



do criteri che ottimizzassero le prestazioni dell'analisi e la sua adattabilità ad eventuali integrazioni future, secondo un modello entità - relazioni. Ad esempio nel caso dell'Anagrafe sono stati creati alcuni flag che permettano di individuare particolari categorie di contribuenti in modo da facilitarne in seguito l'analisi o eventualmente l'esclusione dall'analisi stessa. Alcuni esempi sono cittadini irreperibili al censimento 2001, cittadini al di sopra dei 110 anni di età, residenti in convivenze, cittadini senza fissa dimora ecc. Si è quindi proceduto alla costruzione delle famiglie: una volta riorganizzati i dati relativi all'Anagrafe sono state ricostruite le famiglie indicate al suo interno (per mezzo dell'apposito numero di famiglia) provvedendo a creare una tabella in cui possano essere registrati i dati aggregati necessari all'analisi. I dati sono stati infine aggregati in base al codice fiscale dei dichiaranti, eliminando tutti i dati che permettano l'identificazione del singolo individuo e utilizzando codici che permettano di risalire all'origine dei dati solo attraverso apposite tabelle di conversione. Ottenuti i dati

aggregati è stato infine possibile estendere la fase di verifica in modo trasversale, confrontando cioè le informazioni presenti nei diversi anni per verificare posizioni considerate anomale.

“Molteplici possono essere le tipologie di analisi da condurre sui dati sia singolarmente considerati sia integrati; – sottolinea **Paolo Mariani**, Facoltà di Scienze Statistiche, Dipartimento di Statistica, Statistica Aziendale - Unimib – ne è un esempio lo spaccato che è già possibile ricavare della società milanese”. Eliminati i casi anomali, la popolazione milanese al 31.12.2000 contava 1.278.662 individui, il 52,6% dei quali donne. Più della metà dei cittadini milanesi hanno età compresa tra i 26 e i 60 anni, i giovani rappresentano il 20% mentre i cittadini più anziani quasi il 30% della popolazione totale.

Incrociando i dati anagrafici con quelli provenienti dal Ministero dell'Economia e delle Finanze, emerge che gli 847.874 dichiaranti rappresentano il 66% della totalità dei cittadini milanesi. Peraltro è possibile che anche i cittadini rimanenti possiedano reddito, perché il data warehouse AMeRiCA non include pensioni



sociali, accompagnamenti o altre forme di reddito non tassate, né i redditi bassi che non prevedono l'obbligo di dichiarazione.

A presentare il modello Unico sono più della metà dei cittadini milanesi e tale proporzione è più elevata per gli uomini. Complessivamente il numero di dichiarazioni presentate dalle donne è superiore a quello per gli uomini ma, il reddito imponibile ad esse associato rappresenta solo il 36% di quello com-



pletivo. Coloro che presentano il modello Unico possiedono mediamente redditi più alti anche se con una certa variabilità interna; al contrario, coloro che presentano gli altri modelli dichiarativi dimostrano di produrre redditi inferiori ma presentano, al loro interno una situazione più omogenea. La maggior parte dei dichiaranti produce principalmente redditi provenienti da lavoro dipendente; confrontando la distribuzione della provenienza per gli uomini rispetto a quella per le donne emerge che in misura maggiore quest'ultime possiedono solo redditi patrimoniali.

Paolo Mariani,
Facoltà di Scienze
Statistiche, Dipartimento
di Statistica, Statistica
Aziendale - Unimib

Mario Mezzananza,
Facoltà di Scienze
Statistiche, Organizzazione
dei Sistemi Informativi
Aziendali - Unimib

Biancamaria Zavarella,
Facoltà di Scienze
Statistiche, Dipartimento
di Statistica, Statistica
Economica - Unimib