UNIVERSITÀ DEGLI STUDI DI TRIESTE
Dipartimento di Ingegneria e Architettura

Corso di misure meccaniche, termiche e collaudi

Prof. Rodolfo Taccani
Prof. Lucia Parussini
Prof. Marco Bogar

a.a.2024-2025

Outline

- Introduzione
- Test d’ipotesi
- Interpretazione del risultato del test d’ipotesi
- Esempi di test di ipotesi

Introduzione

La statistica è una parte fondamentale di un efficiente processo di progettazione e produzione, di verifica , di collaudo.

Fasi:

- progettare esperimenti
- eseguire esperimenti
- raccogliere dati reali
- prendere decisioni basate sui dati

Questi esperimenti possono essere usati

- per verifiche di conformità e prestazioni
- per testare l'effetto del cambiamento di una dimensione, di un materiale o di una fase del processo produttivo

Introduzione

Questi esperimenti possono essere usati

- ❑ per verifiche di conformità e prestazioni
- ❑ per testare l'effetto del cambiamento di una dimensione, di un materiale o di una fase del processo produttivo

Ad esempio verifica delle prestazioni attraverso il confronto dei dati:

- ○ di due strumenti
- ○ di due metodi
- ○ di una procedura in periodi diversi
- ○ di due analisti o laboratori
- ○ con i risultati ottenuti per un campione di riferimento o di controllo con il valore "vero", "target" o "assegnato" di questo campione

Introduzione

Questi esperimenti possono essere usati

- ❑ per verifiche di conformità e prestazioni
- ❑ per testare l'effetto del cambiamento di una dimensione, di un materiale o di una fase del processo produttivo

Di solito si vuole cambiare una di queste variabili per migliorare la capacità del pezzo di essere realizzato entro le tolleranze o le prestazioni richieste per il prodotto.

Ad esempio:

- o il pezzo continua a rompersi durante i test di affidabilità, e volete testare se cambiare il materiale riduce questi fallimenti
- o dovete rilavorare una grande percentuale dei vostri pezzi, ma pensate che cambiare una dimensione renderà le rilavorazioni non necessarie e accelererà il processo

Hypotesis testing

Quando vogliamo interpretare i risultati, è consigliabile usare metodi statistici che forniscono una probabilità sulle risposte.

Questa classe di metodi è chiamata **test di ipotesi** o **test di significatività**.

Cos'è un test d'ipotesi?

Un test di ipotesi è un metodo per testare un presupposto e determinare se il risultato è statisticamente significativo.

Hypotesis testing

ipotesi statistica: affermazione o congettura, riguardante un parametro Θ che caratterizza il modello descrittivo della popolazione, $f(x;\theta)$ con $\theta \in \Theta$, dove Θ è lo spazio parametrico

L'assunzione (che traduce la congettura iniziale) di un test di significatività è chiamata ipotesi nulla, o ipotesi 0 (H_0 in breve). È spesso chiamata l'ipotesi di default.

Una violazione dell'ipotesi del test è spesso chiamata prima ipotesi, ipotesi 1 o H_1 in breve. H_1 è in realtà un'abbreviazione per "qualche altra ipotesi", poiché tutto ciò che sappiamo è che l'evidenza suggerisce che la H_0 può essere rifiutata.

Ipotesi 0 (H_0): ipotesi valida fino a prova contraria (e che si vorrebbe fosse falsa)
Ipotesi 1 (H_1): ipotesi alternativa che si vuole verificare

$$H_0: \theta \epsilon \Theta_0 \subset \Theta \qquad H_1: \theta \epsilon \Theta_1 \subset \Theta$$

Hypotesis testing

Un'ipotesi si dice semplice se specifica completamente la distribuzione, cioè $\theta = \theta_0$; altrimenti si dice composta.

Esempio:
"la durata media dei cuscinetti non è superiore a 56.000 ore"

$$H_0: \mu \leq 56000\ ore \qquad\qquad H_1: \mu > 56000\ ore$$

$\mu = 56000\ ore$ ipotesi semplice
$\mu \neq 56000\ ore$ ipotesi composta
$\mu > 56000\ ore$ ipotesi composta

Hypotesis testing

Per eseguire la verifica d'ipotesi è necessario raccogliere un certo campione di dati:

$$X_1, X_2, \dots, X_n \; i.i.d$$

dei quali si otterrà una certa realizzazione

$$x_1, x_2, \dots, x_n$$

La regola di decisione è un insieme, che chiamiamo regione critica e indichiamo con $\mathcal{G}$ e che definiamo come l'insieme di tutti i risultati sperimentali per i quali rifiutiamo l'ipotesi nulla

$$\mathcal{G} = \{(x_1, \dots, x_n) \; per \; le \; quali \; rifiuto \; H_0\}$$

Interpretazione del test di significatività

Prima di poter rifiutare o meno l'ipotesi nulla a un certo livello di significatività, dobbiamo interpretare il risultato del test.

Ci sono due forme comuni che un risultato di un test di ipotesi statistica può assumere, e devono essere interpretati in modi diversi.

Sono il valore p e i valori critici.

Interpretare il valore di p

Un test di ipotesi statistica restituisce un valore chiamato p o p-value.

Il valore p viene confrontato con un valore di soglia α scelto in precedenza, chiamato livello di significatività.

Un valore comunemente usato per α è 5% o 0,05.
Un valore α più piccolo suggerisce un'interpretazione più robusta dell'ipotesi nulla, come l'1% o lo 0,1%.

Un risultato è statisticamente significativo quando il valore p è inferiore ad α. Questo significa che l'ipotesi predefinita può essere rifiutata.

se p-value $> \alpha$: mancato rifiuto dell'ipotesi nulla (cioè risultato non significativo)
se p-value $\leq \alpha$: rifiuto dell'ipotesi nulla (cioè risultato significativo)

Interpretare il valore di p

Per esempio, se stessimo eseguendo un test per verificare se un campione di dati ha distribuzione normale e calcolassimo un valore p di 0.07, potremmo affermare qualcosa come:

Il test ha trovato che il campione di dati ha distribuzione normale, non riuscendo a rifiutare l'ipotesi nulla a un livello di significatività del 5%.

livello di confidenza = 1 - livello di significatività

Pertanto, si può equivalentemente affermare che:

Il test ha trovato che i dati hanno distribuzione normale, non riuscendo a rifiutare l'ipotesi nulla ad un livello di confidenza del 95%.

Interpretare il valore di p

Il p-value è probabilistico: è una misura della probabilità che il campione di dati sia osservato nel caso in cui l'ipotesi nulla sia vera.

Questo significa che quando interpretiamo il risultato di un test statistico, non sappiamo cosa è vero o falso, solo cosa è probabile.

Rifiutare l'ipotesi nulla significa che c'è sufficiente evidenza statistica che l'ipotesi nulla non sia probabile.

Non rifiutare l'ipotesi nulla significa che non ci sono prove statistiche sufficienti per rifiutare l'ipotesi nulla.

NOTA: se diciamo che "accettiamo" l'ipotesi nulla, il linguaggio suggerisce che l'ipotesi nulla è vera. Invece, è più sicuro dire che "non possiamo rifiutare" l'ipotesi nulla, come dire, non ci sono prove statistiche sufficienti per rifiutarla.

Ricordate che il risultato è probabilistico e che anche un'ipotesi nulla "accettata" ha ancora una piccola probabilità di essere sbagliata.

Interpretare il valore di p

Messa a punto post-hoc

Non significa che potete ricampionare il vostro dominio o sintonizzare il vostro campione di dati e rieseguire il test statistico fino ad ottenere il risultato desiderato.
Non significa nemmeno che potete scegliere il vostro p-value dopo aver eseguito il test.
Questo si chiama p-hacking o hill climbing e significa che il risultato che presenterete sarà fragile e non rappresentativo. Nella scienza, questo è nel migliore dei casi non etico, e nel peggiore una frode.

Interpretare i valori critici

Alcuni test restituiscono una lista di valori critici e i loro livelli di significatività associati, così come una statistica del test.

Questi sono di solito test di ipotesi statistica non parametrici o senza distribuzione.

se $statistica\ test < valore\ critico$: mancato rifiuto dell'ipotesi nulla (cioè risultato non significativo)
se $statistica\ test \geq valore\ critico$: rifiuto dell'ipotesi nulla (cioè risultato significativo)

Interpretare i valori critici

I risultati sono presentati nello stesso modo di un valore p, come livello di significatività o livello di confidenza.

Per esempio, se è stato calcolato un test di normalità e la statistica del test è stata confrontata con il valore critico al livello di significatività del 5%, i risultati potrebbero essere dichiarati come:

Il test ha trovato che il campione di dati era normale, non potendo rifiutare l'ipotesi nulla al livello di significatività del 5%.

oppure:

Il test ha trovato che i dati erano normali, non potendo rifiutare l'ipotesi nulla ad un livello di confidenza del 95%.

Errori nei test di ipotesi statistici

L'interpretazione di un test di ipotesi statistica è probabilistica. Ciò significa che le prove del test possono suggerire un risultato ed essere sbagliate.

Per esempio, se $\alpha = 5\%$, suggerisce che (al massimo) 1 volta su 20 l'ipotesi nulla sarebbe erroneamente rifiutata o non sarebbe stata rifiutata a causa del rumore statistico nel campione di dati.

Dato un piccolo p-value (rifiutare l'ipotesi nulla) può significare che l'ipotesi nulla è falsa (ci abbiamo azzeccato) o è vera ed è stato osservato qualche evento raro e improbabile (abbiamo fatto un errore).
FALSO POSITIVO: crediamo falsamente al rifiuto dell'ipotesi nulla.

Dato un grande p-value (mancato rifiuto dell'ipotesi nulla), può significare che l'ipotesi nulla è vera (ci abbiamo azzeccato) o è falsa e si è verificato qualche evento improbabile (abbiamo commesso un errore).
FALSO NEGATIVO: crediamo falsamente all'ipotesi nulla.

Errori nei test di ipotesi statistici

Errore di tipo I

Il rifiuto errato di un'ipotesi nulla vera o un falso positivo.

Errore di tipo II

Il mancato rifiuto errato di una falsa ipotesi nulla o un falso negativo.

Realtà \ *Decisione*	**Non rifiuto H_0**	**Rifiuto H_0**
H_0 è vera	**Decisione giusta**	**Errore di I tipo**
H_0 è falsa	**Errore di II tipo**	**Decisione giusta**

Un livello di significatività molto piccolo, minimizza la probabilità di uno di questi errori.

Ad esempio, $\alpha = $3e-7 significa che il rifiuto dell'ipotesi nulla è dovuto al caso con una probabilità di 1 su 3,5 milioni di ripetizioni indipendenti degli esperimenti.

Esempi di test di ipotesi

Test del tipo di distribuzione delle variabili (gaussiana)
Test di Shapiro-Wilk
Test chi-quadro di D'Agostino
Test di Anderson-Darling

Test di relazione tra variabili (correlazione)
Coefficiente di correlazione di Pearson
Correlazione di rango di Spearman
Correlazione di rango di Kendall
Test chi-quadro

Confronto delle distribuzioni campionarie
Test di Komogorov-Smirnov

Confronto delle medie dei campioni (parametrico)
Test t-Student
Test t-Student per dati accoppiato
Test di analisi della varianza (ANOVA)

Confronto delle medie dei campioni (non parametrico)
Test U di Mann-Whitney
Test di Wilcoxon Signed-Rank
Test H di Kruskal-Wallis
Test di Friedman

Confronto di due varianze
Test F di Fisher

Guida all'uso dei test statistici

Scelta del test statistico

Che tipo di variabili devo studiare?

Nominali | Ordinali | Continue

Se la variabile è continua è distribuita normalmente?

Quanti sono i gruppi che devo confrontare?

1 gruppo | 2 gruppi | 3 o più gruppi

- **Metodi parametrici.** Basati su media e deviazione standard (o varianza). Si utilizzano perciò nel caso di variabili distribuite normalmente.
- **Metodi non parametrici.** Basati su frequenza e percentili. Si utilizzano perciò nel caso di variabili la cui distribuzione non sia normale.

Guida all’uso dei test statistici

Verifica di ipotesi su medie: due gruppi

indipendenti		dipendenti	
distribuzione gaussiana test parametrico	distribuzione non gaussiana test non parametrico	distribuzione gaussiana test parametrico	distribuzione non gaussiana test non parametrico
t-Student ❑ con varianze omogenee ❑ con varianze non omogenee	Wilcoxson somma dei ranghi	t-Student	Wilcoxson ranghi con segno

Guida all’uso dei test statistici

Verifica di ipotesi su medie: più di due gruppi

- **indipendenti**
  - distribuzione gaussiana
test parametrico
    - Analisi della Varianza
F-test di Fischer
  - distribuzione non gaussiana
test non parametrico
    - Analisi della Varianza
test di Kruskall-Wallis
- **dipendenti**
  - distribuzione gaussiana
test parametrico
    - Analisi della Varianza
per misure ripetute
  - distribuzione non gaussiana
test non parametrico
    - Analisi della Varianza
per misure ripetute
test di Friedman

Guida all’uso dei test statistici

Tipo di quesito	Test parametrico	Variabile dipendente	Variabile indipendente	Esempio
Relazione fra due variabili continue	Regressione lineare	Continua	Continua	Come varia la vibrazione del rotore al variare del numero di giri?
Forza della correlazione fra due variabili continue	Coefficiente di correlazione di Pearson	Continua	Continua	Quanto è forte la correlazione tra la concentrazione nell’aria di particolato sospeso e la distanza dall’inceneritore?
Confronto dei valori medi di una variabile continua tra due gruppi	Test t di Student	Continua	Categorica a 2 livelli	I valori di temperatura misurati con due diversi strumenti di misura differiscono significativamente?
Confronto dei valori medi di una variabile continua misurata in un gruppo in due occasioni diverse	Test t di Student per dati appaiati	Continua	Categorica a 2 livelli	Confronto tra due metodi per prevedere la resistenza di taglio per le travi d'acciaio (Karlsruhe e Lehigh): c’è differenza tra i due metodi (media del dato previsto sul dato osservato)?
Confronto dei valori medi di una variabile continua fra più di due gruppi	Analisi della varianza (ANOVA)	Continua	Categorica a più di 2 livelli	I valori di temperatura misurati con tre diversi strumenti di misura differiscono significativamente?
Associazione fra una variabile dipendente e due o più variabili indipendenti	Regressione lineare multipla	Continua	Continue, categoriche	La vita effettiva di un utensile da taglio dipende dalla velocità di taglio e dall'angolo dell'utensile?

Guida all'uso dei test statistici

Tipo di quesito	Test non parametrico	Variabile dipendente	Variabile indipendente	Esempio
Relazione fra due variabili categoriche	Chi-quadrato	Categorica	Categorica	La percentuale di prodotti non conformi è significativamente diversa tra macchina A e macchina B?
Forza della correlazione fra due variabili	Coefficiente di correlazione di Spearman, Tau-b di Kendall	Continua, ordinale	Continua, ordinale	Quanto è forte l'associazione tra i livelli di usura del cuscinetto e le vibrazioni?
Confronto dei valori di una variabile continua/ordinale tra due gruppi	Mann-Whitney U-test	Continua, ordinale	Categorica a 2 livelli	La concentrazione nell'aria di particolato sospeso misurata in zona A e in zona B non differisce significativamente?
Confronto dei valori di una variabile continua/ordinale misurata in un gruppo in due occasioni diverse	Wilcoxson signed ranks test	Continua, ordinale	Categorica a 2 livelli	C'è una differenza nei tempi di reazione in una stanza con due diverse condizioni di illuminazione?
Confronto dei valori di una variabile continua/ordinale fra più di due gruppi	Kruskal-Wallis test	Continua, ordinale	Categorica a più di 2 livelli	L'ansia da esame influenza i punteggi finali del test?
Associazione fra una variabile dipendente e due o più variabili indipendenti	Regressione logistica	Categorica a 2 livelli	Continue, categoriche	Quali fattori si associano al rischio di avere rottura?

Verifica di ipotesi sulla media con varianza nota (Z-Test)

Restituisce una decisione per il test dell'ipotesi nulla che i dati in $\mathbf{x}$ provengano da una distribuzione normale con media μ_0 e varianza σ_0^2.

Null hypothesis $H_0\colon \bar{x} = \mu_0$

$$z = \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}$$

Alternative hypothesis $H_1\colon \bar{x} > \mu_0$

$$\bar{x} > \mu_0 + z_{1-\alpha}\, \sigma_0/\sqrt{n} \quad \text{ovvero} \quad z > z_{1-\alpha}$$

Alternative hypothesis $H_1\colon \bar{x} < \mu_0$

$$\bar{x} < \mu_0 - z_{1-\alpha}\, \sigma_0/\sqrt{n} \quad \text{ovvero} \quad z < z_{\alpha}$$

Alternative hypothesis $H_1\colon \bar{x} \neq \mu_0$

$$\bar{x} < \mu_0 - z_{1-\alpha/2}\, \sigma_0/\sqrt{n} \;\text{ o }\; \bar{x} > \mu_0 + z_{1-\alpha/2}\, \sigma_0/\sqrt{n} \;\text{ ovvero }\; z < z_{\alpha/2} \text{ o } z > z_{1-\alpha/2}$$

Nota

Si usano gli Z-test per la media anche nel caso in cui si abbiano grandi campioni (in genere ≥ 30) ma non di tipo gaussiano, perché si sfrutta l'ipotesi di asintotica gaussianità della media campionaria. In questo caso ovviamente si tratterà però non di un test esatto, bensì di un test *asintotico*.

Verifica di ipotesi sulla media con varianza incognita (T-Test)

Restituisce una decisione per il test dell'ipotesi nulla che gli n dati in $\mathbf{x}$ provengano da una distribuzione normale con media μ_0 e varianza sconosciuta.

Null hypothesis $H_0: \bar{x} = \mu_0$

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

Alternative hypothesis $H_1: \bar{x} > \mu_0$

$\bar{x} > \mu_0 + t_{1-\alpha,\nu} s_{\bar{x}}$ ovvero $t > t_{1-\alpha,\nu}$

Alternative hypothesis $H_1: \bar{x} < \mu_0$

$\bar{x} < \mu_0 - t_{1-\alpha,\nu} s_{\bar{x}}$ ovvero $t < t_{\alpha,\nu}$

Alternative hypothesis $H_1: \bar{x} \neq \mu_0$

$\bar{x} < \mu_0 - t_{1-\alpha/2,\nu} s_{\bar{x}}$ o $\bar{x} > \mu_0 + t_{1-\alpha/2,\nu} s_{\bar{x}}$ ovvero $t < t_{\alpha/2,\nu}$ o $t > t_{1-\alpha/2,\nu}$

$\nu = n - 1$

Verifica di ipotesi sulla media con varianza incognita (T-Test)

Esempio 1 - Right tailed test- Prova di durezza dei materiali

Un ingegnere ha misurato la durezza Brinell di 25 pezzi di ferro dolce che sono stati ricotti subcriticamente. I dati risultanti erano:

170	167	174	179	179	187	179	183	179
156	163	156	187	156	167	156	174	170
183	179	174	179	170	159	187		

L'ingegnere ha ipotizzato che la durezza Brinell media di *tutti* questi pezzi di ferro dolce sia maggiore di 170. Pertanto, era interessato a testare le ipotesi:

$H_0 : \mu = 170$

$H_A : \mu > 170$

L'ingegnere ha condotto il test t per le ipotesi di cui sopra. Ha ottenuto il seguente risultato:

Descriptive Statistics

N	Mean	StDev	StDev Mean	95% Lower Bound
25	172.52	10.31	2.06	168.99

μ : mean of Brinelli

Test

Null hypothesis $H_0 : \mu = 170$

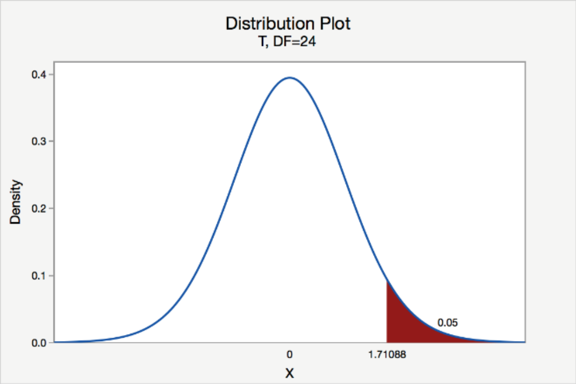
Alternative hypothesis $H_1 : \mu > 170$

t-Value	p-Value
1.22	0.117

Verifica di ipotesi sulla media con varianza incognita (T-Test)

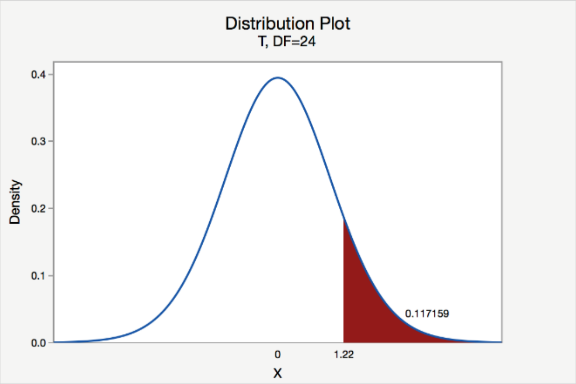
Esempio 1 - Right tailed test- Prova di durezza dei materiali

Se l'ingegnere ha fissato il suo livello di significatività α a 0.05 e ha usato l'approccio del valore critico per condurre il suo test d'ipotesi, rifiuterebbe l'ipotesi nulla se la sua statistica t fosse maggiore di 1.7109):



t = 1.22 ≤ $t_{0.95,24}$=1.7109, quindi non ci sono prove sufficienti, al livello di significatività α = 0.05, per concludere che la durezza Brinell media di tutti questi pezzi di ferro duttile sia maggiore di 170.

Se l'ingegnere usasse l'approccio del valore p per condurre il suo test d'ipotesi, determinerebbe l'area sotto una curva $t_{n-1} = t_{24}$ e a destra della statistica del test t = 1.22:



$p = \int_{1.22}^{\infty} f_S(t|\nu = 24)dt$= 0.117 > 0.05, quindi non ci sono prove sufficienti, al livello α = 0.05, per concludere che la durezza Brinell media di tutti questi pezzi di ferro duttile sia maggiore di 170.

Verifica di ipotesi sulla media con varianza incognita (T-Test)

Esempio 2 - Left tailed test- Altezza dei girasoli

Un biologo era interessato a determinare se le piantine di girasole trattate con un estratto di radici di *Vinca minor* avessero un'altezza media inferiore a quella standard di 15.7 cm. Il biologo ha trattato un campione casuale di n = 33 piantine con l'estratto e successivamente ha ottenuto le seguenti altezze:

11.5	11.8	15.7	16.1	14.1	10.5	9.3	15	11.1	15.2	19
12.8	12.4	19.2	13.5	12.2	13.3	16.5	13.5	14.4	16.7	10.9
13	10.3	15.8	15.1	17.1	13.3	12.4	8.5	14.3	12.9	13.5

Le ipotesi del biologo sono:

$H_0: \mu = 15.7$

$H_A: \mu < 15.7$

Il biologo ha condotto il t-test per le ipotesi di cui sopra. Ha ottenuto il seguente risultato:

Descriptive Statistics

N	Mean	StDev	StDev Mean	95% Upper Bound
33	13.664	2.544	0.443	14.414

µ: mean of Height

Test

Null hypothesis $H_0: \mu = 15.7$

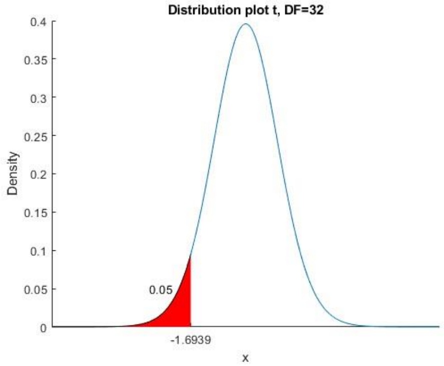
Alternative hypothesis $H_1: \mu < 15.7$

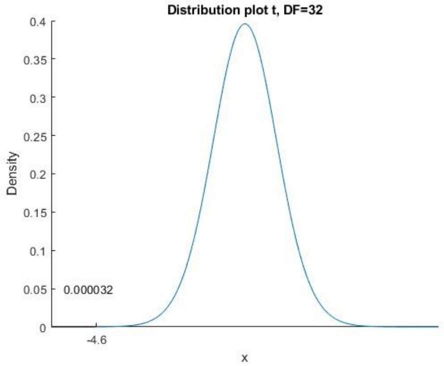
t-Value	p-Value
-4.60	0.000

Verifica di ipotesi sulla media con varianza incognita (T-Test)

Esempio 2 - Left tailed test- Altezza dei girasoli

t-Value	p-Value
-4.60	0.000





Al livello di significatività $\alpha = 0.05$, t=-4.60 < $t_{0.05,32}$=-1.6939

Al livello di significatività $\alpha = 0.05$, $p = \int_{-\infty}^{-4.60} f_S(t|\nu = 32)dt$=0.000032< 0.05

Ci sono prove sufficienti, al livello $\alpha = 0.05$, per rifiutare l'ipotesi nulla e concludere che l'altezza media di tutte queste piantine di girasole è inferiore a 15.7 cm.

Verifica di ipotesi sulla media con varianza incognita (T-Test)

Esempio 3 - Two tailed test- Spessore della gomma

Un produttore sostiene che lo spessore della gomma alla menta che produce è di 7.5 centesimi di pollice. Uno specialista del controllo qualità verifica regolarmente questa affermazione. In una produzione, ha preso un campione casuale di n = 10 pezzi di gomma e ne ha misurato lo spessore. Ha ottenuto:

7.65	7.60	7.65	7.70	7.55
7.55	7.40	7.40	7.50	7.50

Le ipotesi dello specialista del controllo qualità sono:

$H_0: \mu \;\; = 7.5$

$H_A: \mu \;\; \neq 7.5$

Lo specialista del controllo ha condotto il t-test per le ipotesi di cui sopra. Ha ottenuto il seguente risultato:

Descriptive Statistics

N	Mean	StDev	StDev Mean	95% CI for µ
10	7.550	0.1027	0.0325	(7.4765, 7.6235)

µ: mean of Thickness

Test

Null hypothesis $\quad H_0: \mu \; = 7.5$

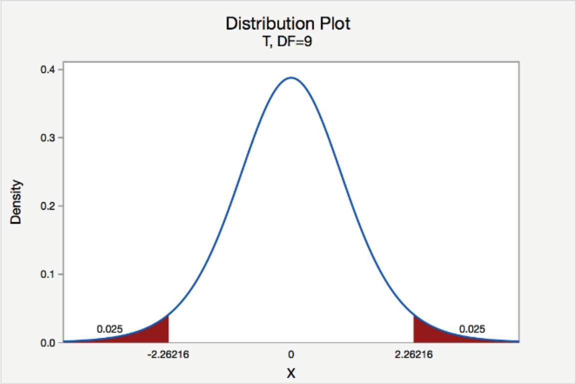
Alternative hypothesis $\quad H_0: \mu \; \neq 7.5$

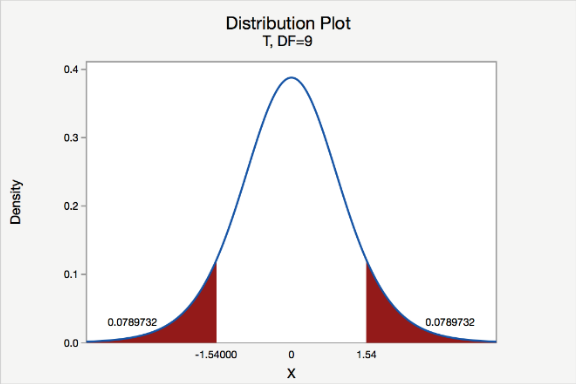
t-Value	p-Value
1.54	0.158

Verifica di ipotesi sulla media con varianza incognita (T-Test)

Esempio 3 - Two tailed test- Spessore della gomma

t-Value	p-Value
1.54	0.158





Al livello di significatività $\alpha = 0.05$, $t_{0.025,9}$=-2.2616 < t=1.54 < $t_{0.975,9}$=2.2616

Al livello di significatività $\alpha = 0.05$, p=$\int_{-\infty}^{-1.54} f_S(t|\nu = 9)dt + \int_{1.54}^{\infty} f_S(t|\nu = 9)dt$=0.158 > 0.05

Non ci sono prove sufficienti, al livello $\alpha = 0.05$, per rifiutare l'ipotesi nulla e concludere che lo spessore medio di tutti i pezzi di gomma alla menta verde differisce da 7.5 centesimi di pollice.

Test chi-quadro sulla varianza – caso a media nota

Restituisce una decisione per il test dell'ipotesi nulla che i dati in $\mathbf{x}$ provengano da una distribuzione normale con varianza σ_0^2 e media μ_0.

Null hypothesis $\quad H_0: s^2 = \sigma_0^2$

$$\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \mu_0)^2}{\sigma_0^2}$$

Alternative hypothesis $\quad H_1: s^2 > \sigma_0^2$

$$s^2 > \sigma_0^2 \chi_{1-\alpha,\nu}^2/(n-1) \quad \text{ovvero} \quad \chi^2 > \chi_{1-\alpha,\nu}^2$$

Alternative hypothesis $\quad H_1: s^2 < \sigma_0^2$

$$s^2 < -\sigma_0^2 \chi_{1-\alpha,\nu}^2/(n-1) \quad \text{ovvero} \quad \chi^2 < \chi_{\alpha,\nu}^2$$

Alternative hypothesis $\quad H_1: s^2 \neq \sigma_0^2$

$$s^2 < -\sigma_0^2 \chi_{1-\alpha/2,\nu}^2/(n-1) \;\text{ o }\; s^2 > \sigma_0^2 \chi_{1-\alpha/2,\nu}^2/(n-1) \;\text{ ovvero }\; \chi^2 < \chi_{\alpha/2,\nu}^2 \;\text{ o }\; \chi^2 > \chi_{1-\alpha/2,\nu}^2$$

$\nu = n - 1$

Nota A differenza degli Z-test sulla media, i chi-quadro test sulla varianza non possono essere utilizzati se il campione fornito non ha distribuzione normale: in assenza di un campione gaussiano, anche di grandi dimensioni, questi test non valgono.

Test chi-quadro sulla varianza – caso a media incognita

Restituisce una decisione per il test dell'ipotesi nulla che i dati in $\mathbf{x}$ provengano da una distribuzione normale con varianza σ_0^2 e media sconosciuta.

Null hypothesis $H_0: s^2 = \sigma_0^2$

$$\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_0^2}$$

Alternative hypothesis $H_1: s^2 > \sigma_0^2$

$s^2 > \sigma_0^2 \chi^2_{1-\alpha,\nu}/(n-1)$ ovvero $\chi^2 > \chi^2_{1-\alpha,\nu}$

Alternative hypothesis $H_1: s^2 < \sigma_0^2$

$s^2 < -\sigma_0^2 \chi^2_{1-\alpha,\nu}/(n-1)$ovvero $\chi^2 < \chi^2_{\alpha,\nu}$

Alternative hypothesis $H_1: s^2 \neq \sigma_0^2$

$s^2 < -\sigma_0^2 \chi^2_{1-\alpha/2,\nu}/(n-1)$ o $s^2 > \sigma_0^2 \chi^2_{1-\alpha/2,\nu}/(n-1)$ ovvero $\chi^2 < \chi^2_{\alpha/2,\nu}$ o $\chi^2 > \chi^2_{1-\alpha/2,\nu}$

$\nu = n - 1$

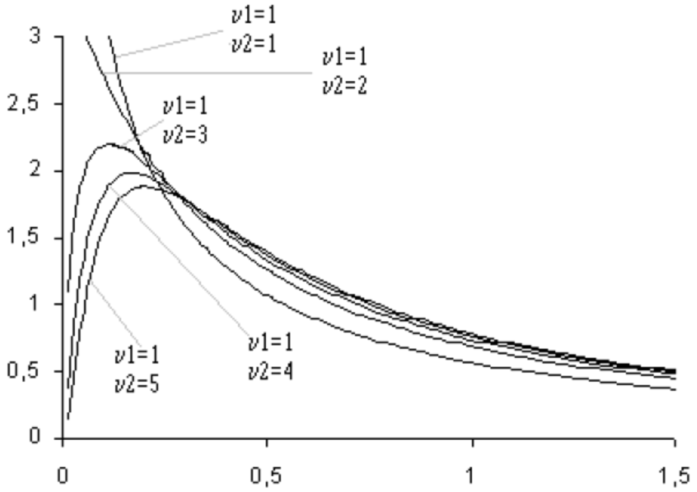
F-Test sulla varianza

Consideriamo una popolazione normalmente distribuita con parametri μ e σ^2. Scegliamo due numeri interi n_1 e n_2. Estraiamo ora due campioni *indipendenti* di dimensione n_1 e n_2 (successivamente, e con reimmissione dopo la formazione del primo), e calcoliamo le corrispondenti varianze s_1^2 e s_2^2.
A causa dell'indipendenza dei due campioni si dice che le due stime s_1^2 e s_2^2 della varianza σ^2 sono indipendenti. Di queste due varianze calcoliamo il rapporto:

$$F = \frac{s_1^2}{s_2^2}$$

Trattandosi di due stime di una medesima varianza è ragionevole pensare che questo rapporto sia vicino ad 1. Dopo aver reintrodotto gli elementi estratti nella popolazione, procediamo all'estrazione di due nuovi campioni, sempre di dimensione n_1 e n_2, e calcoliamone ancora le varianze s_1^2 e s_2^2. Ancora una volta calcoliamo il rapporto F delle due varianze. Ripetiamo ancora indefinitamente questa operazione. Otterremo per questa via una distribuzione di valori della statistica F.

F-Test sulla varianza



La distribuzione F dipende dai gradi di libertà
$\nu_1 = n_1 - 1$
$\nu_2 = n_2 - 1$

Si è fin qui parlato di una popolazione normalmente distribuita con parametri μ e σ^2; in realtà questa condizione è un po' più restrittiva del necessario: in realtà il campionamento può avvenire anche da due differenti popolazioni, sempre distribuite normalmente, con medie μ_1 e μ_2 differenti ma con la stessa varianza σ^2. La condizione di omogeneità delle varianze è nota col termine di *omoschedasticità*.
In sintesi, la statistica F è il rapporto di due stime indipendenti della comune varianza di due popolazioni normalmente distribuite e omoschedastiche.

F-Test sulla varianza

Null hypothesis $H_0: s_1^2 = s_2^2$

$$F = \frac{s_1^2}{s_2^2}$$

dove s^2 più grande deve essere il numeratore per convenzione.

Alternative hypothesis $H_1: s_1^2 > s_2^2$ $F > F_{1-\alpha,\nu_1,\nu_2}$

Alternative hypothesis $H_1: s_1^2 \neq s_2^2$ $F > F_{1-\alpha/2,\nu_1,\nu_2}$

$\nu_1 = n_1 - 1$

$\nu_2 = n_2 - 1$

F-Test sulla varianza

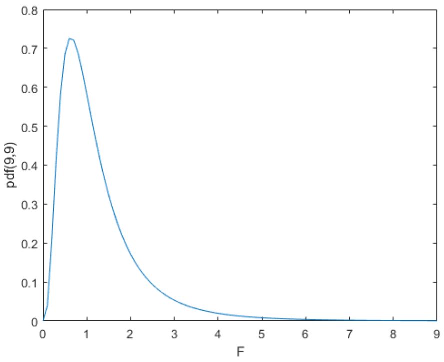
Esempio: *two-sided test*

La tabella fornisce le serie di dati ottenute da due analisti. Verificare se c'è una differenza significativa nella precisione tra il lavoro dell'analista A e quello dell’analista B.

A	B
10.2	9.7
10.7	9
10.5	10.2
9.9	10.3
9	10.8
11.2	11.1
11.5	9.4
10.9	9.2
8.9	9.8
10.6	10.2

$$H_0: s_A^2 = s_B^2$$
$$H_1: s_A^2 \neq s_B^2 \qquad F \geq F_{1-\alpha/2,\nu_A,\nu_B}$$

$$F = \frac{s_A^2}{s_B^2} = \frac{0.74}{0.46} = 1.62$$



$\nu_A = n_A - 1$=9
$\nu_B = n_B - 1$=9
$F_{1-\alpha/2,\nu_A,\nu_B}$=4.03

$F_{1-\alpha/2,\nu_A,\nu_B}$supera il valore calcolato F e l'ipotesi nulla (nessuna differenza) è accettata. Si può concludere con una confidenza del 95% che non c'è una differenza significativa nella precisione tra il lavoro dell'analista A e quello dell’analista B.

F-Test sulla varianza

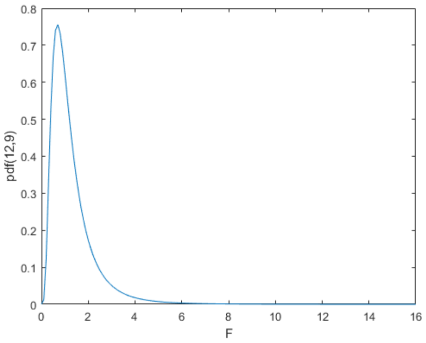
Esempio: *one-sided test*

La determinazione del contenuto di carbonato di calcio con il metodo standard di Scheibler è confrontata con il metodo semplice e più rapido di "neutralizzazione dell'acido" usando uno stesso campione. I risultati sono riportati nella tabella. A causa della natura del metodo rapido sospettiamo che produca una precisione inferiore a quella ottenuta con il metodo Scheibler e possiamo, quindi, eseguire il test F unilaterale.

A	B
2.5	1.7
2.4	1.9
2.5	2.3
2.6	2.3
2.5	2.8
2.5	2.5
2.4	1.6
2.6	1.9
2.7	2.6
2.4	1.7
	2.4
	2.2
	2.6

$$H_0: s_A^2 = s_B^2$$
$$H_1: s_B^2 > s_A^2 \qquad F > F_{1-\alpha,\nu_B,\nu_A}$$

$$F = \frac{s_B^2}{s_A^2} = \frac{0.16}{0.01} = 15.75$$



$\nu_A = n_A - 1$=9
$\nu_B = n_B - 1$=12
$F_{1-\alpha,\nu_B,\nu_A}$=3.07

$F > F_{1-\alpha,\nu_B,\nu_A}$ e l'ipotesi nulla (nessuna differenza) è rifiutata. Si può concludere (con una confidenza del 95%) che per questo unico campione la precisione del metodo di titolazione rapida è significativamente peggiore di quella del metodo Scheibler.

Test di omogeneità sulle medie per dati accoppiati (T-Test)

Il test t di Student può essere applicato anche per verificare un cambiamento fra una prima ed una seconda misura, effettuata sullo stesso individuo o sullo stesso campione, con diversi strumenti di misura o ripetuta a distanza di tempo. Nel test t di Student per dati appaiati la variabile in esame è rappresentata dalla differenza fra le coppie di osservazioni.

Consideriamo un numero n di coppie di dati $\mathbf{x}$ ottenuti alla prima determinazione, ed altrettanti dati $\boldsymbol{y}$ ottenuti alla seconda determinazione.

$$H_0: \mu_x = \mu_y$$

Sia d_i la differenza (presa con il segno) fra il valore del primo e il valore del secondo elemento della coppia, ovvero $d_i = x_i - y_i, i = 1, \dots, n$.

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \qquad s_d = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1}} \qquad s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

Test di omogeneità sulle medie per dati accoppiati (T-Test)

Null hypothesis $H_0: \bar{d} = \Delta$ $\Delta = \mu_x - \mu_y$

$$t = \frac{\bar{d} - \Delta}{s_{\bar{d}}}$$

Alternative hypothesis $H_1: \bar{d} > \Delta$

$\bar{d} > \Delta + t_{1-\alpha,\nu} s_{\bar{d}}$ ovvero $t > t_{1-\alpha,\nu}$

Alternative hypothesis $H_1: \bar{d} < \Delta$

$\bar{d} < \Delta - t_{1-\alpha,\nu} s_{\bar{d}}$ ovvero $t < t_{\alpha,\nu}$

Alternative hypothesis $H_1: \bar{d} \neq \Delta$

$\bar{d} < \Delta - t_{\alpha/2,\nu} s_{\bar{d}}$ o $\bar{d} > \Delta + t_{1-\alpha/2,\nu} s_{\bar{d}}$ ovvero $t < t_{\alpha/2,\nu}$ o $t > t_{1-\alpha/2,\nu}$

$\nu = n - 1$

NOTA: Se i dati sono numerosi si può utilizzare lo Z-Test, dove di solito si considerano numerosi i campioni con $n \geq 30$ (ma non è una regola fissa).

Test di omogeneità sulle medie di due campioni indipendenti (Student's T-Test)

Restituisce una decisione per il test dell'ipotesi nulla che gli $n_a$dati in $\mathbf{x}_a$ e gli n_b dati in $\mathbf{x}_b$ abbiano la stessa media (**confronto tra le medie di due campioni tra loro indipendenti**, se la varianza dei due gruppi è omogenea).

Null hypothesis H$_0$: $\bar{x}_a$=$\bar{x}_b$

$$t = \frac{|\bar{x}_a - \bar{x}_b|}{s} \qquad s = s_p\sqrt{\frac{1}{n_a} + \frac{1}{n_b}} \qquad s_p = \sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2}}$$

s_p("pooled" standard deviation) è la deviazione standard media delle deviazioni standard dei due campioni, cioè la radice quadrata della varianza che si ottiene sommando le devianze dei due campioni e dividendo per la somma dei gradi di libertà ($\nu = n_a + n_b - 2$).

Alternative hypothesis H$_1$: $\bar{x}_a > \bar{x}_b$ $\quad \bar{x}_a > \bar{x}_b + t_{1-\alpha,\nu}s$

Alternative hypothesis H$_1$: $\bar{x}_a < \bar{x}_b$ $\quad \bar{x}_a < \bar{x}_b - t_{1-\alpha,\nu}s$

Alternative hypothesis H$_1$: $\bar{x}_a \neq \bar{x}_b$ $\quad \bar{x}_a < \bar{x}_b - t_{\alpha/2,\nu}s$ o $\mu_a > \mu_b + t_{1-\alpha/2,\nu}s$

Test di omogeneità sulle medie di due campioni indipendenti (Cochran's T-Test)

Restituisce una decisione per il test dell'ipotesi nulla che gli $n_a$dati in $\mathbf{x}_a$ e gli n_b dati in $\mathbf{x}_b$ abbiano la stessa media (**confronto tra le medie di due campioni tra loro indipendenti**, se la varianza dei due gruppi non è omogenea).

Null hypothesis H$_0$: $\bar{x}_a = \bar{x}_b$

$$t = \frac{|\bar{x}_a - \bar{x}_b|}{s} \qquad s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

Alternative hypothesis H$_1$: $\bar{x}_a > \bar{x}_b$ $\quad \bar{x}_a > \bar{x}_b + t_{1-\alpha,\nu}\, s$
Alternative hypothesis H$_1$: $\bar{x}_a < \bar{x}_b$ $\quad \bar{x}_a < \bar{x}_b - t_{\alpha,\nu} s$
Alternative hypothesis H$_1$: $\bar{x}_a \neq \bar{x}$ $\quad \bar{x}_a < \bar{x}_b - t_{\alpha/2,\nu} s$ o $\bar{x}_a > \bar{x}_b + t_{1-\alpha/2,\nu} s$

$$\nu = round\left[\frac{\left(s_a^2/n_a + s_b^2/n_b\right)^2}{(s_a^2/n_a)^2/(n_a - 1\,) + (s_b^2/n_b)^2/(n_b - 1\,)}\right]$$

Test di omogeneità sulle medie di due campioni indipendenti (T-Test n_a, $n_b \geq 30$)

Restituisce una decisione per il test dell'ipotesi nulla che gli $n_a$dati in $\mathbf{x}_a$ e gli n_b dati in $\mathbf{x}_b$ abbiano la stessa media (**confronto tra le medie di due campioni tra loro indipendenti**, se $n_a, n_b \geq 30$).

Null hypothesis H$_0$: $\bar{x}_a$=$\bar{x}_b$

$$t = \frac{|\bar{x}_a - \bar{x}_b|}{s} \qquad s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

Alternative hypothesis H$_1$: $\bar{x}_a > \bar{x}_b$ $\quad \bar{x}_a > \bar{x}_b + t_{1-\alpha,\nu} s$

Alternative hypothesis H$_1$: $\bar{x}_a < \bar{x}_b$ $\quad \bar{x}_a < \bar{x}_b - t_{1-\alpha,\nu} s$

Alternative hypothesis H$_1$: $\bar{x}_a \neq \bar{x}_b$ $\quad \bar{x}_a < \bar{x}_b - t_{\alpha/2,\nu} s$ o $\bar{x}_a > \bar{x}_b + t_{1-\alpha/2,\nu} s$

$$\nu = n_a + n_b - 2$$

Analisi della varianza (ANOVA)

Spesso si presenta la necessità di prendere in considerazione esperimenti od osservazioni relative **a più di due gruppi** individuati sulla base di un fattore di interesse.

L'**analisi della varianza** è una tecnica che consente di confrontare le medie di più di due gruppi (popolazioni):

- quando i gruppi sono definiti sulla base di un singolo fattore si parla di **analisi della varianza** a un fattore o **a una via** (*one way Anova*)

- nell'**analisi della varianza a più vie** si studiano più fattori che possono creare delle differenze tra le medie

E' una **estensione a più gruppi del test t-Student.**

Analisi della varianza (ANOVA)

$SST = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}\right)^2$ DEVIANZA TOTALE

$SSA = \sum_{i=1}^{p} n_i (\bar{y}_i - \bar{y})^2$ DEVIANZA TRA GRUPPI

$SSE = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i\right)^2$ DEVIANZA ALL'INTERNO DEI GRUPPI

La variabilità all'interno dei p gruppi (SSE) è considerata un errore casuale, mentre la variabilità tra i p gruppi (SSA) è attribuibile alle differenza tra i gruppi, ed è anche chiamata *effetto del trattamento:*

SST = SSA + SSE

In questo contesto l'ipotesi nulla che si è interessati a verificare è che **le medie di tutti gruppi siano uguali tra loro**, contro l'ipotesi alternativa che almeno una sia diversa.

Analisi della varianza (ANOVA)

A questo livello non è però possibile fare confronti, perché le devianze hanno un **numero di addendi diverso**.

Ad ognuna delle devianze sono associati i seguenti gradi di libertà:

- la devianza **totale** ha $n_t - 1$ gradi di libertà
- la devianza **tra gruppi** ha $p - 1$ gradi di libertà
- la devianza **all'interno dei gruppi** ha $n_t - p$ gradi di libertà

Dividendo ciascuna devianza per i rispettivi gradi di libertà si ottengono le **VARIANZE,** cioè le medie dei quadrati:

Varianza tra i gruppi

$$MSA = \frac{SSA}{p-1}$$

Varianza all'interno dei gruppi

$$MSE = \frac{SSE}{n_t - p}$$

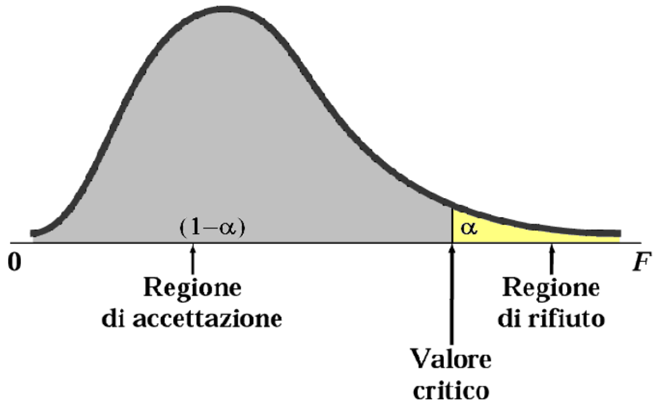
Analisi della varianza (ANOVA)

Per verificare l'ipotesi di uguaglianza delle medie utilizzo il **test F di Fischer** che confronta MSA e MSE:

$$F = \frac{MSA}{MSE}$$

Il valore critico della F viene determinato in funzione del livello di significatività α del test.

valori critici si individuano nelle tavole della distribuzione F in base ai gradi di libertà e al livello di significatività scelto



Analisi della varianza (ANOVA)

Esempio: One way ANOVA

In un esperimento sull'inibizione della crescita dei tumori nel topo mediante trattamenti con due preparati (A e B) si sono ottenuti i pesi dei tumori (centigrammi) dopo una settimana dal trapianto in tre gruppi di 7, 4, 5 topi tenuti rispettivamente come controllo, sotto trattamento A e sotto trattamento B.

Controllo	A	B
52	22	31
30	25	28
27	17	43
42	32	40
25		36
38		
29		

Sulla base dei risultati ottenuti, mostrati in tabella, si valuti l'efficacia dei trattamenti utilizzando l'analisi della varianza ad una via.

Analisi della varianza (ANOVA)

Esempio: One way ANOVA

$$p = 3 \quad n_c = 7 \quad n_A = 4 \quad n_B = 5$$

$$\bar{y} = 32.3 \quad \bar{y}_c = 34.7 \quad \bar{y}_A = 24.0 \quad \bar{y}_B = 35.6$$

$$s_c^2 = 95.2 \quad s_A^2 = 39.3 \quad s_B^2 = 38.3$$

$$SSA = \sum_{i=1}^{p} n_i(\bar{y}_i - \bar{y})^2 = n_c(\bar{y}_c - \bar{y})^2 + n_A(\bar{y}_A - \bar{y})^2 + n_B(\bar{y}_B - \bar{y})^2 = 370.8$$

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 = \sum_{i=1}^{p}(n_i - 1)s_i^2 = 842.6$$

$$p - 1 = 2 \qquad n_c + n_A + n_B - p = 13$$

$$MSA = \frac{SSA}{p-1} = 185.4$$

$$MSE = \frac{SSE}{n_c + n_A + n_B - p} = 64.8$$

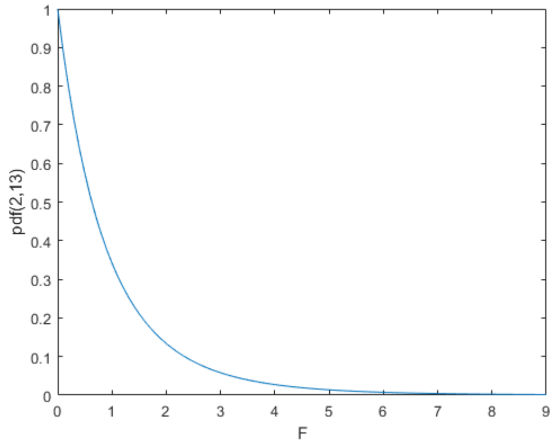
Analisi della varianza (ANOVA)

Esempio: One way ANOVA

$$F = \frac{MSA}{MSE} = 2.86$$

$$F_{0.975,2,13} = 4.97$$

$$p = \int_{2.86}^{\infty} f_F(F|2,13)dF = 0.0935 > 0.05$$



Non c'è evidenza per rifiutare l'ipotesi nulla e affermare che c'è differenza tra i tre gruppi, con un livello di significatività di 0.05

Nel caso in cui si ottenga p < 0.05 vuol dire che **almeno un gruppo** si differenzia dagli altri.

Analisi della varianza (ANOVA)

Se si vuole sapere **quanti e quali** siano diversi si possono fare **test *t* multipli** fra le varie coppie di campioni.

Per tenere conto del fatto che si stanno facendo confronti multipli, si applicheranno opportune correzioni al livello di significatività.

Un semplice metodo, ad hoc, consiste nell'applicare la ***correzione di Bonferroni***. Tale correzione viene, in generale, utilizzata quando è necessario eseguire molti test di ipotesi con lo stesso database. Infatti, eseguendo un gran numero di test, ciascuno con livello di significatività α=0.05, alcuni test forniranno un risultato positivo anche in assenza di qualunque effetto reale.

L'idea alla base della correzione di Bonferroni è che, se si esegue un numero di test di significatività pari a c, per ottenere un livello complessivo di errore di Tipo I pari ad α, si deve semplicemente dichiarare che ciascun **test sarà significativo** se il valore ottenuto di p sarà minore di α/c.

Esempio

Avendo prefissato α =0.05, se si vogliono verificare 5 ipotesi in un solo esperimento (diciamo 5 diversi trattamenti contro un controllo), non si accetterà un risultato come significativo se il valore di p nei vari test non è minore di 0.01.

Chi-Square Goodness-of-Fit Test

Restituisce una decisione per il test dell'ipotesi nulla che i dati in $\mathbf{x}$ provengano da una distribuzione normale.

Null hypothesis $H_0: p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$$\chi^2 = \sum_{j=1}^{m} \frac{\left(n_j - np_j\right)^2}{np_j}$$

Alternative hypothesis $H_1: p(x) \neq \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $\quad \chi^2 > \chi^2_{1-\alpha,\nu}$

$\nu = m - 3$

Chi-Square Goodness-of-Fit Test

Esempio: Si prende un campione di 100 torte prodotte in una panetteria e si misura la massa di ogni torta (in grammi). Applica il test del chi-quadro per esaminare se l'insieme di dati formato dall'insieme di 100 misurazioni di massa mostrato qui sotto è conforme a una distribuzione gaussiana:

487	504	501	515	491	496	482	502	508	494	505	501	485	503	507	494	489	501	510	491
503	492	483	501	500	493	505	501	517	500	494	503	500	488	496	500	519	499	495	490
503	500	497	492	510	506	497	499	489	506	502	484	495	498	502	496	512	504	490	497
488	503	512	497	480	509	496	513	499	502	487	499	505	493	498	508	492	498	486	511
499	504	495	500	484	513	509	497	505	510	516	499	495	507	498	514	506	500	508	494

Applicando la regola di Sturgis, il numero raccomandato di bin m per n punti è dato da:

$$m = 1 + 3.3log_{10}(n) = 7.6 \rightarrow 8$$

Le misure di massa coprono l'intervallo 480-519. Quindi, sceglieremo larghezze dei bin di dati di 5 grammi, con i confini dei bin fissati a 479.5, 484.5, 489.5, 494.5, 499.5, 504.5, 509.5, 514.5, e 519.5 (i confini sono imposti in modo che non ci sia ambiguità in quale bin ogni particolare valore dei dati stia). Il passo successivo comporta il conteggio del numero di misurazioni in ciascun intervallo. Questi sono i valori $n_j, j = 1, \dots, 8$. I risultati di questo conteggio sono esposti nella tabella seguente:

j	1	2	3	4	5	6	7	8
Data range	479.5-484.5	484.5-489.5	489.5-494.5	494.5-499.5	499.5-504.5	504.5-509.5	509.5-514.5	514.5-519.5
Measurements in range (n_j)	5	8	13	23	24	14	9	4

Chi-Square Goodness-of-Fit Test

Esempio:

n_j ≥4 per tutti j, quindi possiamo ora procedere al calcolo dei valori n_j'. Questi sono i numeri attesi di misurazioni in ogni intervallo di dati per una distribuzione gaussiana. Il punto di partenza per questo calcolo è conoscere il valore medio e la deviazione standard delle 100 misurazioni di massa:

$\bar{x}$=499.53 s=8.389

j	$x_{j,inf}$	$x_{j,sup}$	n_j	$z_{j,inf} = (x_{j,inf} - \bar{x})/s$	$z_{j,sup} = (x_{j,sup} - \bar{x})/s$	$normcdf(z_{j,inf})$	$normcdf(z_{j,sup})$	p_j	n_j'	$\frac{(n_j - n_j')^2}{n_j'}$
1	479.5	484.5	5	-2.388	-1.792	0.000	0.037	0.0366	3.66	0.4912
2	484.5	489.5	8	-1.792	-1.196	0.037	0.116	0.0793	7.93	0.0006
3	489.5	494.5	13	-1.196	-0.600	0.116	0.274	0.1585	15.85	0.5113
4	494.5	499.5	23	-0.600	-0.004	0.274	0.499	0.2242	22.42	0.0151
5	499.5	504.5	24	-0.004	0.592	0.499	0.723	0.2247	22.47	0.1048
6	504.5	509.5	14	0.592	1.188	0.723	0.883	0.1595	15.95	0.2373
7	509.5	514.5	9	1.188	1.785	0.883	0.963	0.0802	8.02	0.1210
8	514.5	519.5	4	1.785	2.381	0.963	1.000	0.0372	3.72	0.0215

$\chi^2 = 1.5028$

Chi-Square Goodness-of-Fit Test

Esempio:

Il passo finale è controllare se questo valore di χ^2 è maggiore di quello che ci si aspetterebbe per una distribuzione gaussiana. Questo comporta la ricerca di χ^2 nella tabella sottostante. Prima di fare questo, dobbiamo specificare il numero di gradi di libertà, $\nu = m - 3$.

Degree of Freedom	Probability of Exceeding the Critical Value								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
	Not Significant							*Significant*	

Per un livello di significatività α = 0.05, $\chi^2 \leq \chi^2_{1-\alpha,5}$=11.07 e $p = \int_{1.503}^{\infty} f_{\chi^2}(t|\nu = 5)dt$= 0.913 > 0.05 . Non ci sono prove sufficienti, al livello di significatività α = 0.05, per rifiutare l'ipotesi nulla e concludere che l'insieme di dati non è conforme a una distribuzione gaussiana.

Test di Kolmogorov-Smirnov (buon adattamento)

Il test di Kolmogorov-Smirnov è un test di buon adattamento:

$$H_0: X \sim F_0 \qquad H_1: X \nsim F_0$$

che si esegue partendo da un campione del tipo:

$$X_1, \dots, X_n \; i.i.d. \sim F$$

La distribuzione indicata nell'ipotesi nulla F_0 deve essere una distribuzione di probabilità continua completamente specificata.

Funzione di ripartizione empirica

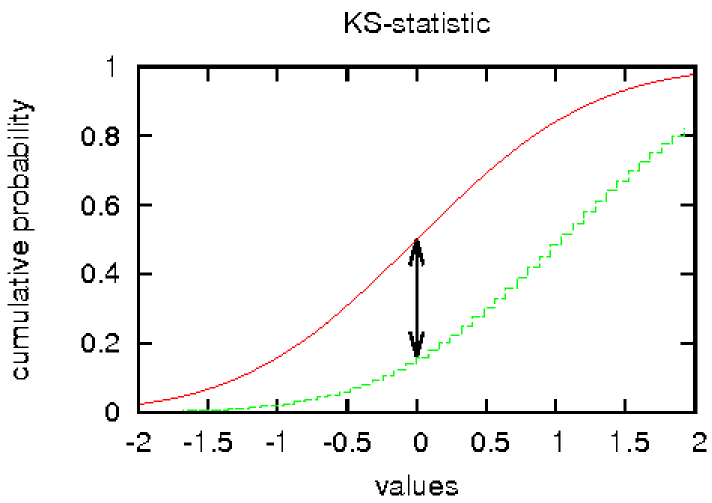
$$\hat{F}_n(x) = \begin{cases} 0 & se\ x \leq X_1 \\ \frac{k}{n} & se\ X_k \leq x \leq X_{k+1} \\ 1 & se\ x \geq X_n \end{cases}$$

o in forma più compatta

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x}$$

con $I_{X_i \leq x}$ funzione indicatrice

Test di Kolmogorov-Smirnov (buon adattamento)



Test di Kolmogorov-Smirnov (buon adattamento)

Ricordiamo che:

distribuzione binomiale

$$Bin(n,p) = f_b(k|n,p) = \begin{cases} \binom{n}{k} p^k q^{n-k} & per \;\; k = 0, \dots, n \\ 0 & altrimenti \end{cases}$$

$$\mu = np = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}$$

$$\sigma^2 = npq$$

Test di Kolmogorov-Smirnov (buon adattamento)

Siccome la variabile aleatoria: $I_{X_i \leq x}$ è una variabile binomiale con distribuzione binomiale:

$$Bin\big(n, F_n(x)\big)$$

abbiamo:

$$E\big[\hat{F}_n(x)\big] = E\left[\frac{Bin\big(n, F_n(x)\big)}{n}\right] = \frac{nF_n(x)}{n} = F_n(x)$$

$$Var\big[\hat{F}_n(x)\big] = Var\left[\frac{Bin\big(n, F_n(x)\big)}{n}\right] = \frac{nF_n(x)\big(1 - F_n(x)\big)}{n^2} = \frac{F_n(x)\big(1 - F_n(x)\big)}{n}$$

Possiamo affermare che $\hat{F}_n(x)$ è uno stimatore puntualmente non distorto di $F_n(x)$ e, poiché la varianza tende a 0 per $n \to \infty$, si ha di conseguenza anche la proprietà di consistenza in media quadratica.

$\hat{F}_n(x)$ è una stima campionaria della vera funzione di ripartizione $F_n(x)$

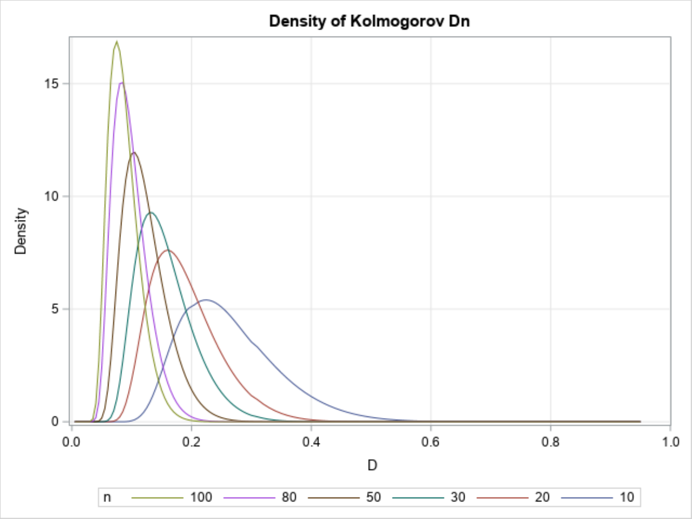
Test di Kolmogorov-Smirnov (buon adattamento)

Se $\hat{F}_n(x)$ e $F_0(x)$ sono *vicine*, cioè *sufficientemente simili*, non si rifiuta l'ipotesi nulla.

Se $\hat{F}_n(x)$ e $F_0(x)$ sono *lontane*, cioè se *molto dissimili*, si rifiuta l'ipotesi nulla.

$$D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_0(x) \right|$$

Rifiuta H_0 se $D_n > q_{KS,1-\alpha}$



	Livello di significatività (α)				
n	**0,01**	**0,05**	**0,10**	**0,15**	**0,20**
2	0,92930	0,84189	0,77639	0,72614	0,68377
3	0,82900	0,70760	0,63604	0,59582	0,56481
4	0,73421	0,62394	0,56522	0,52476	0,49265
5	0,66855	0,56327	0,50945	0,47439	0,44697
6	0,61660	0,51926	0,46799	0,43526	0,41035
7	0,57580	0,48343	0,43607	0,40497	0,38145
8	0,54180	0,45427	0,40962	0,38062	0,35828
9	0,51330	0,43001	0,38746	0,36006	0,33907
10	0,48895	0,40925	0,36866	0,34250	0,32257
11	0,46770	0,39122	0,35242	0,32734	0,30826
12	0,44905	0,37543	0,33815	0,31408	0,29573
13	0,43246	0,36143	0,32548	0,30233	0,28466
14	0,41760	0,34890	0,31417	0,29181	0,27477
15	0,40420	0,33760	0,30397	0,28233	0,26585
16	0,39200	0,32733	0,29471	0,27372	0,25774
17	0,38085	0,31796	0,28627	0,26587	0,25035
18	0,37063	0,30936	0,27851	0,25867	0,24356
19	0,36116	0,30142	0,27135	0,25202	0,23731
20	0,35240	0,29407	0,26473	0,24587	0,23152
25	0,31656	0,26404	0,23767	0,22074	0,20786
30	0,28988	0,24170	0,21756	0,20207	0,19029
35	0,26898	0,22424	0,20184	0,18748	0,17655

Test di Kolmogorov-Smirnov (buon adattamento)

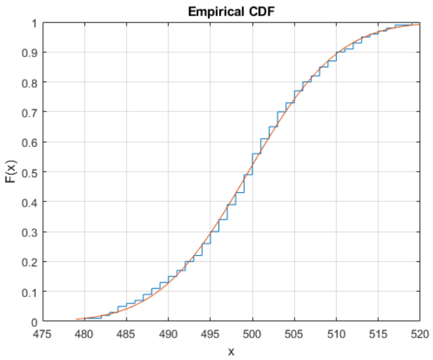
Esempio: Si prende un campione di 100 torte prodotte in una panetteria e si misura la massa di ogni torta (in grammi). Applica il test del chi-quadro per esaminare se l'insieme di dati formato dall'insieme di 100 misurazioni di massa mostrato qui sotto è conforme a una distribuzione gaussiana:

487	504	501	515	491	496	482	502	508	494	505	501	485	503	507	494	489	501	510	491
503	492	483	501	500	493	505	501	517	500	494	503	500	488	496	500	519	499	495	490
503	500	497	492	510	506	497	499	489	506	502	484	495	498	502	496	512	504	490	497
488	503	512	497	480	509	496	513	499	502	487	499	505	493	498	508	492	498	486	511
499	504	495	500	484	513	509	497	505	510	516	499	495	507	498	514	506	500	508	494

Il valore medio e la deviazione standard delle 100 misurazioni di massa sono:
$\bar{x}$=499.5 s=8.4

$$D_n = \sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_0(x)\right| = 0.0448 \quad p = 0.9827$$

$q_{KS,1-0.05}$=0.1340



Non ci sono prove sufficienti, al livello di significatività $\alpha = 0.05$, per rifiutare l'ipotesi nulla e concludere che l'insieme di dati non è conforme a una distribuzione gaussiana.

Test di Kolmogorov-Smirnov (buon adattamento)

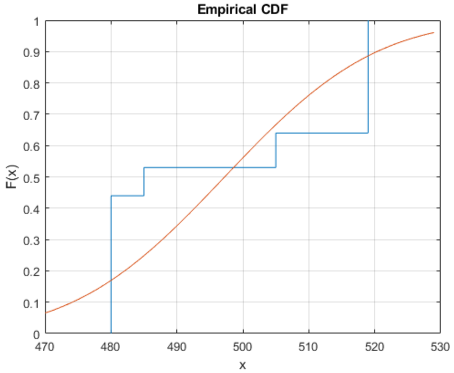
Esempio: Si prende un campione di 100 torte prodotte in una panetteria e si misura la massa di ogni torta (in grammi). Applica il test del chi-quadro per esaminare se l'insieme di dati formato dall'insieme di 100 misurazioni di massa mostrato qui sotto è conforme a una distribuzione gaussiana:

480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480
480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480	480
480	480	480	480	485	485	485	485	485	485	485	485	485	505	505	505	505	505	505	505
505	505	505	505	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519
519	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519	519

Il valore medio e la deviazione standard delle 100 misurazioni di massa sono:
$\bar{x}$=497.2 s=18.0

$D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_0(x) \right|$=0.2816 p=2e-7

$q_{KS,1-0.05}$=0.1340



Ci sono prove sufficienti, al livello di significatività α = 0.05, per rifiutare l'ipotesi nulla e concludere che l'insieme di dati non è conforme a una distribuzione gaussiana.

Test sui dati accoppiati – test di indipendenza (dati gaussiani)

coefficiente di correlazione lineare $\boldsymbol{\rho}$ tra due variabili aleatorie X e Y:

$$\boxed{\rho = \frac{cov(X,Y)}{\sqrt{\sigma_x^2 \sigma_y^2}}}$$

- $|\rho| \leq 1$
- $|\rho| = 1$, allora con probabilità 1 le due variabili sono una la trasformazione lineare dell'altra: $Y = aX + 1$
- $\rho = 0$, allora le due variabili X e Y sono dette *scorrelate*
- Se X e Y sono indipendenti, allora: $E(X \cdot Y) = E(X) \cdot E(Y) \; \rightarrow \; \rho = 0$ Non è però vero il viceversa.

Test sui dati accoppiati – test di indipendenza (dati gaussiani)

Campione gaussiano bidimensionale

Un campione accoppiato gaussiano bidimensionale è un campione del tipo:

$$(X_1, Y_1) \dots (X_n, Y_n)\ i.i.d. \sim N\left(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho\right)$$

$$f_{XY}(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma_y}\right)\right)}$$

Se $\rho = 0$

$$f_{XY}(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)} = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

Si nota quindi che, nel caso in cui si abbia un campione congiuntamente gaussiano (ma non nel caso generale): X e Y *sono indipendenti se e solo se* $\rho = 0$.

Di conseguenza, per eseguire un test di indipendenza tra due variabili aleatorie gaussiane è sufficiente eseguire un test d'indipendenza sul loro coefficiente di correlazione.

Test sui dati accoppiati – test di indipendenza (dati gaussiani)

Null hypothesis $\rho = 0$

Stimiamo ρ: $\hat{\rho} = R = \frac{\widehat{cov}(X,Y)}{\sqrt{s_x^2 s_y^2}} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$

$$ST = \frac{R}{\sqrt{1 - R^2}}\sqrt{n-2} \sim t_{n-2}, \qquad n \geq 3$$

Alternative hypothesis $H_1: \rho > 0$ $\quad ST \geq t_{1-\alpha,n-2}$

Alternative hypothesis $H_1: \rho < 0$ $\quad ST \leq t_{\alpha,n-2}$

Alternative hypothesis $H_1: \rho \neq 0$ $\quad |ST| \geq t_{1-\alpha/2,n-2}$

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di

Ingegneria e Architettura