



Sostituisci Outliers

Custom Procedure - Dialog

Descrizione

Questa procedura sostituisce i valori mancanti con la media dei valori del gruppo a cui appartengono.

NB: la sintassi è ricavabile premendo il tasto “incolla” nella finestra di dialogo. La sintassi può essere personalizzata e modificata sulle specifiche esigenze dell’utente.

Input richiesti:

- variabile/i di raggruppamento casi.
- variabile per il calcolo della media del gruppo.
- percorso nel filesystem in cui salvare il dataset risultante dalla procedura e i dataset intermedi di calcolo.
- estremo inferiore e superiore inclusi definiti dall’utente per identificare un valore come valido.
(estremi entrambi obbligatori)

Output:

Un nuovo dataset con i valori mancanti sostituiti con la media dei soli casi validi del gruppo di appartenenza

Esempio

Aprire il file **SostituisciOutliers_data.sav**

Il file contiene 4 prodotti e 2 tipologie di prodotto e i relativi prezzi.

I prezzi sono da considerarsi validi se sono compresi tra i 100 e i 200 Euro (definito dall’utente).

I valori non validi saranno sostituiti con la media del gruppo di appartenenza (selezionato dall’utente come variabile singola o insieme di più variabili).

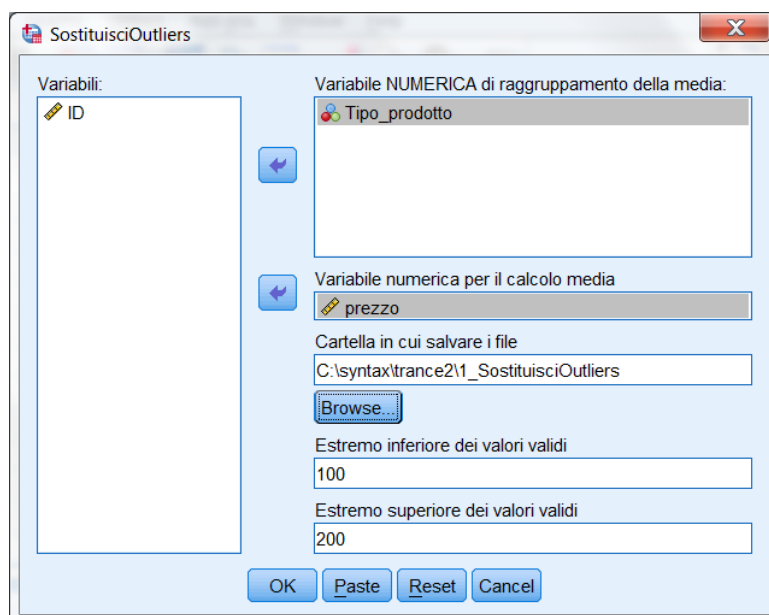
Il file contiene 3 variabili:

ID	Tipo_prodotto	prezzo
99	1	200
101	1	400
201	2	100
301	2	200

1. ID: corrisponde all'ID del prodotto - serial number
2. Tipo_prodotto: la tipologia del prodotto (1 o 2).
3. Prezzo prodotto.

In questo esempio sono state assegnati i seguenti parametri:

- Raggruppamento: Tipo_prodotto. (la media sarà calcolata per i casi validi dei gruppi 1 e 2).
- Media: Prezzo. (la media calcolata per il prezzo).
- Estremi valori validi: 100 e 200 (estremi inclusi).



Il risultato che si ottiene è il seguente:

Sono presenti 2 nuove variabili:

VALID_MEAN_PREZZO: contiene la media del gruppo per tipologia del prodotto.

PREZZO_NEW: il nuovo valore del prezzo quando il caso ha un valore non valido.

ID	Tipo_prodotto	prezzo	valid_mean_prezzo	prezzo_new
99	1	200	200,00	200,00
101	1	400	200,00	200,00
201	2	100	150,00	100,00
301	2	200	150,00	200,00

Il caso non valido è il secondo e il **PREZZO_NEW** è stato sostituito con la media dei prezzi validi della tipologia prodotto=1, che è pari a 200. Per gli altri valori, che sono validi, il **PREZZO_NEW** corrisponde al **PREZZO**.