Come si gestiscono eventuali dati mancanti nella "vista variabili"?
Nella vista variabili è possibile definire valori mancanti **aggiuntivi** rispetto al campo vuoto delle variabili numeriche, il cosiddetto valore “mancante di sistema”, che se la variabile è numerica è già definito di default.
In altre parole, per IBM SPSS Statistics il valore 0 e il valore mancante sono due cose diverse.

Come risolverebbe la questione dei valori che da 0,2 sono diventati 2000?
Con la funzione *Calcola con condizione*.
Nel caso visto in cui ad avere la virgola erano i valori stringa “<0,02” si può utilizzare l’opzione SE ed esplicitando una condizione sul simbolo “<” presente davanti al numero.
In pratica è stato usato questo comando: IF “Char.Substr(varstr,1,1) = ‘<’ THEN valore num./1000

Per dare la procedura più corretta di sostituzione (caso CONCETRAZIONE) quali sono i passi giusti per evitare i numeri in centinaia/migliaia generati sulla nuova variabile?
La cosa migliore sarebbe stata uniformare le cose nel file di partenza e sistemare le eventuali incoerenze tra ‘,’ e ‘.’ con un Trova e Sostituisci.
Se invece il problema si è trascinato in IBM SPSS Statistics si può risolvere attraverso la funzione *Calcola* usando l’opzione del calcolo condizionale (ovvero esplicitando la condizione SE).

Potreste consigliare dei testi di riferimento?
Ce ne sono moltissimi anche se riguardano soprattutto la parte di analisi e ben pochi contemplano la parte di preparazione dati che pure è importantissima.
Un testo che a me piace molto è “Biostatistics: The Bare Essentials with SPSS (Inglese)” Geoffrey R. Norman, David L. Streiner
Di cui credo potrebbe esistere anche una versione in italiano
Con la ristrutturazione, si costruisce un nuovo dataset o si trasforma quello originale?
Si costruisce un nuovo dataset a partire da quello originale

Dopo una ristrutturazione simile a quella operata è buona norma procedere con il calcolo delle correlazioni tra variabili?
Diciamo che la ristrutturazione si rende necessaria se voglio fare un’analisi di correlazione che non riuscirei a fare a partire dal formato mostrato.
Questo allo scopo di capire cosa?
L’analisi di correlazione serve ad evidenziare le relazioni esistenti tra le variabili.
Inoltre, quanto possono incidere nel calcolo le celle con dati mancanti e se in qualche misura devono essere sostituiti con qualche procedura automatica?
Questo è un tema molto delicato. Sicuramente i valori mancanti possono creare diversi problemi nelle analisi di regressione/correlazione.
L’ideale sarebbe evitare di considerare nella propria analisi variabili che abbiamo TROPPI valori mancanti (es. più del 40% dei valori).
In situazioni intermedie è possibile sostituire i valori mancanti con una media o magari stimando quei valori attraverso altre variabili.

Io sto lavorando su un dataset relativo ad un questionario che ha il valore 7 per le domande saltate. Come si fa a gestire questa casistica?
Vista Variabili. Definire il valore 7 come mancante discreto

E ‘possibile visualizzare le etichette dei valori di riferimento principali del boxplot? mediana, quartili, ...
Non come etichetta ma attraverso linee di riferimento

Come si trasforma in logaritmica?

La scala dell'asse Y del grafico a scatola può essere modificata da lineare a logaritmica attraverso il Chart Editor modificandone le Proprietà

Sulla raccolta visuale non è meglio fidarsi della divisione in percentile che arbitrariamente agire con la classificazione visuale?
A volte la raccolta visuale è più efficace della divisione tramite percentili perché talvolta dall'istogramma è possibile individuare dei "Natural Breaks" nella distribuzione dei dati che potrebbero delimitare più efficacemente dei percentili le classi di ns interesse.
Quando i Natural breaks non sono evidenti la classificazione basata sui percentili è probabilmente la scelta migliore.