimmetrica, e le tabelle dei valori critici riportati in tabella si riferiscono al lato destro della distribuzione, conviene sempre mettere a numeratore nel calcolo di F dai dati (F_{calc}) la varianza maggiore

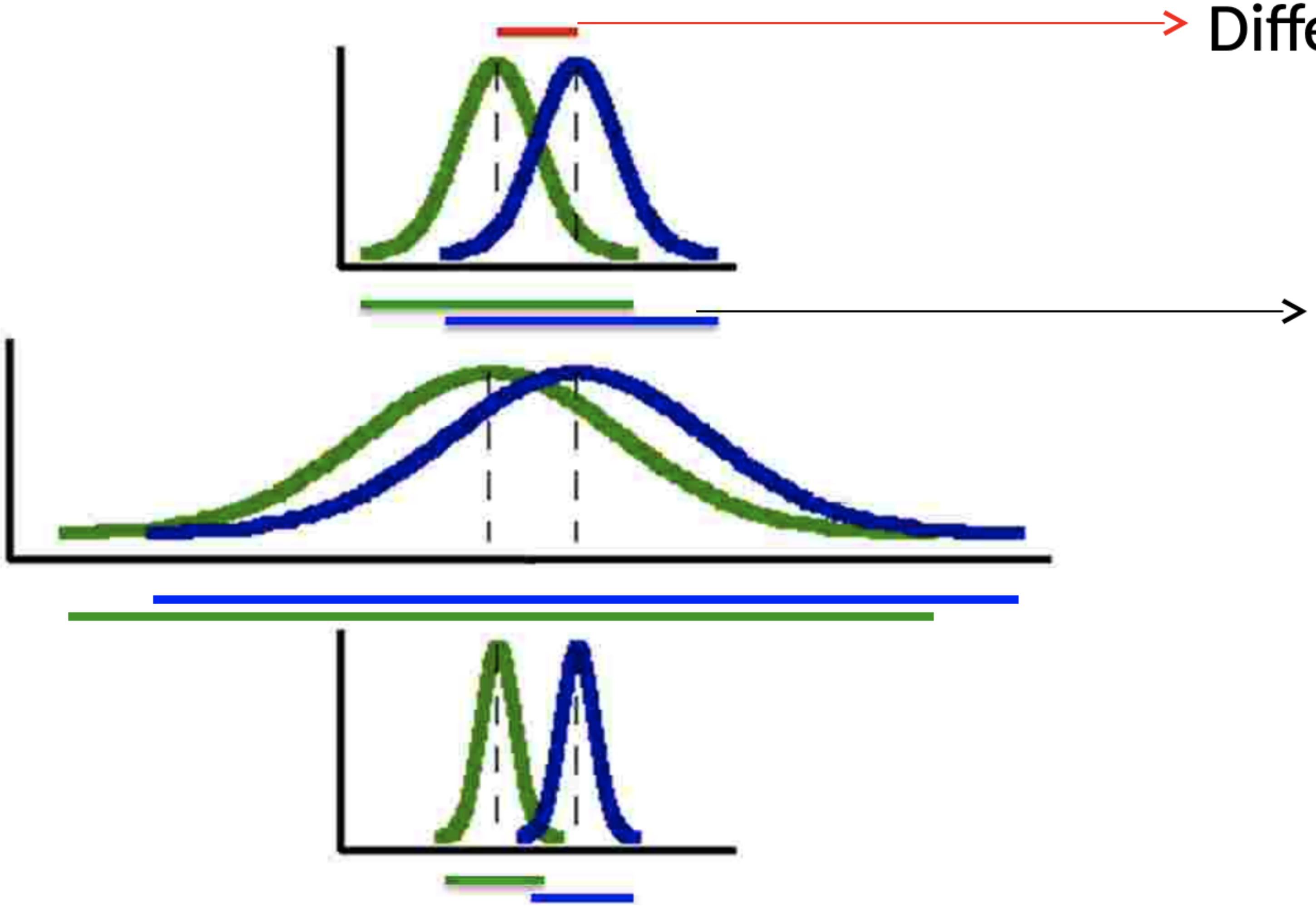
Distribuzione F con i gradi di libertà 9 e 8 con coda a destra. In rosso il valore ottenuto con i dati a disposizione: accetto l’ipotesi nulla



Test statistici

2. T-test

$$t_{calcolato} = \frac{\text{Misura legata alla differenza fra le medie}}{\text{Misura di variabilità dentro i gruppi}}$$



Differenza medie

Variabilità dei gruppi

Come scegliere il test t giusto a partire dalle assunzioni

Indipendenza

NO

Test t appaiato

$$t = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}}$$

SÌ

Test t non appaiati

$$s_2^2 = s_1^2$$

Test t per pop. omoschedastiche

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s_2^2 \neq s_1^2$$

Test t per pop. eteroschedastiche

Welch t-test

Come scegliere il test t giusto a partire dalle assunzioni

Indipendenza

NO

Test t appaiato

$$t = \frac{\overline{D}}{\frac{S_D}{\sqrt{n}}}$$

SÌ

Test t non appaiati

$s_2^2 = s_1^2$

Test t per pop. omoschedastiche

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$s_2^2 \neq s_1^2$

Test t per pop. eteroschedastiche
Welch t-test

Vogliamo testare l'ipotesi che date due popolazioni (due gruppi), la loro media è uguale.

1) l'ipotesi nulla è quindi:

$$H_0\colon \mu_1 = \mu_2$$
$$H_1\colon \mu_1 \neq \mu_2$$

2) Le ipotesi aggiuntive sono che i dati ottenuti campionando le due popolazioni siano realizzazioni di due v.a. Normali con $pop1 \sim N(\mu_1, \sigma_1^2)$ e $pop2 \sim N(\mu_2, \sigma_1^2)$

3). Dobbiamo trovare una statistica test la cui distribuzione teorica è nota quando è vera l'ipotesi nulla

$$t = \frac{\mu_1 - \mu_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

t una v.a. con densità di probabilità t-Student:

$$t \sim T(n_1 + n_2 - 2)$$

Definita quindi dalla conoscenza del numero di campioni della popolazione 1 (n) e del numero di campioni della popolazione 2 (m)

Test su 2 gruppi: t test

Esempio: abbiamo i seguenti due gruppi

	DATASET 1	DATASET 2
1	639	650
2	646	633
3	650	631
4	641	637
5	641	642
6	637	638
7	659	640
8	650	634
9	640	626
10	635	636
11		640

vogliamo sapere se i due gruppi sono uguali ossia vogliamo testare l'ipotesi nulla:

H0 i due gruppi sono uguali

Se è vera l'ipotesi che i due gruppi sono estratti dalla stessa popolazione, le loro varianze sono entrambe stime della varianza di popolazione ipotizzo che i due gruppi siano misure di una variabile che ha distribuzione graussiana e applico il t-test, ossia calcolo:

$$t = \frac{m_i - m_j}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$$

$$s_p^2 = \frac{(n_i - 1)s_i^2 + (n_j - 1)s_j^2}{(n_i + n_j - 2)}$$

$s_i^2, s_j^2 =$ varianza del gruppo i e del gruppo j

$m_i, m_j =$ media del gruppo i e del gruppo j

$n_i, n_j =$ N° di dati del gruppo i e del gruppo j

	DATASET 1	DATASET 2
1	639	650
2	646	633
3	650	631
4	641	637
5	641	642
6	637	638
7	659	640
8	650	634
9	640	626
10	635	636
11		640
Media	**643.8**	**637**
Varianza	**54.4**	**39.6**

$$s_p^2 = \frac{(10-1)\cdot 54.4 + (11-1)\cdot 39.6}{(10+11-2)} = 46.61$$

$$t = \frac{m_i - m_j}{s_p\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}} = \frac{643.8-637}{\sqrt{46.61}\cdot\sqrt{\frac{1}{10}+\frac{1}{11}}} = 2.2796$$

$$df = 10+11-2 = 19$$

$$p\_value = 5\%$$

	Probabilità di valori maggiori, P								
ν	0,50	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	1,000	3,078	6,314	12,706	31,821	63,657	127,321	318,309	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,335	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850

$$t = \frac{m_i - m_j}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} = \frac{643.8 - 637}{\sqrt{46.61} \cdot \sqrt{\frac{1}{10} + \frac{1}{11}}} = 2.2796 > 2.093 \, (\text{soglia})$$

Rifiuto l’ipotesi nulla e quindi affermo che i due gruppi sono significativamente diversi

Come scegliere il test t giusto a partire dalle assunzioni

Indipendenza

NO

Test t appaiato

$$t = \frac{\overline{D}}{\frac{S_D}{\sqrt{n}}}$$

SÌ

Test t non appaiati

$s_2^2 = s_1^2$

Test t per pop. omoschedastiche

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$s_2^2 \neq s_1^2$

Test t per pop. eteroschedastiche
Welch t-test

WELCH t TEST

Vogliamo sempre testare l'ipotesi che date due popolazioni (due gruppi), la loro media è uguale.

1) l'ipotesi nulla è quindi:

$$H_0\colon \mu_1 = \mu_2$$
$$H_1\colon \mu_1 \neq \mu_2$$

2) Le ipotesi aggiuntive sono che i dati ottenuti campionando le due popolazioni siano realizzazioni di due v.a. Normali con $pop1 \sim N(\mu_1, \sigma_1^2)$ e $pop2 \sim N(\mu_2, \sigma_2^2)$

3). Dobbiamo trovare una statistica test la cui distribuzione teorica è nota quando è vera l'ipotesi nulla

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t una v.a. con densità di probabilità t-Student:

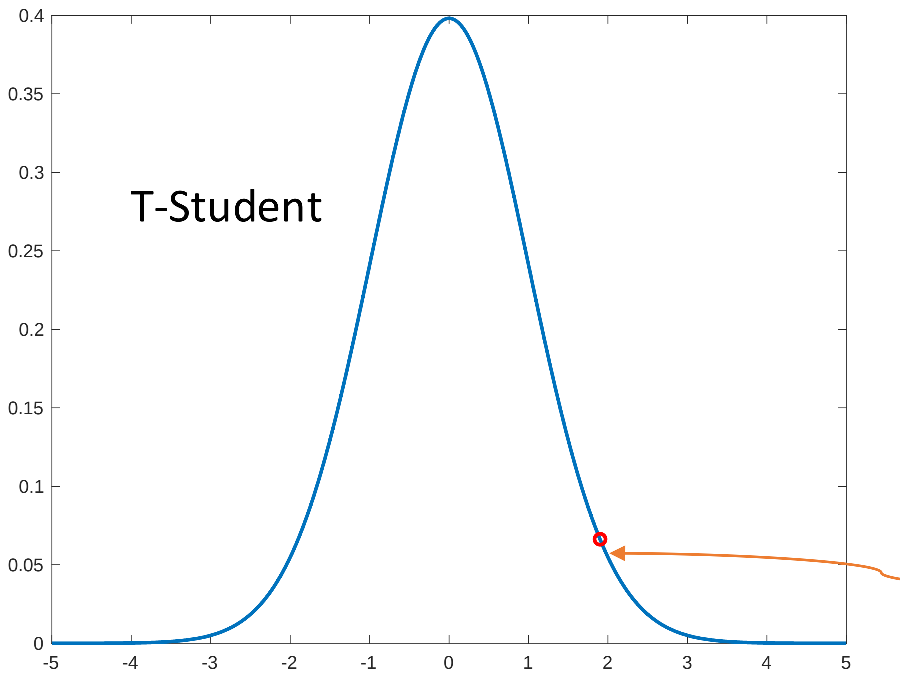
$$t \sim T(\upsilon) \qquad \upsilon \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$$

Welch's *t*-test è più robusto del test t (anche detto t-Student)

Può essere usato anche nel caso di varianze uguali

====================

Usando i dati contenuti nel file Matlab e applicando il test sempre per la variabile age il risultato è:



P_value= 5.99e-02

Statistica t calcolata per il data set

Come scegliere il test t giusto a partire dalle assunzioni

Indipendenza

NO

Test t appaiato

$$t = \frac{\overline{D}}{\frac{S_D}{\sqrt{n}}}$$

SÌ

Test t non appaiati

$$s_2^2 = s_1^2$$

Test t per pop. omoschedastiche

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s_2^2 \neq s_1^2$$

Test t per pop. eteroschedastiche
Welch t-test

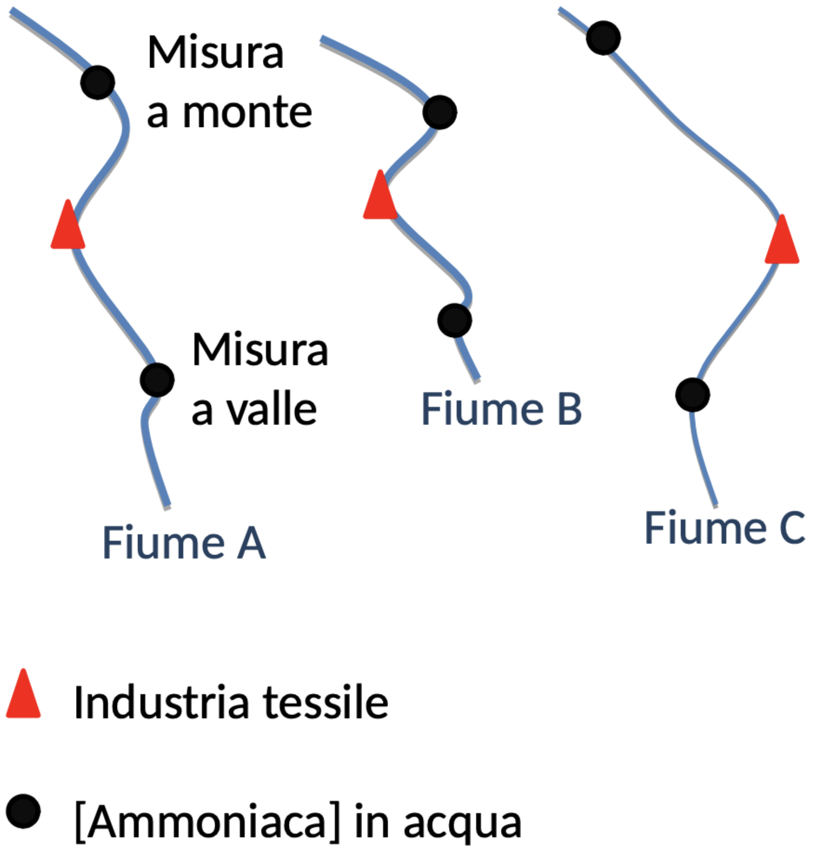
t-TEST per variabili appaiate

Variabili appaiate: quando vogliamo studiare una differenza fra due o più misurazioni della stessa variabile sulla stessa persona ad es. prima e dopo uno stimolo particolare, prima e dopo un intervento educativo, lo stesso test a distanza di tempo.

1. Misure ripetute

Studente	Prima	Dopo
A	22	23
B	23	24
C	24	24
D	25	25
E	20	21
F	18	18
G	18	18
H	19	20

2. Correlazione nello spazio



Vogliamo sempre testare l'ipotesi che date due misure la loro media è uguale.

1) l'ipotesi nulla è sempre:

$$H_0\colon \mu_1 = \mu_2$$
$$H_1\colon \mu_1 \neq \mu_2$$

2) Le ipotesi aggiuntive sono che i dati sono ottenuti campionando la stessa popolazione in due momenti diversi e siano realizzazioni di due v.a. Normali con camp$1 \sim N(\mu_1, \sigma_1^2)$ e camp$2 \sim N(\mu_2, \sigma_2^2)$

3). Dobbiamo trovare una statistica test la cui distribuzione teorica è nota quando è vera l'ipotesi nulla

$$t = \frac{D}{\frac{S_D}{\sqrt{n}}}$$

$$D = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^{camp1} - x_i^{camp2}\right) = media\ delle\ differenze$$

$$differenze S_d = Deviazione\ standard\ delle$$

$$n = numero\ di\ coppie$$

t una v.a. con densità di probabilità t-Student:

$$t \sim T(n-1)$$

TEST PARAMETRICI

L'ANALISI DELLA VARIANZA (ANOVA)

ANOVA = ANalysis Of VAriance

E’ un metodo statistico molto potente e flessibile per valutare le medie di più di due gruppi/popolazioni con una singola analisi

Si una per studiare variabili aleatorie quantitative

L’ipotesi nulla riguarda le medie ma viene testata operando sulle varianze

ANOVA AD UN FATTORE

ONE-WAY ANOVA

Quando si può usare l'ANOVA?

Per il confronto tra gruppi, è corretto calcolare l'ANOVA solo quando si verificano le seguenti condizioni:

1. **normalità della distribuzione** della/e variabile/i dipendente/i

2. **omoschedasticità** (o omogeneità delle varianze dei gruppi). In questo caso, puoi utilizzare il test di Levene.

Se entrambe le assunzioni sono rispettate, allora puoi procedere con il test dell'ANOVA.

Test di Shapiro-Wilk (test di normalità ➔ H0: la variabile è normale)

Indicato anche per piccoli campioni (n≤50), è basato sul confronto di due diversi modi di stima della devianza: uno basato sui ranghi e uno sull'uso della formula della devianza campionaria

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$x_{(i)} = i - esimo\ valore\ più\ piccolo\ (rango\ i)\ del\ campione$$
$$\bar{x} = media\ aritmetica$$
$$a_i = valori\ noti\ una\ volta\ fissato\ il\ numero\ di\ campioni$$
$$(il\ calcolo\ è\ basato\ sulla\ conoscenza\ dei\ momenti\ centrali\ di\ ordine\ 1\ e\ delle\ loro\ covarianze)$$

Osservando la statistica test, al numeratore troviamo uno stimatore non parametrico e al denominatore il consueto stimatore parametrico, ossia la varianza campionaria. **Il valore del test varia tra 0 e 1**.

Qualora il valore della statistica W sia troppo piccolo, il test rifiuta l'ipotesi nulla (H0) che i valori campionari siano distribuiti come una variabile casuale normale.
La statistica test sotto l'ipotesi nulla segue una distribuzione non nota, per cui sono stati tabulati i valori critici di questa distribuzione.

Test di Kolmogorov-Smirnov (test di normalità ➔ H0: la variabile è normale)

Il test di Kolmogorov-Smirnov è un test non parametrico che verifica la forma delle distribuzioni campionarie. In particolare, si basa sulla maggiore differenza verticale tra la funzione di distribuzione empirica F (x) e quella ipotetica F_{Ho}(x).
F_{Ho}(x) viene assunta avere una distribuzione normale con media μ e deviazione standard σ note.

La statistica test è definita in tre maniere differenti a seconda del tipo di ipotesi alternativa che si vuole considerare.

L'ipotesi nulla sarà sempre la seguente: $H0$: $F(x)$=$F_{Ho}(x)$, mentre delle tre ipotesi alternative che si possono considerare a noi interessa solitamente la statistica del test a una coda: calcolata come la distanza tra la funzione di ripartizione di riferimento e la funzione di ripartizione empirica del campione.

Nella sua formulazione esatta prevede che le variabili siano continue. Non richiede di per sé alcuna ipotesi sulla distribuzione campionaria, salvo nel caso a un campione, in cui viene testata una distribuzione a propria scelta.

Test di omogeneità della varianza

Ci sono vari test per valutare se le varianze di k gruppi di dati sono uguali o meno. I più noti sono:

Test di Levene
Test di Bartlett

Il test di Levene è ritenuto da vari statistici più robusto, rispetto alla non normalità della distribuzione, di quanto siano i test di rapporti tra varianze e del test di Bartlett.
Il test di Levene deve la sua diffusione anche all'inserimento in alcuni pacchetti statistici, che lo impongono come verifica preliminare di validità al test t di Student e all'ANOVA.

Del metodo di Levene esistono molte versioni, ma le più diffuse sono tre. La prima è la proposta originaria di Levene. Le altre due, che ne rappresentano delle modifiche, sono attribuite a Morton B. Brown e Alan B. Forsythe. In esso, al posto della media indicata da Levene, suggeriscono di utilizzare la mediana oppure la media trimmed al dieci per cento.

In modo più specifico, per la ten percent trimmed mean si intende la media del gruppo, ma dopo che da esso sono stati eliminati il 10% dei valori maggiori e il 10% dei valori minori. La scelta del 10% oppure di un'altra qualsiasi percentuale è puramente arbitraria.

Diversi autori sono molto critici sull'uso dei test per l'omogeneità della varianza. Infatti essi sono- fortemente alterati dalla non normalità della distribuzione- e con pochi dati è impossibile verificare se le varie distribuzioni campionarie possano essere ritenute prossime alla normale.

il test t permette di confrontare solo due gruppi

l’ANOVA permette di confrontare un numero qualsiasi di gruppi

Ad esempio, ipotizziamo che l’obiettivo sia confrontare il punteggio medio conseguito all’esame di statistica tra chi ha frequentato il corso in presenza e chi online. In questo caso, i gruppi sono solo due (frequentanti in presenza e frequentanti online). Pertanto, per il confronto delle medie si può usare indifferentemente il test t o l’ANOVA.

Ipotizziamo ora di voler approfondire l’analisi, suddividendo lo stesso campione di studenti in tre gruppi: chi ha frequentato solo in presenza, chi ha frequentato solo online, e chi ha frequentato un po’ online ed un po’ in presenza. In questo caso, per il confronto delle medie non si può più usare il test t ma si deve necessariamente ricorrere all’ANOVA.

<u>Un esempio con dati sperimentali:</u> la **variabile** *altezza* viene misurata in individui suddivisi in 4 gruppi; i gruppi sono sottoposti a diversi trattamenti per il **fattore** *ph*

- Supponiamo di voler capire se il ph influisce sulla crescita di una certa pianta
    - Quattro gruppi di 20 piante ciascuno vengono coltivati a 4 livelli diversi di ph, per esempio a ph 5, 6, 7, 8, per un certo periodo
    - Le piante sono inizialmente di altezza uguale, e vengono comunque distribuite a caso nei 4 gruppi a ph diverso
    - Alla fine dell'esperimento viene misurata l'altezza di ciascuna piantina (quindi, 80 valori in tutto)
    - Le medie di altezza tra sono diverse? Ovvero, il ph ha influito sulla crescita?

<u>Un esempio con dati osservazionali:</u> la **variabile** *peso* viene misurata in individui che provengono da 4 gruppi; i 4 gruppi differiscono per il **fattore** *origine geografica*

- Supponiamo di voler capire se animali di una certa specie di coleottero hanno caratteristiche di peso diverse in 4 diverse aree isolate (per esempio 4 località molto distanti)
    - Campioniamo 20 individui in ciascuna delle 4 popolazioni diverse
    - Ognuno degli 80 individui viene pesato
    - I pesi medi nelle 4 diverse aree di provenienza sono diversi? Se sì, quale potrebbe essere la causa? Per esempio le caratteristiche ecologiche?

Ipotesi nulla e alternativa nell'ANOVA

- In entrambe i casi, l'ipotesi nulla è quella di uguaglianza tra le medie nelle 4 popolazioni da cui provengono i campioni:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

- L'ipotesi alternativa è invece quella che prevede che ALMENO una media sia diversa. Non è quindi esprimibile in simboli in maniera semplice, ma NON è

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \quad \underline{\text{NO}}$$

- Inizialmente, quindi, l'ANOVA non cerca di identificare quale popolazione è diversa da quale altra, ma solamente se esiste almeno una media diversa da un'altra nelle popolazioni analizzate.

- Ovviamente l'ANOVA si applica nello stesso modo a 2,3,4,5,...k gruppi
  - Per k = 2, equivale a svolgere un test t

- Per testare l'ipotesi nulla che la media di una variabile in k popolazioni sia la stessa, si suddivide la variabilità totale della variabile

- La variabilità totale viene suddivisa in due componenti:

    1. La variabilità all'interno dei gruppi
    2. La variabilità tra i gruppi

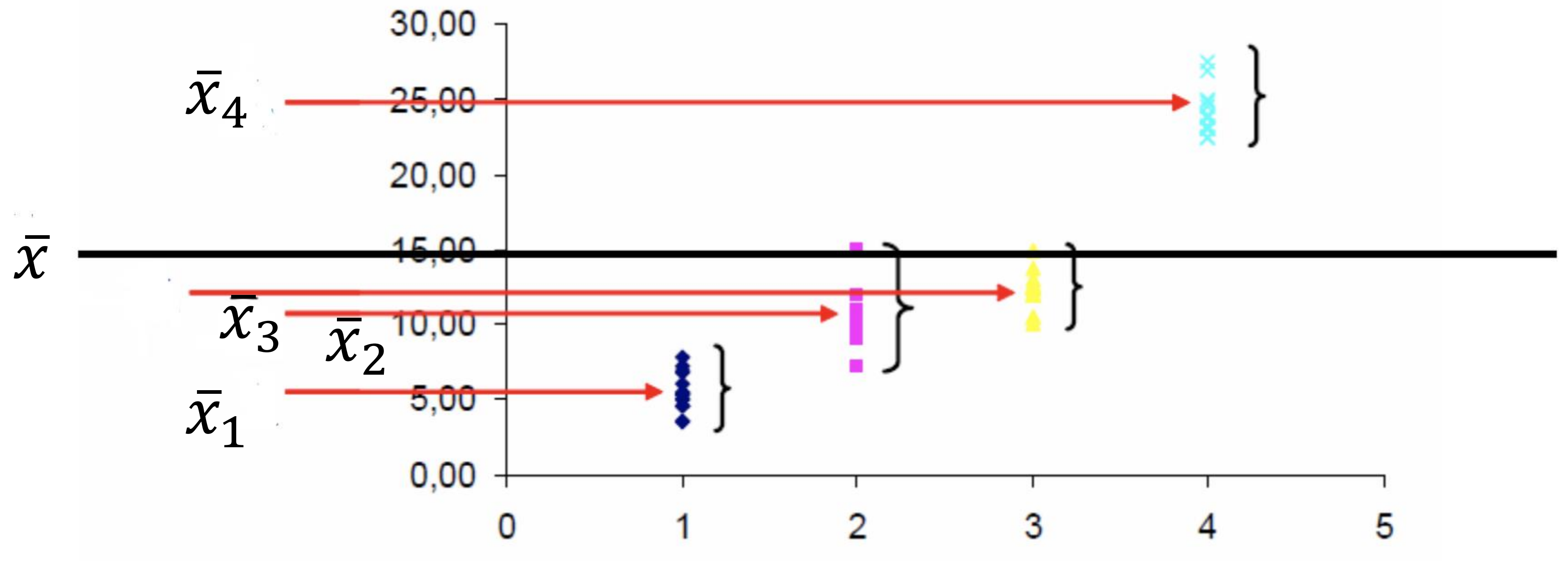
- Per vedere questa scomposizione, definiamo prima le medie dei k gruppi con i simboli:

$$\bar{x}_i = media\ aritmetica\ del\ gruppo\ i$$

- Definiamo anche la media generale con:

$$\bar{x}$$

E' semplicemente la media calcolata mettendo insieme tutti i dati di tutti i gruppi
Attenzione! NON è la media delle k medie calcolate nei singoli gruppi



Dall’esempio possiamo facilmente vedere che la variabilità complessiva può essere divisa in due componenti:

1. La variabilità entro gruppi, cioè quanto mediamente i singoli valori sono distanti dalla media del loro gruppo di appartenenza.
2. La variabilità tra gruppi, cioè quanto mediamente sono distanti le medie dei diversi gruppi dalla media generale

E' piuttosto intuitivo capire che più ci allontana dall'ipotesi nulla (H0: tutte le medie sono uguali) e più la variabilità tra gruppi diventerà grande, e rappresenterà una importante frazione della variabilità totale

E' anche intuitivo capire che la variabilità entro gruppi della variabilità totale non dipende dalla differenza tra i gruppi. E' una componente che considera semplicemente il fatto che non tutte le osservazioni, anche se appartenenti allo stesso gruppo o sottoposte allo stesso trattamento, sono uguali. E' anche definita come variabilità dell'errore del campionamento.

La componente della variabilità entro gruppi viene definita nell'ANOVA come media dei quadrati degli errori (MSE: Mean Square Error)

E' semplicemente la media pesata delle varianze calcolate all'interno dei gruppi, ovvero un'estensione a k gruppi della varianza comune già vista nel test (dove k=2). E' quindi una varianza, chiamata anche varianza dell'errore.

$$MSE = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{\sum_{i=1}^{k}(n_i - 1)} = \frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{n_T - k}$$

n_i = numerosità dell'i-esimo gruppo
n_T = numerosità totale (somma di tutti gli n_i)
s_i^2 = varianza campionaria dell'i-esimo gruppo

Il numeratore viene anche chiamato SSE o *somma dei quadrati dell'errore* o anche *devianza dell'errore*
Il denominatore rappresenta i gradi di libertà di questa componente della variabilità totale

La componente della variabilità tra gruppi viene definita nell'ANOVA come media dei quadrati tra gruppi (MSB: Mean Square Between groups).

Dipende da quanto sono distanti le medie dei gruppi dalla media generale, ma considera anche le numerosità dei singoli gruppi. E' anche questa una varianza, chiamata anche varianza tra gruppi.

$$MSB = \frac{\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2}{k-1}$$

Il numeratore di MSB viene chiamato *SSB*, o *somma dei quadrati tra gruppi*, o anche *devianza tra gruppi*

Il denominatore di MSB rappresenta i gradi di libertà di questa componente della variabilità totale (ci sono k gruppi, e quindi k-1 gradi di libertà)

Come già detto, più ci si allontana dall'ipotesi nulla e più tende a crescere la componente della variabilità tra gruppi. Quindi, *più ci si allontana dall'ipotesi nulla e più MSB tende a crescere*.

E' possibile dimostrare che quando è vera l'ipotesi nulla MSB tende ad essere uguale MSE: $MSB \approx MSE$

Ovviamente, se è vera l'ipotesi alternativa (almeno una media è diversa dalla altre), MSB sarà maggior di MSE (mai minore)

Poichè sia MSB che MSE sono due varianze, e il valore di MSB/MSE atteso quando è vera l'ipotesi nulla è 1, l'F di Fisher è la statistica test adatta all'ANOVA. In altre parole, dopo aver calcolato MSB e MSE, posso calcolare

$$\boldsymbol{F = \frac{MSB}{MSE}}$$

Come abbiamo detto più volte, l'ipotesi alternativa (almeno una media è diversa) prevede la deviazione di F solo verso valori >1 (cioè MSB>MSE). Quindi, anche se l'ipotesi alternativa nell' l'ANOVA non è unidirezionale, prevede deviazioni solo in una direzione della distribuzione nulla di Fisher.

- Una tabella utile per riassumere i risultati dell'ANOVA è la seguente

Servono anche per definire la v.a. F

Origine della variazione	Gradi di libertà	SS	MS	F	P-value
Tra gruppi	$k-1$	SSB	$MSB = SSB/(k-1)$	$F_{calc} = MSB/MSE$	$P(F>F_{calc})$
Entro gruppi	n_T-k	SSE	$MSE = SSE/(n_T-k)$		
Totale	n_T-1	SSTO			

- I gradi di libertà e la somma dei quadrati (SS) godono della proprietà additiva, ma non le medie dei quadrati (MS)
  - Questa proprietà può essere utile

- Nell'ultima colonna, se non dispongo di un calcolatore che mi permette di determinare il P-value, posso riportare il valore critico di F per l'α scelto e quindi se il valore calcolato supera quello critico, indicare $P< \alpha$, altrimenti $P> \alpha$.

LE ASSUNZIONI DELL'ANOVA

- Sono le assunzioni del test t, ma estese a tutti i gruppi:
  - La variabile deve avere una distribuzione normale in tutte le popolazioni corrispondenti ai gruppi campionati
  - Le varianze in tutte le popolazioni corrispondenti ai gruppi campionati deve essere uguale

- Ovviamente, come sempre, per ciascun gruppo il campione deve rappresentare un insieme di misure estratte a caso dalla corrispondente popolazione

- E' necessario verificare che queste assunzioni vengano soddisfatte

- Fortunatamente però, l'ANOVA è un'analisi piuttosto robusta a violazioni di queste assunzioni, soprattutto se i campioni hanno circa le stesse numerosità

- Prima di vedere <u>come</u> si procede nell'ANOVA, vediamo <u>perchè</u> svolgere un'ANOVA

- Per esempio, con 3 popolazioni da confrontare (per esempio, tre livelli di pH) non potrei semplicemente fare 3 test t? O con 4 popolazioni 6 test t?
    - Come si calcola il numero di test a coppie?

- No, perchè

    1. Sembra logico prima di tutto testare l'ipotesi nulla che prevede che tutti i gruppi siano uguali

    2. Non posso semplicemente fare tanti test t perchè aumenterebbe molto l'errore complessivo di primo tipo

<u>Il problema dei test multipli e l'errore complessivo di primo tipo</u>

- Se scegliamo in un singolo test un livello di significatività α, sappiamo che esiste una probabilità α di rifiutare un'ipotesi nulla vera (errore di primo tipo)

- Questo significa anche che se facciamo 100 test nei quali l'ipotesi nulla è sempre vera, 5 volte (mediamente) la rifiutiamo erroneamente

Una chiave per i test parametrici sulle medie