



RESEARCH DATA CENTER

REX - Remote Execution / Sistema di elaborazione a distanza

Indicazioni per favorire il rilascio degli output del sistema di elaborazione a distanza REX

Novembre 2023



I ricercatori che lavorano con i microdati messi a disposizione dalla Banca d'Italia tramite il sistema di elaborazione a distanza sono tenuti a garantire che i risultati delle loro analisi rispettino gli standard di confidenzialità, ovvero, in sintesi, che non sia possibile risalire in alcun modo all'unità di base alla quale l'informazione è riferita. In assenza del rispetto della confidenzialità, riferito non solo alla singola elaborazione, ma anche all'insieme delle elaborazioni svolte da uno stesso utente, non sarà possibile fornire al ricercatore i risultati delle proprie elaborazioni.

Il tentativo di violare le condizioni di confidenzialità può condurre alla sospensione e alla revoca dell'utenza.

Si riporta di seguito una lista non esaustiva di regole e linee guida che favoriscono il rispetto dei requisiti di confidenzialità e il rilascio dei risultati delle elaborazioni e alle quali i ricercatori sono quindi tenuti ad attenersi. Il giudizio ultimo sulla possibilità di rilasciare l'output per la successiva pubblicazione è comunque del Research Data Center, che valuta il rispetto delle condizioni di confidenzialità su un complesso insieme di variabili.

Regole per il rispetto dei requisiti di confidenzialità

Per evitare di violare le regole di confidenzialità e, di conseguenza, non ottenere l'output richiesto, nella redazione dei programmi di elaborazione vanno evitati:

- qualsiasi calcolo che possa avere come risultato una singola unità statistica, inclusi i comandi che stampano un singolo record o parti di esso o che provano a esportare dati;
- la stampa a schermo di statistiche calcolate su un'unica unità statistica, inclusi minimo, massimo, mediana e quantili;
- la creazione di distribuzioni di frequenza su variabili continue;
- la stampa di qualsiasi tipo di grafico;
- la stampa a schermo dei residui e dei valori previsti e il calcolo delle loro statistiche d'ordine;
- comandi per connettere sorgenti esterne o cambiare la directory di lavoro;
- il calcolo di statistiche aggregate su meno di 3 unità statistiche; a tal fine, si richiede di mostrare nell'output il numero di osservazioni sottostanti ai calcoli.

Linee guida generali

I principi che seguono hanno lo scopo di rendere la lettura dell'output più semplice al fine di facilitare il processo di controllo dello stesso (e per questa via favorire il rilascio dei risultati delle elaborazioni):

- il codice sottoposto deve essere chiaro, commentato, di facile lettura (ben indentato) e le variabili devono essere definite con nomi autodescrittivi e non ambigui;
- il numero di job sottoposti e il volume dei risultati attesi deve essere limitato a quanto indispensabile per la ricerca. Output molto lunghi (superiori alle 2500 linee) saranno rifiutati;
- il numero di osservazioni incluse nei modelli statistici dovrà essere adeguatamente superiore al numero dei parametri;
- per modelli con output molto lunghi o un ampio numero di parametri (per esempio con molte dummy o effetti fissi) è consigliabile usare comandi per l'abbreviazione dell'output (ad esempio l'opzione "quietly" di Stata).

La lista di criteri sopra indicata è valida per ogni dataset e può essere soggetta a modifiche e integrazioni quando necessario. Per alcuni dataset specifici possono essere previsti vincoli addizionali.

Regole aggiuntive per specifici archivi

Tavole ancillari agli archivi con dati bancari (BANCI)

Nel caso di elaborazioni effettuate associando la mappa dei gruppi bancari, l'unità statistica di riferimento diventa il gruppo (variabile `idba_capogruppo_ps`). Pertanto, il calcolo di statistiche aggregate deve interessare almeno 3 gruppi.