

DIVIDI/DISTINGUI:**Dati → suddividi file**

- Per raggruppare i risultati
- Si può distinguere le statistiche descrittive (es. per genere, età, diploma, ...)
- Funzione che ripeterà le singole analisi per sottocampioni, producendo un output di due tipi:
 - Confronta gruppi = presenta i risultati insieme per comparare i gruppi
 - Distingui risultati per gruppo = mostra tutti i risultati del primo gruppo, del secondo, ecc
- Rivedere alcune statistiche suddividendole per sottogruppi
- Suddividi file ti mostra le frequenze della variabile suddivisa per una variabile filtro

SELEZIONA CASI:**Dati → seleziona casi**

- Selezionare nelle matrici di dati solo quei casi che presentano determinate caratteristiche e utilizzare per le proprie analisi/trasformazioni solo un determinato sottocampione
- Nell'output segnala la presenza di un filtro. Nella matrice i casi non selezionati vengono barrati. Viene creata una nuova variabile "Filter" il cui valore 1 rappresenta la selezione del caso e il valore 0 l'esclusione del caso. Si attiva una scritta nella barra delle funzioni in basso a dx (filtro attivato)
- Ti permette di escludere una parte del campione e utilizzare un sottocampione
- Tutto il lavoro di analisi/trasformazione viene operato su quel determinato campione
- Per disattivarla: selezionare "tutti i casi" [ricordarsi di disattivarla quando le analisi sono terminate, sennò gli output si riferiranno al sottocampione selezionato]

RICODIFICA VARIABILI:**Trasforma → ricodifica variabili**

- Attribuire nuovi valori ai valori di una variabile
- Utile quando è necessario creare dei gruppi in una variabile
- Creare una variabile che presenti un determinato valore per ognuno dei casi della variabile che ci interessa
- Anche inversione di una variabile: se vogliamo sommare due variabili, per esempio una che misura ottimismo e l'altra pessimismo, ottenendo un indice di ottimismo, dobbiamo invertire la scala pessimismo
- Assegnare un nuovo valore ad uno vecchio, utile per raggruppare dei valori/associare un valore numerico nuovo a un vecchio valore
- Troviamo:
 - Ricodifica nelle stesse variabili = sovrascrive i vecchi valori con i nuovi, perdendo il dato originale [**trasforma → ricodifica nelle stesse variabili**]
 - Ricodifica in nuove variabili = crea una nuova variabile con nuovi valori [**trasforma → ricodifica in variabili differenti**]
 - Ricodifica automatica = per le variabili testuali, ordina automaticamente in ordine alfabetico la variabile interessata, attribuisce il valore 1 ai casi mancanti e un valore progressivo a tutte le risposte successive [**trasforma → ricodifica automatica**]

CALCOLA VARIABILE:**Trasforma → calcola variabile**

- Costruire nuove variabili su funzioni matematiche/formule
- Elaborare una nuova variabile a partire da una formula o nuova variabile partendo da altre variabili
- Molto usata per costruire scale e indici (composti spesso da somma/media di altre variabili)
- Usata anche per normalizzazione di una variabile asimmetrica o standardizzazione di un punteggio
- Quando dobbiamo calcolare un punteggio sulla base di una qualsiasi formula
- Controllare valori mancanti/anomali; attenzione a variabili che devono essere invertite

ANALISI DESCRIPTIVE:

- Quando vogliamo capire come è fatta una variabile = la distribuzione, le sue caratteristiche
- Analisi descrittive monovariate (una sola variabile per volta): ti indicano le caratteristiche statistiche di una variabile + capire come le persone rispondono alle diverse modalità di risposta di una domanda
- Prima di ogni analisi: consigliare esplorare le analisi, sia per le frequenze e le descrittive generali, sia per la normalità. Se la distribuzione non fosse normale: adottare trasformazioni per ridurre gli effetti della non normalità
- Tre funzioni:
 - Frequenze = indicatori e in particolare la frequenza di ogni modalità; specifica per variabili qualitative [**analizza → statistiche descrittive → frequenze**]
 - Descrittive = veloce, sintetica, non richiede particolari configurazioni; specifica per variabili quantitative [**analizza → statistiche descrittive → descrittive**]
 - Esplora = più completa delle altre, offre anche indicatori di normalità, outliers (valori anomali), missing (valori mancanti), rappresentazioni grafiche; scopo di conoscere (esplorare) le variabili (in particolari quelle quantitative). Fondamentale come analisi preliminare a qualsiasi test basato sulle medie [**analizza → statistiche descrittive → esplora**]

T-TEST PREMESSA:

- Per comprendere il rapporto tra due variabili quantitative, quando vogliamo confrontarne le medie
- Variabili usate: ad intervalli o a rapporti (ogni tanto anche a livello ordinale)
- Prima di un t-test: esplorare le frequenze delle variabili; valutare la presenza di eventuali valori anomali ed escluderli dalle analisi
- Tre applicazioni del t-test:
 - T-test a campione unico = quando abbiamo una variabile continua e un possibile parametro di confronto (esempio: la media di un altro gruppo, una media attesa, un singolo valore per testarne l'appartenenza alla medesima popolazione) [**analizza → confronta medie → t-test a campione singolo**]

- T-test per campioni indipendenti = quando abbiamo una variabile dipendente continua e una variabile che indica due differenti sottogruppi (es: genere, m/f), per verificare se i due campioni sono estratti dalla stessa popolazione [analizza → confronta medie → t-test a campioni indipendenti]
- T-test per campioni appaiati = quando vogliamo verificare se due rilevazioni compiute sullo stesso campione presentano differenze significative o se possono essere riferite ai parametri della medesima popolazione (esempio: pre e post test effettuati sullo stesso gruppo dopo un periodo di addestramento [analizza → confronta medie → t-test campioni appaiati])
- T-test richiede tre condizioni:
 1. La variabile dipendente deve essere il più possibile normale
 2. Le varianze delle due popolazioni devono essere uguali
 3. Le osservazioni in ciascun gruppo devono essere indipendenti l'una dall'altra, sia nella singola variabile, sia tra le due variabili (vale solo per t-test per campioni indipendenti)
- Prima di qualsiasi t-test: **analizza → statistiche descrittive → esplora** = individuare valori anomali, conoscere la distribuzione di frequenza delle variabili utilizzate, comprendere come può essere composta la media delle variabili utilizzate

T-TEST A CAMPIONE UNICO:**Analizza → confronta medie → t-test campione singolo**

- Utilizzato con variabile numerica + valore di riferimento (ad esempio la media di risposte corrette che sono state fornite nella stessa prova in un diverso campione, proveniente da un altro paese, scuola o dall'intera popolazione di riferimento)
- Ipotesi nulla (H0): la media della popolazione da cui è estratto il campione non differisce da quella della popolazione di riferimento, quindi i due valori non sono significativamente differenti ($p > .05$ = differenza non significativa)
- Ipotesi alternativa (H1): le due medie differiscono fra loro, quindi i due valori sono significativamente differenti ($p < .05$ = differenza significativa)

T-TEST PER CAMPIONI INDIPENDENTI:**Analizza → confronta medie → t-test campioni indipendenti**

- Utilizzato con variabile numerica + variabile dicotomica che indica due gruppi
- Ipotesi nulla (H0): le medie delle popolazioni da cui sono stati estratti i due gruppi non sono differenti ($p > .05$ = differenza non significativa)
- Ipotesi alternativa (H1): i parametri dei due gruppi sono differenti ($p < .05$ = differenza significativa)
- Utilizzato quando si vuole verificare se esistono differenze tra due gruppi, indicati in una variabile dicotomica: utilizzabile per verificare, per esempio, se esistono differenze significative nel voto di laurea tra maschi e femmine

T-TEST PER CAMPIONI APPAIATI:**Analizza → confronta medie → t-test campioni appaiati**

- Utilizzato per confrontare due variabili appaiate rilevate sullo stesso campione
- Ipotesi nulla (H0): le due prove (prima e dopo) non sono significativamente differenti ($p > .05$ = differenza non significativa)
- Ipotesi alternativa (H1): le due prove hanno medie significativamente differenti ($p < .05$ = differenza significativa) e quindi è avvenuto un cambiamento tra il prima e dopo
- T-test che permette di confrontare due variabili assimilabili, cioè che hanno le stesse caratteristiche. Le due variabili devono avere gli stessi minimi e massimi teorici

ANOVA A UNA VIA:**Analizza → confronta medie → anova**

- Analisi della varianza, utilizzata quando i gruppi sono più di due
- Utilizzata quando abbiamo una variabile quantitativa (variabile indipendente) + una variabile nominale che indica i gruppi
- Ipotesi nulla (H0): le differenze all'interno dei gruppi non dipendono dal fattore individuato e pertanto i gruppi provengono dalla medesima popolazione ($p > .05$ = differenza non significativa)
- Ipotesi alternativa (H1): le differenze all'interno dei gruppi possono dipendere dal fattore individuato e pertanto almeno uno dei gruppi provenienti da popolazioni differenti ($p < .05$ = differenza significativa tra i gruppi)
- Utilizziamo Anova quando vogliamo confrontare le differenze nelle medie di più di due gruppi

CHI-QUADRO, TAVOLE DI CONTINGENZA:**Analizza → statistiche descrittive → tavole di contingenza**

- Chi-quadro = distribuzione statistica utilizzata per effettuare confronti basati sulla frequenza
- Utilizzato quando abbiamo due variabili qualitative
- Tavole di contingenza = tabelle in cui vengono incrociate due variabili
- Ipotesi nulla (H0): le due variabili sono indipendenti ($p > .05$ = non esiste una relazione) → l'ipotesi nulla utilizzata nel chi-quadro è la non dipendenza delle due variabili, ossia l'assenza di relazione tra le frequenze osservate
- Ipotesi alternativa (H1): esiste un qualche tipo di relazione tra le due variabili ($p < .05$ = esiste una relazione) → l'ipotesi alternativa è che ci sia relazione e abbia una certa intensità
- Relazione misurata confrontando la frequenza osservata (le risposte dei nostri soggetti) con la frequenza teorica/attesa, elaborata sulla base della probabilità di indipendenza nelle frequenze
- Se non c'è relazione: possiamo ipotizzare che le frequenze si distribuiscono indipendentemente l'una dall'altra; all'opposto se esiste relazione, osserveremo un andamento diagonale dei valori più frequenti nelle due celle
- Uso le tavole di contingenza quando ho due variabili qualitative (nominali oppure ordinali) su cui puoi lavorare solo a livello di frequenze. Le tavole permettono di incrociare le due variabili e capire come si distribuiscono
- Prima di usare le tavole:
 1. Controllare mediante analisi → statistiche descrittive → frequenze se esistono modalità di risposta con pochi casi, che possono essere esclusi con dati → seleziona casi
 2. Controllare eventuali variabili mancanti
 3. Il chi-quadro tende ad aumentare (quindi a diventare significativo) con l'aumentare della numerosità. Chiedendo PHI e V di Cramer (analisi → statistiche descrittive → tavole di contingenza → statistiche) posso capire se la significatività è dovuta alla numerosità: più il risultato si avvicina

allo 0, più l'effetto può essere dovuto proprio alla numerosità. Nelle tabella 2x2, l'indice PHI può essere anche utilizzato come indice di dimensione dell'effetto e indica quando è intensa la correlazione tra le due variabili, comportandosi come la correlazione

CHI-QUADRO CON MODELLO EQUIPROBABILE:

Analizza → test non parametrici → finestre di dialogo legacy → chi-quadro

- Il chi-quadro viene utilizzato per effettuare confronti basati sulle frequenze
- L'uso non parametrico permette di lavorare una variabile qualitativa
- Variabile: sia nominale che ordinale
- Ipotesi nulla (H0): le risposte sono state scelte a caso ($p > .05$ = non esistono differenze significative) → l'ipotesi nulla è l'equiprobabilità delle risposte, ossia che tutte le risposte possibili non siano influenzate da assolutamente nulla e quindi tutte con uguale probabilità
- Ipotesi alternativa (H1): esiste un criterio di scelta non casuale ($p < .05$ = esistono differenze significative) → l'ipotesi alternativa è che le risposte si distribuiscono secondo un altro criterio, non casuale
- Confrontata la frequenza osservata (risposte dei soggetti) con frequenza attesa, elaborata sulla base dell'equiprobabilità delle frequenze
- Chi-quadro con modello di equiprobabilità: lo usiamo quando si ha una variabile qualitativa e vogliamo verificare che la distribuzione delle diverse modalità di risposta sia statisticamente significativa rispetto ad una scelta casuale (mancanza di influenza) delle risposte

CHI-QUADRO, USO NON PARAMETRICO, MODELLO TEORICO:

Analizza → test non parametrici → finestre dialogo legacy → chi-quadro

- Confronta le medie attese con un modello teorico, una distribuzione ben definita
- Modello teorico: può provenire da una teoria oppure dalla distribuzione osservata in altri studi e ricerche analoghe
- Utilizzabile quando ho a disposizione una distribuzione teorica specifica e quando ho variabile qualitativa + distribuzione teorica
- Variabili nominali o ordinali
- Ipotesi nulla (H0): le due distribuzioni sono identiche ($p > .05$ = non esistono differenze significative tra i due gruppi) → le risposte si distribuiscono secondo il modello teorico che stiamo utilizzando
- Ipotesi alternativa (H1): le due distribuzioni sono differenti ($p < .05$ = esistono differenze significative tra i due gruppi) → le risposte si distribuiscono secondo un altro criterio
- Confronto frequenza osservata (risposte soggetti) con frequenza attesa, elaborata sulla base dell'ipotesi nulla
- Ci permette di confrontare la nostra distribuzione con la distribuzione teorica basata su un modello di distribuzione
- Utilizzabile anche quando si ha variabile qualitativa e si vuole verificare che la distribuzione delle diverse modalità di risposta sia statisticamente significativa rispetto ad un modello teorico di partenza

BINOMIALE:

Analizza → test non parametrici → finestre di dialogo legacy → binomiale

- Molto usata con variabile dicotomica e si vuole valutare la distribuzione rispetto ad un modello teorico, che può essere equiprobabile (50%-50%) oppure con diversa distribuzione
- Equivale al calcolo del chi-quadro per due sole categorie
- Si usa con variabile qualitativa dicotomica + distribuzione teorica o equiprobabile
- Ipotesi nulla (H0): le risposte si distribuiscono secondo il modello teorico utilizzato [corrisponde al valore p della prima categorie] = le distribuzioni sono identiche ($p > .05$ = non esistono differenze significative tra i due gruppi)
- Ipotesi alternativa (H1): le risposte si distribuiscono secondo un altro criterio = le due distribuzioni sono differenti ($p < .05$ = esistono differenze significative tra i due gruppi)
- Confrontata la frequenza osservata con quella teorica, di cui esprimiamo la proporzione (la probabilità)
- Si può usare la binomiale in due modi:
 1. Verificare se la distribuzione delle due modalità di risposta di una variabile dicotomica è casuale o no
 2. Verificare se il campione per una certa variabile dicotomica appartiene ad una certa popolazione, di cui conosco la distribuzione

CORRELAZIONE:

Analizza → correlazione → bivariata

- Per comprendere se due variabili covariano insieme = al mutare di una corrisponde la variazione dell'altra
- Quando ho due variabili quantitative
- Ipotesi nulla (H0): nella popolazione non esiste relazione tra le due variabili → la correlazione è nulla nella popolazione da cui sono state estratte le due variabili, quindi la coppia di variabili osservate non hanno una relazione ($p > .05$)
- Ipotesi alternativa (H1): nella popolazione esiste una relazione e ha una certa intensità → le coppie di variabili hanno una relazione ($p < .05$)
- Se c'è correlazione, può essere:
 - Positiva = al crescere di una delle due cresce anche l'altra, o al decrescere di una decresce l'altra
 - Negativa = al crescere di una, l'altra decresce

LIVELLI DI SCALA (SCALE DI MISURA)

- Livello nominale = permette di assegnare un'etichetta ad ogni valore assunto da una variabile (categoria) quindi solo dei nomi (es: maschio, femmina). NO media, d-s, IQR, percentili e intervallo. SI moda
- Livello ordinale = permette di ordinare le categorie di una variabile secondo un determinato criterio (es: magro, normale, sovrappeso, obeso). NO media, d-s, IQR, intervallo. SI moda
- Livelli ad intervalli/a rapporti = utilizzano una qualsiasi unità di misura specifica per quella variabile (es: età utilizzata, anno, temperatura); si differenziano per l'origine (o punto zero) relativa (scala ad intervallo) o assoluta (scala a rapporti); il numero corrispondente all'anno (2022) dipende dal momento in cui facciamo iniziare il conteggio (nascita di Cristo) ed è quindi un'origine relativa, la nostra età personale ha invece un'origine assoluta (anno di nascita)

TIPOLOGIE DI SCALA

- Livelli nominale e ordinale = utilizzano numeri, ma in realtà a questi livelli non esiste una vera e propria unità di misura e i valori che una variabile assume sono di tipo qualitativo (colore dei capelli: castano, biondo, ...; corporatura: mingherlino, normale, sovrappeso, obeso)

- Livelli intervallo e rapporti = fanno uso di veri numeri, essendo basati su unità di misura; di tipo quantitativo