jamovi (autore: Vittorio Albertoni)

Premessa

Il ricorso della statistica all'informatica risale all'elaborazione dei dati del censimento USA del 1890, per la prima volta effettuata con l'uso delle macchine tabulatrici di Herman Hollerith. L'elaborazione meccanica dei dati ridusse i tempi di conta e di classificazione ma non aveva nulla di paragonabile alla moderna analisi statistica.

Il primo pacchetto software di tipo «moderno» risale al 1961 e si chiamava BMDP (Bio MeDical Package) e il suo sviluppo è cessato nel 2017.

I giganti del settore, tuttora distribuiti e utilizzati, sono SPSS, il cui primo rilascio risale al 1968 e dal 2009 è stato acquisito e riprogrammato in Java da IBM e da questa attualmente distribuito, e SAS, il cui primo rilascio risale al 1971 ed è attualmente distribuito da SAS Institute.

Tutto ciò nel campo del software commerciale, rilasciato con licenza proprietaria.

Nel campo del software libero dobbiamo arrivare al 1996 per vedere la nascita del linguaggio R [1] , sulle ceneri del linguaggio S, sviluppato in casa Bell Labs dal 1976.

Il software presentato in questo manualetto si ispira e si pone come contraltare libero di SPSS e SAS e, programmato in C++ e Javascript, utilizza il linguaggio R per le elaborazioni statistiche.

La sua nascita si deve agli sviluppatori Jonathon Love, Damian Dropmann e Ravi Selker, già sviluppatori del progetto JASP presso l'Università di Amsterdam, che riguarda un software pure open source di cui jamovi è praticamente una copia.

I suoi sviluppatori raccomandano di scriverne il nome in lettere minuscole, nonostante si tratti dell'acronimo di Just Another Multivariate Open-source Visual Interface.

E' distribuito dal 2017 ed offre maggiore estensibilità tramite moduli, una gestione dati superiore e una più stretta integrazione con R rispetto al suo simile JASP, distribuito dal 2015.

jamovi è un ottimo strumento didattico e di analisi. Non serve per complessi modelli multilevel, modelli misti con strutture di covarianza particolari, analisi con specificazioni non standard o altamente personalizzate. In questi casi è meglio R puro o, meglio ancora, Python.

Altro importante limite è dato dal fatto che jamovi lavora totalmente in memoria RAM per cui non è adatto per computer con meno di 4 GB di RAM e per affrontare analisi su grandi masse di dati, meno che mai se pensiamo ai big data, campo in cui dobbiamo ricorrere a R ottimizzato o all'insuperabile Python.

[1] In proposito si veda il manualetto «r» allegato all'articolo «La più grande calcolatrice per la statistica» pubblicato sul mio blog all'indirizzo *www.vittal.it* nel novembre 2022.

Indice

1 Installazione **3**

2 Come funziona **3**

3 A cosa serve **4**
3.1 Variabili 4
3.2 Dati 5
3.3 Analisi 6
3.4 Grafici 10
3.5 Modifica 11

1 Installazione

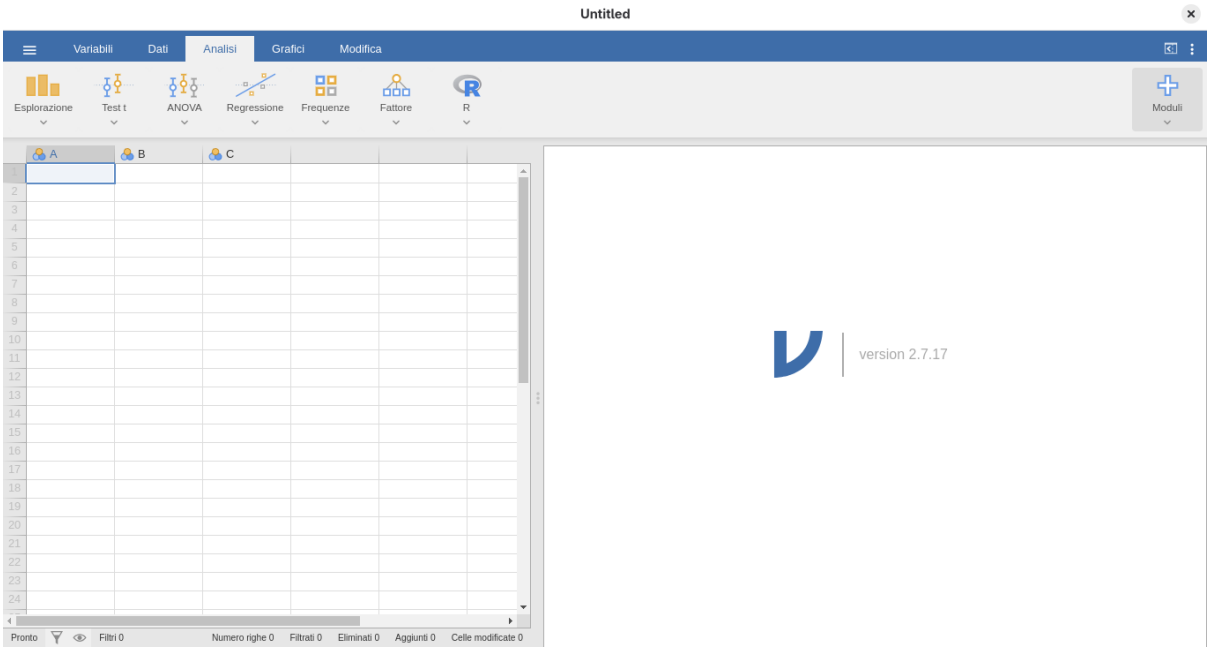
Troviamo jamovi all'indirizzo *https://www.jamovi.org/*, dove esso si presenta con le esaltanti affermazioni di essere avvincente alternativa a costosi prodotti statistici come SPSS e SAS e di essere concepito e costruito dalla comunità scientifica per la comunità scientifica.

Se apriamo la pagina DOCS accediamo ad un buon manuale in lingua inglese.

Da menu PRODUCTS ▷ JAMOVI DESKTOP andiamo all'indirizzo
https://www.jamovi.org/download.html
dove troviamo gli installer per Windows e Mac e il package flatpak per Linux.

2 Come funziona

jamovi si apre con questa finestra



che corrisponde alla scelta della linguetta ANALISI nella barra dei menu.

In corrispondenza a questa scelta la sottostante barra degli strumenti contiene icone corrispondenti ai tipi di analisi possibili.

Richiamo l'attenzione sull'ultima icona sulla destra del gruppo, che si riferisce al linguaggio R e che apre un editor per questo linguaggio, utilizzabile per scrivere comandi in linguaggio R per fare cose non previste dai richiami visuali attraverso le icone.

Essa non è presente per default ma si può avere caricando preventivamente un modulo apposito, ed è ciò che consiglio di fare per avere una completezza di strumenti di base.
A questo scopo si clicca sul pulsante con l'icona e la scritta MODULI in fondo a destra della barra degli strumenti, si sceglie LIBRERIE JAMOVI e si installa il primo modulo che compare nella lista, RJ EDITOR.

Sempre agendo sull'icona e la scritta MODULI in fondo a destra della barra degli strumenti possiamo installare altri moduli in arricchimento delle prestazioni di base di jamovi, già notevoli e più che sufficienti per la comunità di dilettanti evoluti cui dedico i miei manualetti.

In corrispondenza alle altre linguette (VARIABILI, DATI, GRAFICI e MODIFICA) abbiamo altre icone corrispondenti alle azioni possibili.

In ogni caso, delle due grandi finestre sottostanti le barre del menu e degli strumenti, quella di sinistra è destinata alle elaborazioni da impostare e quella di destra è destinata a mostrare il risultato delle elaborazioni.

Cliccando sulla prima icona della barra del menu, quella contrassegnata da tre barrette orizzontali sovrapposte, apriamo un menu da cui possiamo avviare un nuovo progetto, possiamo salvare il progetto in lavorazione e possiamo esportare l'intero contenuto della finestra dei risultati in formato PDF. Agendo sulla scelta del formato di esportazione possiamo esportare coerenti contenuti del progetto in formati adatti ai fogli di calcolo Calc e Excel e a molti altri formati, compresi quelli per interscambio con i software SPSS e SAS.

Per esportare zone selezionate della finestra dei risultati occorre selezionarle agendo con il cursore del mouse e, cliccando destro all'interno dell'area selezionata, scegliere il formato del file esportato.

Cliccando sulla voce APRI dello stesso menu possiamo aprire progetti salvati o importare dati da altre fonti, compresi i fogli di calcolo Calc e Excel e altri formati, compresi quelli per interscambio con i software SPSS e SAS.

3 A cosa serve

Come ormai si sarà compreso, jamovi è un software che ci consente di fare tante cose utilizzando il motore di calcolo statistico R in modo visuale, cioè senza scrivere codice.

Con le estensioni installabili possiamo fare altre cose, sempre in formato visuale.

Se tutto ciò non basta abbiamo a disposizione l'editor per scrivere istruzioni in linguaggio R ed eseguirle grazie all'editor installabile, come abbiamo visto, attraverso modulo aggiuntivo.

Da sapere che per lavorare con jamovi non abbiamo bisogno di installare anche R sul computer, in quanto jamovi ha al suo interno un motore R più che sufficiente per analisi anche molto impegnative.

Certo che, installando R puro e R ottimizzato possiamo andare anche oltre e, come ho detto in premessa, affrontare anche elaborazioni nel campo dei big data, sempre sapendo che non arriveremo mai a fare ciò che si può fare con Python.

C'é chi dice che R non è un motore big data ma è un cervello statistico e che si usa per pensare, non per spostare montagne.

Passiamo ora in rassegna tutto ciò che possiamo fare con jamovi in modo visuale, aprendo, una per una, le linguette della barra del menu.

3.1 Variabili

Nella finestra di apertura di jamovi, che è riprodotta a pagina 3, abbiamo già, nella finestra di lavoro sulla sinistra, una sorta di foglio elettronico destinato all'inserimento di dati con predisposte tre colonne, destinate a ricevere i dati di altrettante variabili, denominate A, B e C.

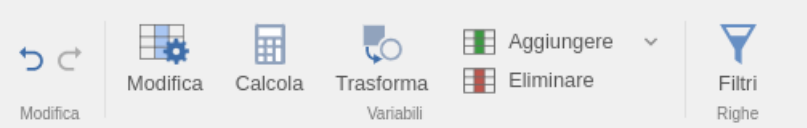
Rammento che, in statistica, una variabile è praticamente sempre una serie di dati e viene ospitata da jamovi, come avviene in altri software per elaborazioni statistiche e nei fogli di calcolo, in una colonna di un foglio elettronico.

Per default, le tre variabili presenti nel foglio in apertura del programma sono inizializzate al tipo di misura `Nominale` e al tipo di dato `Intero`. Si tratta di una scelta salomonica che usa un tipo di dato elaborabile (il numero intero) lasciando aperta la strada che lo può utilizzare come indicatore qualitativo (misura nominale).

Per analisi semplici ci possiamo accontentare, tanto più che, se inseriamo nella colonna della variabile che ci viene proposta per default una etichetta di testo anziché un numero intero, l'inizializzazione della variabile si adatta automaticamente al tipo di misura `Nominale` e al tipo di dato `Testo` e, se inseriamo nella colonna della variabile che ci viene proposta per default un numero con cifre decimali, l'inizializzazione della variabile si adatta automaticamente al tipo di misura `Continua` e al tipo di dato `Decimale`.

Tutte queste cose le possiamo vedere con doppio click sulla casella contenente l'intestazione della colonna del foglio elettronico.

Dal momento, poi, che, sempre agendo sul foglio elettronico che ci viene proposto in apertura del programma cliccando sull'intestazione bianca di colonne non ancora assegnate a variabili, possiamo aggiungere nuove variabili, anche calcolate e trasformate, potremmo anche fare a meno di ciò che ci viene proposto cliccando sulla linguetta di menu VARIABILI, azione attraverso la quale ci viene proposta questa barra di strumenti



Il ricorso a questo strumento serve a creare un data set con tipi di misura e tipi di dato precostituiti e non assoggettabili ad adattamenti automatici in relazione al dato inserito.

Come dire che, per analisi serie e complicate, è sempre bene avvalersi di questo strumento per intelaiare il data set.

Il primo pulsante sulla sinistra (icona con la scritta MODIFICA) serve per annullare o ripristinare modifiche effettuate prima.

L'altro pulsante, il secondo da sinistra (icona pure con la scritta MODIFICA), ci dà modo di scrivere una descrizione per le variabili e, con riferimento alla variabile selezionata, assegnarle un nome e fissarne tipo di misura (potendo scegliere tra Nominale, Ordinale e Continua) e il tipo del dato (potendo scegliere tra Intero, Decimale e Testo).

Procedendo verso destra abbiamo il pulsante corrispondente all'icona con la scritta CALCOLA, che ci dà modo di rendere variabile calcolata quella selezionata inserendo la formula per effettuare il calcolo coinvolgendo altre variabili, come fossimo in un foglio di calcolo.

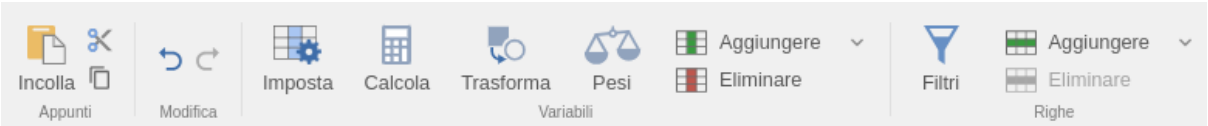
A seguire il pulsante corrispondente all'icona con la scritta TRASFORMA, che ci dà modo di creare una nuova variabile applicando una formula di trasformazione ai dati della variabile selezionata.

Poi tasti per aggiungere nuove variabili o eliminare variabili.

Infine un tasto per filtrare i dati, corrispondente all'icona con l'imbuto e la scritta FILTRO, che ci dà modo di eliminare dalla colonna della variabile selezionata dati corrispondenti ad una determinata condizione, creando una variabile filtro che si aggiunge a quelle presenti nel dataset. Eliminando questa variabile filtro tutto torna come prima.

3.2 Dati

Cliccando sulla seconda linguetta da sinistra della barra del menu, nominata DATI, mostriamo il dataset in un foglio elettronico su cui possiamo intervenire attraverso le azioni sottese ai pulsanti con le icone che compaiono in questa barra di strumenti



Alcune azioni sono simili a quelle già viste prima per la creazione delle variabili, con la differenza che qui, oltre a creare, modificare o trasformare una variabile agendo su un elenco di variabili lo facciamo agendo sul foglio elettronico che le contiene, potendo anche inserire le serie che compongono le variabili stesse.

Di nuovo troviamo il primo raggruppamento di icone che sottendono azioni di taglia, copia e incolla, la possibilità di aggiungere o eliminare oltre che colonne anche righe e la possibilità di pesare le componenti di una variabile.

A quest'ultimo scopo dobbiamo creare una variabile contenente i pesi da attribuire a varie righe di una variabile e poi, cliccato sul pulsante contrassegnato dall'icona con la scritta PESI, indicare la variabile cui attribuire questi pesi. In questo modo le analisi che coinvolgono quella variabile sono condotte attribuendo alla variabile stessa i pesi indicati. Eliminando la variabile contenente i pesi tutto torna come prima.

3.3 Analisi

La terza linguetta della barra del menu, nominata ANALISI, apre finalmente la strada al core del nostro software e cliccando su di essa ci ritroviamo questa barra di strumenti

* * *

Il primo strumento, ESPLORAZIONE, presenta questo menu

Scegliendo la voce DESCRITTIVE apriamo una finestra nella quale dobbiamo indicare la colonna/variabile o le colonne/variabili di cui indicare le statistiche. Se indichiamo più colonne le statistiche vengono indicate, colonna per colonna affiancate.
Per default vengono indicate media, mediana, deviazione standard, minimo e massimo. Se vogliamo altre statistiche possiamo sceglierle aprendo il sotto-menu STATISTICHE nella stessa finestra.
Sempre in questa finestra abbiamo un sottomenu GRAFICI che lasciamo agli esperti.

Per i meno esperti, a proposito di grafici, può essere utile scegliere una delle due voci di menu SCATTER PLOT e PARETO PLOT.

Nel primo caso, dovendo indicare dati numerici sia sull'asse delle ascisse sia sull'asse delle ordinate nella finestra che si apre, otteniamo un grafico con punti che segnano le corrispondenze tra le varie ascisse e le varie ordinate.
Aprendo la voce GENERAL OPTIONS in questa finestra possiamo scegliere se indicare l'interpolante (qui chiamata REGRESSION LINE), in modalità LINEAR o SMOOTH.
Sempre in questa finestra abbiamo le voci PLOT & AXIS TITLES, AXIS e LEGEND aprendo le quali abbiamo modo di intervenire per completare il grafico con titoli e legende.

Nel secondo caso, dovendo indicare dati di misura nominale o ordinale sull'asse delle ascisse e dati numerici sull'asse delle ordinate nella finestra che si apre, otteniamo una sorta di grafico a barre, detto di Pareto, arricchito della curva delle percentuali cumulative.

* * *

A seguire abbiamo lo strumento TEST T, che presenta questo menu

La prima scelta, TEST T A CAMPIONI INDIPENDENTI, serve per valutare la significatività della differenza tra medie campionarie osservate su campioni separati e indipendenti.
Come a dire verificare se il peso medio di pomodori coltivati in un'area concimata è significativamente diverso dal peso medio di pomodori coltivati in un'area non concimata.
In questo caso i dati delle osservazioni campionarie vanno inseriti in una sola colonna del foglio elettronico dei dati, come si trattasse di una sola variabile, affiancata da altra colonna in cui inserire un indicatore di raggruppamento per indicare il campione di appartenenza del dato inserito nella prima colonna (l'indicatore può essere il numero 1 per un campione e 2 per l'altro).
Nella finestra in cui inserire i dati, si inserisce il nome della prima colonna, l'unica contenente i dati delle rilevazioni campionarie, nella zona VARIABILI DIPENDENTI e il nome della colonna con gli indicatori di raggruppamento nella zona VARIABILE DI RAGGRUPPAMENTO.

La seconda scelta, TEST T A CAMPIONI ACCOPPIATI, serve per valutare la significatività della differenza tra medie campionarie su osservazioni fatte sullo stesso campione in momenti separati.
Come a dire verificare se la media delle misurazioni della pressione sanguigna a diverse ore effettuate prima di un trattamento è diversa da quella delle misurazioni effettuate alle stesse ore dopo un trattamento, sullo stesso paziente.
In questo caso abbiamo i dati delle misurazioni in due diverse colonne/variabili del foglio elettronico che contiene i dati e indichiamo i loro nomi nella zona VARIABILI ASSOCIATE nella finestra in cui inserire i dati.

La terza scelta, TEST T A CAMPIONE SINGOLO, serve per valutare la significatività della differenza tra una media campionaria e un certo valore.
Come a dire verificare se la media delle misurazioni del peso del contenuto di alcuni flaconi è significativamente diversa da quella indicata sull'etichetta dei flaconi.
In questo caso abbiamo una sola variabile da indicare nella zona VARIABILI DIPENDENTI della finestra in cui inserire i dati, quella che contiene le misurazioni campionarie, e dobbiamo inserire il valore con sui instaurare il confronto nella zona IPOTESI della stessa finestra, nella finestrella VALORE DEL TEST.

Nelle finestre per l'inserimento dei dati gli statistici professionisti trovano una miriade di possibilità di arricchire questi tipi di analisi basate sul test di Student.

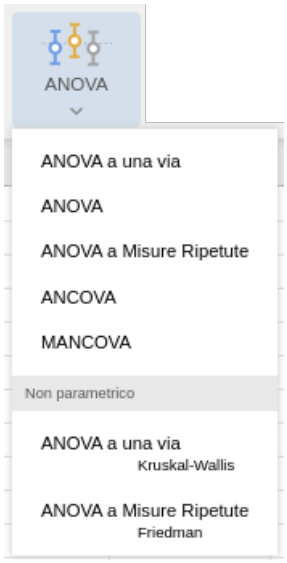
* * *

Il successivo strumento è etichettato ANOVA, che sta per ANalysis Of VAriance.
Nonostante il suo nome esso, come il test t che abbiamo visto prima, serve per valutare la significatività della differenza tra medie.
Questa apparente stranezza è dovuta al fatto che utilizza un adattamento del test F di Fisher-Snedecor, originariamente dedicato all'analisi della varianza.
Nella versione originale il test F si basa sul rapporto tra due varianze campionarie per verificare se esse possono dirsi accidentalmente o significativamente diverse.
Nell'adattamento utilizzato da ANOVA il test F si basa invece sul rapporto tra varianza tra gruppi e varianza entro i gruppi, cioè rapportando il quanto le medie di più gruppi sono lontane tra loro (tra gruppi) e il quanto i punti interni ai gruppi sono sparsi attorno alla loro media (entro i gruppi).
Per questa via, basandosi totalmente su analisi di varianze, si arriva a valutare la significatività o meno della differenza tra medie.

ANOVA nasce per superare due limiti dell'analisi attraverso il test t.
Il primo dato dal fatto che esso è applicabile solo a due gruppi per volta.
Il secondo, dato dal fatto che per essere certi dell'attendibilità dei risultati raggiunti applicando il test t dovremmo prima verificare che le varianze dei gruppi non siano significativamente diverse.

Con ANOVA possiamo confrontare più di due gruppi contemporaneamente, superando così il primo limite, e abbiamo a disposizione più strumenti di analisi, tra cui quello per superare anche il secondo limite.

Il suo menu si presenta così



L'applicazione normale risponde alla voce ANOVA.

La voce ANOVA A UNA VIA SERVE quando si vuole verificare se le medie di tre o più gruppi sono significativamente diverse in funzione di un solo fattore, per esempio per valutare l'effetto di diversi fertilizzanti sulla crescita di una pianta.

La voce ANOVA A MISURE RIPETUTE serve quando gli stessi soggetti vengono misurati più volte, come prima, durante e dopo un terapia sullo stesso paziente.

ANCOVA sta per Analisi della Covarianza, e serve per confrontare le medie dei gruppi aggiustate per una variabile continua.

MANCOVA sta per Analisi Multivariata della Covarianza ed è l'evoluzione multivariante di ANCOVA.

* * *

Proseguendo verso destra troviamo lo strumento REGRESSIONE, che presenta questo menu

La voce MATRICE DI CORRELAZIONE ci dà modo di ottenere il coefficiente di correlazione tra due variabili, nel caso di più variabili a coppie in una tabella a matrice ed anche di ottenere piccoli grafici che visualizzano l'interpolante. Gli indicatori che possiamo ottenere sono quelli di Pearson, di Spearman e di Kendall.

La voce CORRELAZIONE PARZIALE ci consente di ottenere il coefficiente di correlazione tra due variabili eliminando il possibile effetto di un'altra o di altre variabili, dette variabili di controllo. Per esempio per misurare la correlazione tra l'altezza (X) e la circonferenza del torace

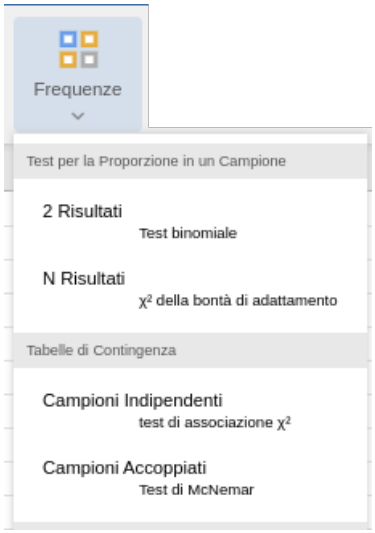
(Y) di un gruppo di persone eliminando l'influenza del peso (Z) che verosimilmente è correlato sia all'altezza che al torace.

Attraverso la voce REGRESSIONE LINEARE otteniamo gli elementi per determinare l'equazione della retta di regressione (intercetta e coefficienti) che sintetizza la relazione funzionale esistente tra una variabile dipendente e una o più variabili indipendenti (regressione multipla).
Le variabili indipendenti possono essere di tipo numerico, ed è questo il caso tipico, e jamovi le definisce Covariate ma possono anche essere di tipo nominale, nel qual caso sono definite Fattori.
Attraverso una variabile di pesatura possiamo attribuire pesi diversi alle variabili indipendenti.

Abbiamo poi accesso a tre varianti di REGRESSIONE LOGISTICA, a 2 RISULTATI se la variabile dipendente ha due sole classi (si/no, sano/malato), a N RISULTATI, utilizzata quando la variabile dipendente ha più di due classi non ordinate (auto/treno/bici/bus), e a RISULTATI ORDINALI. quando la variabile dipendente ha più di due classi ordinate (insufficiente/sufficiente/buono/ottimo).

* * *

Abbiamo poi lo strumento FREQUENZE, che presenta questo menu



I TEST PER LA PROPORZIONE IN UN CAMPIONE servono a verificare se la proporzione o la distribuzione di frequenze osservate in un campione sono diverse da una proporzione teorica o attesa o da una distribuzione teorica o attesa.
Se la variabile ha solo due categorie (si/no, promosso/bocciato) e basta confrontare la proporzione tra le due si ricorre al test binomiale a 2 RISULTATI.
Se la variabile ha più di due categorie diventa una distribuzione vera e propria e si ricorre al test a N RISULTATI, altrimenti noto come test del χ^2 della bontà di adattamento.

Una tabella di contingenza (o a doppia entrata) è uno strumento statistico utilizzato per mettere in relazione due o più variabili categoriali, organizzando i dati in righe e colonne per mostrare le frequenze congiunte.
L'analisi delle tabelle di contingenza (tabelle a doppia entrata) in jamovi si effettua nel modulo TABELLE DI CONTINGENZA che ci offre l'alternativa tra CAMPIONI INDIPENDENTI, applicabile nel caso ogni soggetto compaia una sola volta e i gruppi siano diversi tra loro, e CAMPIONI ACCOPPIATI, applicabile se i dati provengono dagli stessi soggetti misurati più volte.

* * *

Infine abbiamo lo strumento FATTORE, che presenta questo menu

L'ANALISI DI AFFIDABILITÀ serve a verificare quanto una scala misura in modo coerente e stabile ciò che vuole misurare, cioè controlla se gli item di una scala "vanno nella stessa direzione" e producono risultati consistenti.
Si usa per capire se le domande della scala sono coerenti tra loro, se ogni item contribuisce davvero alla misura del costrutto e se il punteggio totale della scala è affidabile.
Se una scala è poco affidabile, i risultati possono dipendere più dal caso o da errori di misura che dal fenomeno studiato.

L'ANALISI DELLE COMPONENTI PRINCIPALI è orientata alla riduzione dei dati con l'obiettivo di riassumere molti dati in poche componenti, creando nuove variabili come combinazione delle variabili originali.

L'ANALISI FATTORIALE ESPLORATIVA si applica quando non si sa quanti fattori esistono e lo si vuole scoprire dai dati,

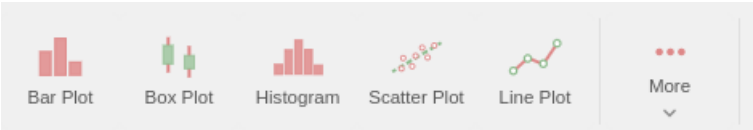
L'ANALISI FATTORIALE CONFERMATIVA serve per verificare i risultati dell'analisi fattoriale esplorativa.

* * *

Avendo caricato il relativo modulo secondo quanto abbiamo visto nel Capitolo 2, abbiamo anche a disposizione l'Editor R, in due modalità:
. RJ EDITOR, semplice e diretto, permette di scrivere ed eseguire codice R puro; i risultati vengono mostrati come output testuale o grafico. Serve per analisi personalizzate veloci a chi conosce già R e vuole usare jamovi solo come ambiente dati o per testare codice o fare analisi non presenti nei moduli standard.
. RJ EDITOR +, più avanzato e orientato allo sviluppo di analisi integrate. Serve per creare analisi riutilizzabili o sviluppare moduli o procedure personalizzate.

3.4 Grafici

La quarta linguetta della barra del menu, nominata GRAFICI ci offre la possibilità di tracciare alcune tipologie di grafico e cliccando su di essa ci ritroviamo questa barra di strumenti

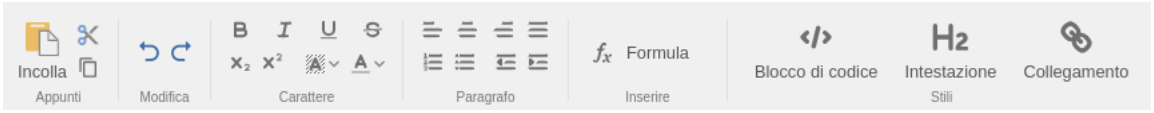


Troviamo tre tipi di grafico adatti alla visualizzazione di frequenze e distribuzioni di frequenze, BAR PLOT, BOX PLOT e HISTOGRAM e due tipi di grafico per visualizzare la relazione intercorrente tra due variabili numeriche, SCATTER PLOT e LINE PLOT.

Cliccando su MORE troviamo la possibilità di costruire un grafico PARETO PLOT. Si tratta di un grafico a barre che si presta a visualizzare grandezze (asse Y) riferite a variabili qualitative (asse X) con la curva rappresentante la percentuale cumulativa.

3.5 Modifica

Abbiamo infine la linguetta MODIFICA, cliccando sulla quale apriamo questa barra di strumenti



Questa volta si tratta di strumenti che non hanno nulla a che vedere con la statistica, ma servono a formattare un testo, come fossero da utilizzare in un editor di testo.

E' con questi strumenti che possiamo intervenire nella finestra in cui abbiamo i risultati delle elaborazioni per inserire annotazioni a chiarimento delle elaborazioni stesse.