

# DOMANDE-ESERCIZI

e risposte

Così come le domande non sono orientate ad avere un numerino, così queste risposte sono, in realtà, delle spiegazioni del tipo di ragionamento da fare.

Prima di iniziare a lavorare in SPSS, sarà meglio verificare che sia impostato in modo da visualizzare il nome delle variabili e non le etichette. Usate **Modifica | Opzioni** e nel pannello **Generale**, impostate **Elenchi di variabili su Visualizza nomi**.

## Usando xEsercizi1.sav

A) Da un punto di vista statistico, useresti RESID come variabile indipendente?

La variabile RESID è una variabile nominale e, in quanto tale, è possibile utilizzarla come una variabile indipendente. Tuttavia bisogna prima capire se la sua distribuzione nel campione raccolto la rende effettivamente utilizzabile. Calcolando le frequenze di RESID (**Analizza | Statistiche descrittive | Frequenze**) vediamo che tre quarti del campione cade in una categoria, c'è una categoria con due risposte e le altre due hanno 10 e 28 risposte. A questo punto la variabile non è utilizzabile perché (ad es. in un'anova) confronteremmo un gruppo di 159 casi con uno di 28.

B) Ha senso calcolare un chi quadro fra PADRE e MADRE?

Osservando le due variabili vediamo che sono entrambe nominali, quindi un chi-quadro sulla loro tabella di contingenza avrebbe senso. Tuttavia se facciamo veramente la tabella di contingenza (**Analizza | Statistiche descrittive | Tavole di contingenza**), vediamo che ci sono molte celle vuote o con valori bassi. Se facciamo calcolare il chi-quadro, vediamo che SPSS ci avverte che il 78% delle celle ha un valore atteso inferiore a 5. Quindi il chi quadro calcolato non si approssima alla distribuzione di chi-quadro e non è valido. Non si può calcolare il chi-quadro fra queste due variabili senza prima compattare alcune categorie (o eliminarle).

C) Studiando le correlazioni c'è qualche possibilità che le variabili da FAV ad AC possano servire per spiegare BENEVOL?

Si tratta di un numero elevato di variabili (circa 30) quindi ci sono possibilità teoriche che qualcuna possa spiegare BENEVOL. Tuttavia dobbiamo considerare che, per poter effettuare una regressione multipla, le variabili di tipo X devono avere correlazione almeno media con la dipendente e, inoltre, che le variabili X dovrebbero essere poco correlate fra loro per evitare i problemi di multicollinearità.

Facciamo allora una matrice di correlazione fra BENEVOL (indicata per prima) e le variabili comprese fra FAV e AC (**Analizza | Correlazione | Bivariata**). Vediamo che parecchie correlazioni con BENEVOL non sono significative e che quelle significative sono piuttosto basse (da .145 a .249). Elimino tutte le variabili non significative e rifaccio le correlazioni (non è obbligatorio, ma i risultati si leggono meglio). E questa volta guardo anche le correlazioni fra le X. Ci sono molte variabili che correlano fra loro con valori superiori a quelli con cui correlano con BENEVOL. Quindi abbiamo parecchie variabili multicollineari, ad es. FAV, UNFAV e COMMUN. E' possibile fare una regressione multipla, ma bisognerà stare molto attenti alle variabili da utilizzare e non dovremo aspettarci che il modello spieghi una percentuale consistente di varianza.

# Usando il file Matt.sav

I dati provengono da una ricerca di psicologia della religione. Le variabili da D1 a D13 sono spiegate nel file SPSS. Per le altre: H e K misurano una tipologia di attaccamento al partner (H) e Dio (K); G è la scala di orientamento religioso; A la scala di fondamentalismo; R di autoritarismo.

## 1. La variabile D1 è l'anno di nascita.

- (a) Calcola l'età, considerando che il questionario è stato compilato nel 2006.

Usando **Trasforma | Calcola variabile**, mettiamo **ETA** come variabile di destinazione e **2006-D1** come espressione numerica, premete **OK**. Se invece premete **Incolla** otterrete questo: `COMPUTE eta = 2006 - D1.`

- (b) Crea una variabile ordinale **FETA** che raggruppi i giovani (dai 18 ai 30 anni), gli adulti (dai 31 ai 55 anni) e gli anziani (da 56 anni in su).

Basta usare **Trasforma | Ricodifica in variabili differenti** e inserire il nome della nuova variabile e le ricodifiche. Sostanzialmente l'incolla sarà: `RECODE eta (18 thru 30=1) (31 thru 55=2) (56 thru Highest=3) INTO feta .`

- (c) Qual è la distribuzione per genere e per fasce d'età?

Dovremmo considerare **D2** e **FETA**, in questo si tratta di fare una tabella di contingenza. Se invece consideriamo **ETA**, si tratta di fare le medie suddivise per genere o per fasce d'età.


Nel primo caso, useremo **Analizza | Statistiche descrittive | Tavole di contingenza**, chiedendo le percentuali di riga e/o di colonna. già che ci siamo possiamo anche chiedere il chi-quadro. Dai dati vediamo che mentre nei giovani e negli adulti abbiamo più o meno un equilibrio di maschi e di femmine, fra gli anziani ci sono parecchi maschi in più. Questa differenza è sufficiente per dire che le due variabili non sono indipendenti e che quindi la distribuzione non è casuale? Guardando il chi-quadro vedo che non è significativo e quindi la distribuzione per genere e per età è casuale.


Nel secondo caso, useremo **Analizza | Confronta medie | Medie**. **ETA** sarà la variabile dipendente, **FETA** e **D2** le indipendenti. Dai dati vediamo che non vi è sostanziale differenza di età (attorno a 41 anni), mentre (logicamente ed ovviamente!) vi sono medie sensibilmente diverse per fasce d'età.

## 2. La variabile D12 misura l'orientamento politico da sinistra (1) a destra (10). Mentre la variabile D13 indica le persone che non hanno un orientamento politico (valore 2).

- (a) Se una persona ha dichiarato di non avere un orientamento politico (**D13=2**), la variabile **D12** dev'essere posta a missing.

Possiamo fare due cose: creare una nuova variabile che contenga la versione modificata oppure modificare direttamente **D13**.

Nel primo caso prima dobbiamo copiare il contenuto della variabile attuale in quella nuova; usiamo **Trasforma | Calcola variabile**, clicchiamo su  per svuotare la finestra di dialogo (che contiene i calcoli precedenti), usiamo **D12n** come nome della nuova variabile e **D12** come espressione.

Adesso possiamo usare **Trasforma | Ricodifica nelle stesse variabili**, mettendo **D12N** nelle variabili da ricodificare. Innanzitutto non dobbiamo ricodificare tutta la variabile, ma solo quando **D13** è uguale a 2: quindi premiamo il pulsante ,

selezioniamo **Includi se il caso soddisfa la condizione** e scriviamo D13=2 nello spazio apposito, quindi pulsante **Continua** . Adesso possiamo cliccare su **Valori vecchi e nuovi...** e dire che **Tutti gli altri valori** devono diventare **Mancante di sistema**, quindi pulsante **Aggiungi** e poi **Continua** ; infine **OK** .

Nel secondo caso, usiamo direttamente **Trasforma | Ricodifica nelle stesse variabili** sulla variabile D12 e poi facciamo le stesse cose indicate al paragrafo precedente.

- (b) Qual è la media di D12 suddivisa (separatamente) per genere, fasce d'età (FETA), pratica religiosa (D6)

Di nuovo usiamo **Analizza | Confronta medie | Medie**. Metteremo D12 o D12N (quella ricodificata) fra le dipendenti e D2, feta, D6 fra le indipendenti. Avremo tre tabelle con le medie per ciascun valore delle tre variabili indipendenti.

- (c) Crea una nuova variabile (POLIT, di tipo categoriale) che abbia i seguenti valori: 1=sinistra (D12 da 1 a 3), 2=centro (D12 da 4 a 7), 3=destra (D12 da 8 a 10), 4=nessun orientamento (D13=2)

Se consideriamo che D12 o D12N contiene un valore mancante per D13=2 e se consideriamo che eventuali persone che non hanno risposto alla D12 possono essere considerate come persone senza preferenza politica, possiamo semplicemente ricodificare questa variabile, usando **Trasforma | Ricodifica in variabili differenti**, ricordandosi di azzerare le scelte precedenti tramite il bottone **Ripristina** . POLIT sarà il nome della nuova variabile, quindi inseriremo i valori da 1 a 3 e gli assegniamo valore 1 e così via, terminando assegnando a **Mancante di sistema** il valore 4.

3. Gli item da A1 ad A20 formano la scala di Fondamentalismo di Altemeier & Hunsberger (FONDAM) e si ottiene sommando tutti i 20 item dopo aver invertito A2, A5, A7, A9, A10, A12, A13, A15, A17, A19.

- (a) Come potresti verificare che sono effettivamente questi gli item da ribaltare? Cosa usi e perché?

Nel costruire uno strumento (una scala psicologica) spesso si inverte il senso di una frase per evitare che il rispondente dia risposte casuali o segni sempre lo stesso valore. Per cui, per essere d'accordo con alcuni item bisogna segnare i valori alti, per essere d'accordo con altri item bisogna segnare valori bassi. In alcuni casi vengono chiamati item pro e item contro(scalari). In teoria, a punteggi alti sugli item pro, dovrebbero corrispondere valori bassi su quelli contro e viceversa. La statistica della correlazione ci dà esattamente questa informazione, per cui se corrolo tutti gli item A, dovrei trovare che quelli da invertire hanno correlazioni negative con quelli da non invertire e positive con gli altri da invertire.

Quindi uso **Analizza | Correlazione | Bivariata** selezionando tutte le variabili A (meglio selezionare prima tutte quelle da non invertire e poi quelle da invertire). In questo modo vedo facilmente che tutte le correlazioni fra le pro sono positive, le correlazioni fra le contro sono positive e le correlazioni fra pro e contro sono negative.

- (b) Come puoi fare per ribaltarli? [Fallo!]

Innanzitutto bisogna sapere qual era la scala utilizzata per la misurazione. Se avessimo costruito noi il questionario lo sapremmo, ma non lo sappiamo. Basta però chiedere il valore minimo e il valore massimo per ogni variabile. Usando **Analizza | Statistiche**

**descrittive** | **Descrittive** e scegliendo fra le opzioni il minimo e il massimo, possiamo facilmente vedere che tutti i 20 item hanno una gamma che oscilla da 1 a 9. A questo punto, il valore 1 deve diventare 9, il 2 deve diventare 8 e così via. possiamo usare la ricodifica nelle stesse variabili oppure in variabili differenti (meglio). Tuttavia possiamo anche considerare che  $10-1=9$  e  $10-9=1$ ,  $10-2=8$ ,  $10-3=7$  e così via... Quindi basta sottrarre il valore attuale del caso statistico al valore massimo aumentato di 1.

Quindi costruiamo l'istruzione per A2 usando il menù: **Trasforma** | **Calcola variabile**, A2R sarà la nuova variabile, mentre  $10-A2$  sarà l'espressione. Anziché dare l'OK, usiamo **Incolla**, che produrrà **COMPUTE A2R = 10-A2.**, che si può facilmente copiare e modificare per A5, A7 e così via. Alla fine evidenziamo le righe e **Esegui** | **Selezione**. Ricodificate!

- (c) Calcola la variabile FONDAM (usando gli item ribaltati) e studia la sua distribuzione di normalità (media, ds, curtosi, asimmetria, outliers).

Possiamo usare di nuovo il menù **Trasforma** | **Calcola variabile**: metteremo FONDAM come nome della nuova variabile e  $A1+A2R+A3+A4+A5R...$  come espressione. Per conoscere media, ds, curtosi, asimmetria, outliers possiamo usare **Analizza** | **Statistiche descrittive** | **Esplora** ricordandoci di selezionare **Grafici di normalità con test** dal pannello **Grafici...**. Dai risultati possiamo vedere la media (102,15), la ds (18 arrotondata), la curtosi (1.968) e l'asimmetria (.695). Vediamo quindi che mentre possiamo accettare l'asimmetria, non è accettabile la curtosi. Guardando i grafici Box-plot vediamo che ci sono dei soggetti outliers sia con punteggi molto bassi sia con punteggi molto alti.

- (d) Vi è differenza statistica rispetto alla variabile D6 usata come indipendente?

Per saperlo dobbiamo fare un test inferenziale. La variabile D6 rappresenta la pratica religiosa e ha 4 categorie. Potremmo fare direttamente anova univariata, ma controlliamo prima la distribuzione della variabile calcolando le frequenze. Guardando la tabella delle frequenze, vediamo che c'è una categoria con un solo caso che quindi va eliminato prima di effettuare l'analisi della varianza. Per eliminarlo, lo definiamo come "mancante di sistema". A questo punto con **Analizza** | **Confronta medie** | **ANOVA univariata**, inseriamo D6 come fattore e FONDAM come dipendente. Siccome avremo 3 medie, scegliamo anche un post-hoc (ad es. Tukey). L'anova è significativa, quindi c'è almeno una media diversa dalle altre e l'analisi a post-hoc ci dice quale.

- (e) Altemeier afferma che la sua scala è monofattoriale, mentre alcuni autori europei, con dati di campioni tedeschi e rumeni, hanno trovato una struttura a 2 fattori. Il campione italiano che struttura fattoriale ha?

Si tratta di fare un'analisi fattoriale esplorativa (AFE): **Analizza** | **Riduzione dati** | **Analisi fattoriale**. Inseriamo le 20 variabili A; da **Descrittive...** possiamo selezionare il test KMO e la sfericità di Bartlett, da **Estrazione...** il metodo della massima verosimiglianza, il grafico degli autovalori (possiamo escludere la soluzione non ruotata) e possiamo lasciare impostato il criterio di Kaiser. Dal bottone **Rotazione** possiamo anche non selezionare una rotazione, per rifare successivamente l'analisi, ma conviene chiedere subito la rotazione Varimax. Da **Opzioni** possiamo chiedere che gli item siano ordinati per saturazioni decrescenti nei fattori e possiamo anche chiedere che vengano azzerati i valori inferiori a .20 (scelta del tutto arbitraria).

(Se li avete fatti) KMO è maggiore di .60 e quindi significa che ci sono abbastanza correlazioni medie e medio-alte per poter fare un'AFE; anche il test della sfericità di

Bartlett (significativo) ci dice che la matrice delle correlazioni non è una matrice I (identità) e non vi si avvicina.

4. Gli item da G1 a G14 formano la scala di Orientamento religioso Intrinseco-Estrinseco di Allport, che è suddivisa in 2 sottoscale: orientamento intrinseco (INTRINS) è formato dagli item G1, G3, G4, G5, G7, G10, G12, G14; l'orientamento estrinseco (ESTRIN) dagli item G2, G6, G8, G9, G11, G13.
  - (a) Verifica che non vi siano item da ribaltare
  - (b) Calcola le scale INTRINS e ESTRIN e verifica la loro normalità
  - (c) Vi è bassa o alta correlazione fra le due variabili?
  - (d) Ha senso usare queste due variabili per spiegare FONDAM?
  - (e) Se ha senso, quanta % di varianza viene spiegata da questo modello?
  - (f) Gli autori della scala ipotizzano 2 fattori (INTRINS ed ESTRIN), mentre altri autori ritengono che il fattore ESTRIN sia invece composto da due sottofattori. Fai un'analisi fattoriale esplorativa, scegliendo il metodo e la rotazione e proponi una possibile analisi fattoriale confermativa (vincolata).
  - (g) Usando ReligxLisrel.sav prova a fare una analisi fattoriale esplorativa ordinale e confronta le varie soluzioni
5. La scala di autoritarismo RWA di Altemeier rivista in italiano da Roccato è misurata dagli item che vanno da R3 a R16.
  - (a) Verifica che non vi siano scale da ribaltare
  - (b) Se ce ne sono, ribaltale
  - (c) Calcola il punteggio di Autoritarismo
  - (d) Questo punteggio ha una relazione lineare con l'orientamento politico (D12) e si differenzia per fasce politiche (POLIT)?
  - (e) Ha ragione Roccato a ritenere che la scala sia monofattoriale?
6. Altre domande.
  - (a) Studia le variabili quantitative (vQT) e cerca un modello che possa spiegare il fondamentalismo.
  - (b) Cerca quale delle variabili qualitative (vQL) possano essere ragionevolmente usate come indipendenti e verifica se le quantitative si differenzino statisticamente rispetto alle qualitative
  - (c) In particolare verifica se il fondamentalismo, suddiviso per genere, fasce d'età e pratica religiosa segmenta il campione in gruppi statisticamente differenti
  - (d) Vi sono sottogruppi omogenei per fasce d'età e/o pratica religiosa?

## Comandi SPSS per MATT.sav

Nell'output di SPSS, la voce "Registro" contiene i comando eseguiti. Qui di seguito trovate i comandi come li ho eseguiti io.

```
COMPUTE eta = 2006 - D1 .
RECODE eta (18 thru 30=1) (31 thru 55=2) (56 thru Highest=3) INTO feta .
CROSSTABS
  /TABLES=D2 BY feta
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ
  /CELLS= COUNT ROW COLUMN
  /COUNT ROUND CELL .
MEANS
  TABLES=eta BY D1 feta
  /CELLS MEAN COUNT STDDEV .
COMPUTE D12N = D12 .
DO IF (D13 = 2) .
RECODE
  D12N (ELSE=SYSMIS) .
END IF .
MEANS
  TABLES=D12N BY D2 feta D6
  /CELLS MEAN COUNT STDDEV .
RECODE
  D12N
  (SYSMIS=4) (1 thru 3=1) (4 thru 7=2) (8 thru 10=3) INTO polit .

SAVE OUTFILE='Matt.sav' /COMPRESSED.
```