

TUTORIAL PER SVOLGERE ANALISI STATISTICHE PER LE TESI DI LAUREA

- 1) STUDIARE/RIPASSARE UN PO' DI TERMINOLOGIA DI STATISTICA
- 2) PREPARARE IL DATABASE IN EXCEL
- 3) INSTALLARE IL SOFTWARE GRATUITO "JAMOVİ"
(<https://www.jamovi.org/download.html>)
- 4) IMPORTARE IL DATABASE DA EXCEL SU JAMOVİ
- 5) EFFETTUATE LE ANALISI

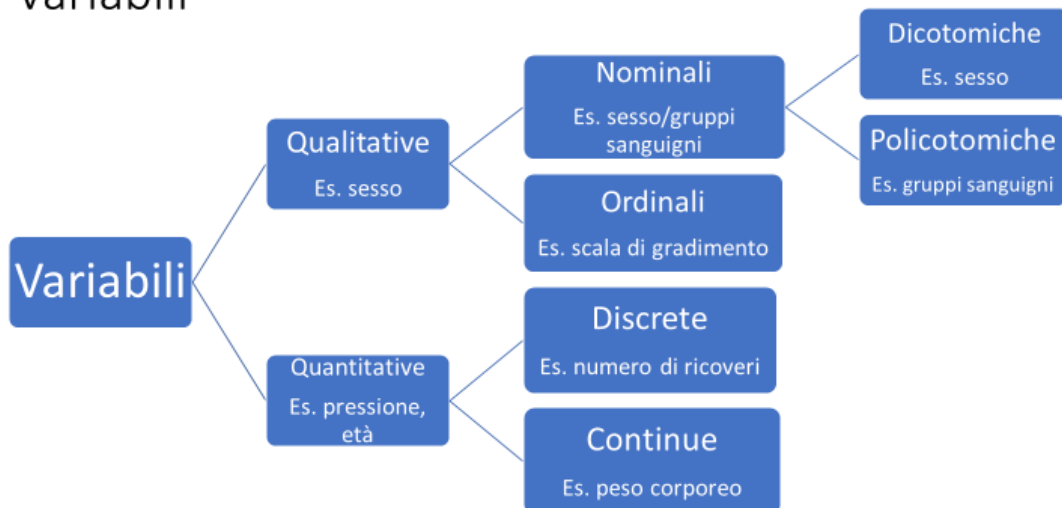
1) RIPASSO DEI CONCETTI PRINCIPALI DI STATISTICA

Definizioni

	Sesso	Età
Maria	F	34
Filippo	M	23
Giulia	F	19

- La tabella complessiva si chiama **database** (raccolta ben strutturata di informazioni)
- Maria, Filippo, Giulia sono **unità statistiche** (tutte le informazioni relative ad una specifica unità statistica sono riportate sulla stessa riga)
- Sesso e età sono **variabili**
- F/M,34,23,19 sono **dati o osservazioni** (realizzazione della variabile, ovvero il valore che assume la variabile su quella specifica unità statistica).
- Nel caso delle variabili che esprimono qualità come il sesso, i «valori» che può assumere la variabile sono detti **modalità o categorie**

Variabili



ESEMPI DI VARIABILI:

Variabili quantitative

Continue:

Peso: 65,7kg 80,2kg 56,1kg ...

Diametro: 2,3cm 1,7cm 2,2cm..

Altezza: 170cm 165cm 183cm..

Discrete:

Accessi al PS: 45 72 121 3 ...

Numero di pz visitati al mese con
sospetto melanoma: 2 45 31 12..

Variabili qualitative

Nominali policotomiche:

Gruppi sanguigni: A, B, AB, 0
Fumo: Si, No, Ex fumatore
Regioni: Toscana, Lombardia, Sicilia,..

Nominali dicotomiche:

Sesso: Maschio, Femmina
Fumo: Si, No
Istologia: nodulo benigno, tumore

Ordinali:

*Qualità ordinabili in modo crescente o
descrescente*

BMI: sottopeso, normopeso,
sovrappeso, obeso
Esposizione al sole: scarsa, media, alta

Score-punteggi-scale

Es.

Scala NRS: 0,1,2,3,4,5,6,7,8,9,10

Indice di Apgar: 0,1,2,3,4,5,6,7,8,9,10

Sono variabili ibride tra variabili ordinali e quantitative

Se con molti valori vengono analizzati come una variabile quantitativa (ma non possono essere applicate tutte le tecniche)

Se i valori sono pochi, per esempio 3, vengono maggiormente assimilate a una variabile ordinale

Frequenze

- **Assolute:** numero di volte in cui si presenta una modalità della variabile
- **Relative:** vengono calcolate come rapporto tra le frequenze assolute e la numerosità campionaria.
- **Percentuali:** vengono calcolate come prodotto delle frequenze relative per 100

Tabelle di contingenza

		Outcome/variabile dipendente		Totale di riga
		Mod1	Mod2	
Exposure /variabile indipendente	Mod1			R_1
	Mod2			R_2
	Totale di colonna	C_1	C_2	Totale (n)

Modalità o categorie della variabile

- Frequenze assolute
- Frequenze relative
- Frequenze percentuali

Numerosità campionaria

TEST DI IPOTESI

Ipotesi di associazione

- Coinvolgono due variabili di tipo qualitativo
- La tecnica descrittiva adeguata per sintetizzare i dati è la tabella di contingenza
- La tecnica grafica più adatta è il diagramma a barre

Test per la differenza tra medie

Test parametrici

Test non parametrici

Test t di Student per dati non
appaiati



Test di Mann-Whitney

Test t di Student per dati appaiati



Test di Wilcoxon

Test per la differenza tra medie

Test parametrici

Test non parametrici

Test ANOVA per dati non
appaiati



Test di Kruskal-Wallis

Test ANOVA per dati appaiati

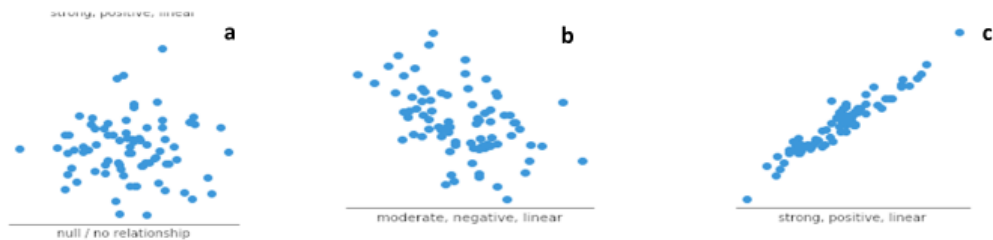


Test di Friedmann

Ipotesi di correlazione

- Coinvolge due variabili di tipo quantitativo
- Una ha funzione di exposure e una di outcome
- La tecnica descrittiva che viene utilizzata è lo scatterplot

Ipotesi di correlazione



- a) Non c'è correlazione, i punti sono una «nuvola casuale»
- b) C'è una correlazione media, intravediamo un andamento e i punti iniziano ad avvicinarsi
- c) C'è una correlazione elevata (quasi perfetta), i punti sono molto vicini ad una linea di tendenza centrale

Ipotesi di correlazione

Coefficienti di correlazione lineare r

Di Pearson

Di Spearman

Di Kendall

- I 3 coefficienti assumono valori tra -1 e 1
- Più il coefficiente si avvicina agli estremi più la correlazione è forte
- A 0 non c'è correlazione
- Per valori compresi tra -1 e 0 la correlazione è negativa
- Per valori compresi tra 0 e 1 la correlazione è positiva

2) PREPARARE IL DATABASE IN EXCEL

Il database deve avere questa forma:

id	Nome Variabile 1	Nome Variabile 2	Nome Variabile 3
1				
2				
...				

Quindi per esempio:

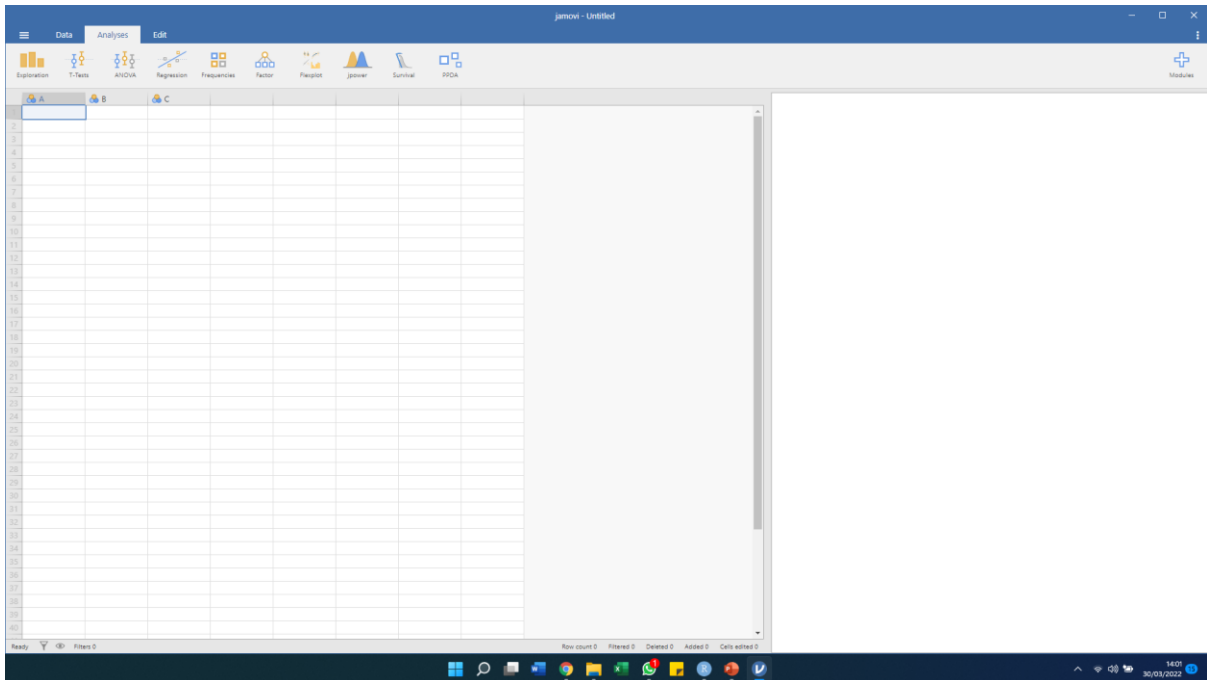
id	Sesso	Età	Età diagnosi	tumore
1	M	25	18	1
2	F	89	88	0
3	M	62	58	1

!!!!AVVERTENZE PER COSTRUIRE UN BUON DATABASE!!!!

- Prima riga nome delle variabili. **NON** ci **devono** essere celle unite.
- Dalla seconda riga si inseriscono i pazienti
- Colonne e righe **devono** essere contigue (per es. non ci deve essere una riga vuota che separa il nome delle variabili dal primo paziente/id)
- **Puoi** scegliere se inserire una dicitura tipo Si/No oppure 1/0
- Le categorie delle variabili qualitative **devono** essere scritte tutte allo stesso modo. Per esempio M, m, M. , m. vengono riconosciute come categorie differenti
- Se il dato è mancante la cella **deve** essere VUOTA (non ha lo stesso significato di «non presente»)
- Le stringhe non sono analizzabili (es. cellula dendritica regolare 2mm)
- Ogni cella deve contenere un unico dato (QUESTO è UN ERRORE -> volume massa: 2cm x 1cm x 1.5 cm)
- Tutte le osservazioni all'interno della stessa variabile **devono** avere la stessa unità di misura e non deve essere riportata all'interno della cella (ES. 5mg/die)
- I decimali **devono** avere tutti o il punto o la virgola in base alla vostra versione di excel

3) SCARICARE JAMOVİ (link nella prima pagina)

Questo è il software



In alto ci sono 4 sezioni:

- 1) Tre linee
- 2) Data
- 3) Analysis
- 4) Edit

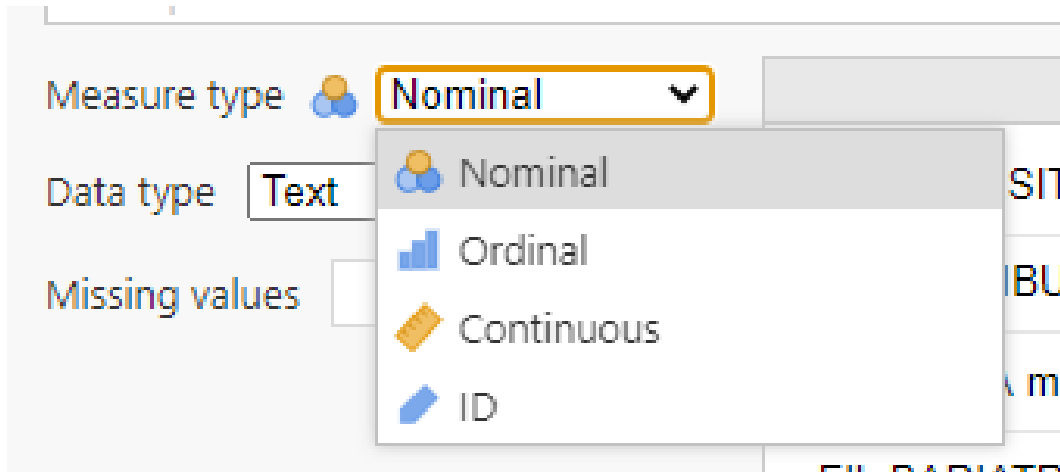
**QUI POTETE TROVARE UNA GUIDA DI JAMOVİ PER
L'IMPORT DEL DATABASE, TRASFORMAZIONE DEI DATI
E ANALISI STATISTICHE**

<https://docs.jamovi.org/>

4) IMPORTARE IL DATABASE

- 1) Cliccare sulle tre lineette in alto a sinistra
- 2) Cliccare su "import" e selezionare il database nella cartellina dove è stato salvato
- 3) CONTROLLARE CHE TUTTE LE VARIABILI SIANO STATE IMPORTATE NEL FORMATO CORRETTO

Questi sono i 3 possibili formati:



NOMINAL=qualitativa

CONTINUOUS=quantitativa

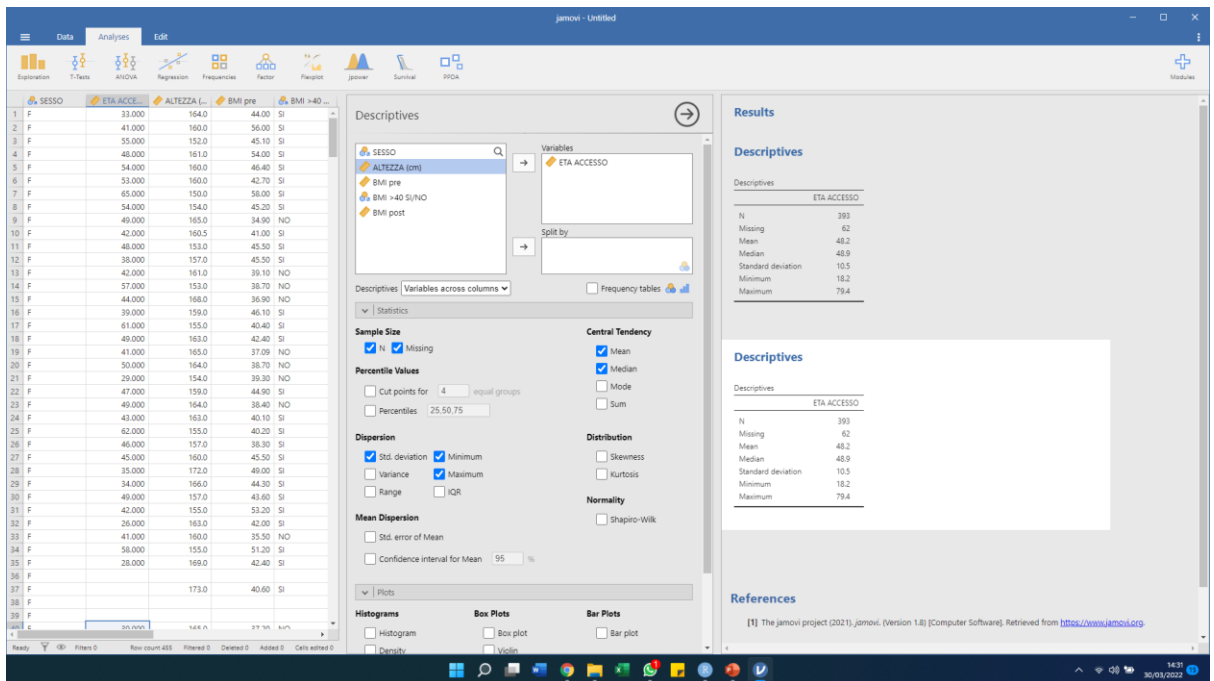
ATTENZIONE!! Se vi accorgete che una variabile quantitativa (ad esempio l'età) è stata importata come qualitativa **RICONTROLLA IL DATABASE** perché avrai messo punti al posto di virgole (o viceversa), simboli per indicare i valori mancanti o note da tenere a mente (ad esempio: - / | ? ! " '), oppure hai inserito SPAZI prima o dopo numeri

Dopo aver corretto il database dovete salvarlo e importarlo nuovamente su JAMOVI fino a quando non sarà tutto corretto.

5) EFFETTUATE LE ANALISI

- 1) **CAPIRE** L'OBIETTIVO DELLA VOSTRA RICERCA E NON PARTIRE ALLA CIECA A FARE COSE
- 2) EFFETTUARE UNA PRELIMINARE ANALISI DESCRITTIVA PER CAPIRE SE CI SONO VALORI ESTREMI O ERRORI

- ANDARE SU JAMOVI E CLICKARE ANALYSES -> EXPLORATION -> DESCRIPTIVES



CON I VARI FLAG POTETE DECIDERE QUALI DESCRITTIVE OTTENERE

- > PER VARIABILI QUANTITATIVE: media e deviazione standard
- > PER VARIABILI ORDINALI, SCORE O PUNTEGGI: mediana e IQR (interquartile range)
- > PER VARIABILI QUALITATIVE: frequenze

- 3) EFFETTUARE L'ANALISI INFERENZIALE:

CONFRONTO TRA:

VARIABILE 1	VARIABILE 2	Test	Percorso jamovi	Note
quantitativa	quantitativa	Correlazione di Pearson	-Regression -correlation matrix -correlation coefficient: Pearson	
ordinale/ score	ordinale/ score	Correlazione di Spearman	-Regression -correlation matrix -correlation coefficient: Spearman	
qualitativa	qualitativa	chi-quadrato	-frequencies Independent sample, chi-square for association	Per le tabelle 2x2 potete calcolare gli Odds Ratio (flag apposito)
qualitativa	Stessa var. Qualitativa misurata a distanza di tempo	McNemar	-frequencies Paired sample, MnNemar	
quantitativa	Qualitativa dicotomica	T-test o Mann-whitney	-T-test - independent sample t-test	1) Fare il test della normalità (assumption check->Normality test) 2) Se questo test ha un $p > 0.05$ fare test T di student altrimenti Mann-Whitney
quantitativa	Stessa var. Quantitativa misurata a distanza di	T-test o Wilcoxon	-T-test -Paired sample t-test	3) Fare il test della normalità (assumption check->Normality test) 4) Se questo test ha un $p > 0.05$ fare

	tempo			test T di student altrimenti Wilcoxon
quantitativa	Qualitativa policotomica	ANOVA	-ANOVA -One-way ANOVA	se l'assunzione di normalità non è rispettata passare al successivo test
Quantitativa o ordinale /score	Qualitativa policotomica	Kruskal-Wallis	-ANOVA -One-way ANOVA Kuskal Test	
quantitativa	Stessa var quantitativa misurata in più tempi	ANOVA per misure ripetute	-ANOVA -Repeated measure ANOVA	se l'assunzione di normalità non è rispettata passare al successivo test
quantitativa o ordinale/ score	Stessa var quantitativa misurata in più tempi	ANOVA per misure ripetute	-ANOVA -Repeated measure ANOVA, Friedmann Test	

Una volta selezionato il test adeguato, basta spostare le variabili coinvolte nella rispettiva cella.

Tutti i test danno anche le informazioni descrittive necessarie per comprendere il fenomeno **(basta scorrere verso il basso e aprire i vari menù a tendina)**

PER CAPIRE la conclusione del test guardare il p-value, se questo è:

- $p < 0.05$ allora il test è significativo, quindi l'associazione/differenza/correlazione che avete osservato è reale
- $p > 0.05$ allora il test è non significativo, quindi l'associazione/differenza/correlazione che avete osservato è frutto del caso.