

нява няколко техники и метода както за редактиране, така и за импутация. Очевидно е обаче, че не може да се намери лесно решение, защото въпросната дейност се случва на ранния етап, а потребителят трябва внимателно да обмисли как различните части от обработването на данни се съчетават помежду си, за да избегне излишни усложнения и загуба на време.

Интернет страницата на EUREDIT съдържа много доклади и някои софтуерни продукти за разработване и оценяване на нови методи за редактиране и импутация (<http://www.cs.york.ac.uk/euredit/>).

### 3.1. Общи статистически програми

Разширените статистически програми SAS, SPSS или STATA предлагат много добри възможности за анализ на данни - от обикновените описателни статистически характеристики (едностранни разпределения, таблици на честотите, брой на липсващите/наличните стойности и т.н.) до сложните многовариантни техники. Те съставят първата част от редактирането и импутацията: проверка на данните и локализиране на екстремните стойности и грешките. Тези програми обаче не предоставят непременно самостоятелна колекция от методи за редактиране и импутация. Съществуват определени методи като регресионните модели или множествена импутация-SAS, но един цялостен подход ще изисква програмиране, особено що се отнася до процесите на редактиране. А колкото повече процеси на автоматизирано редактиране и импутация се изискват, толкова по-мощна ще бъде задачата. От прагматична гледна точка положително би било обработката на данните да се запази в един-единствен софтуерен продукт, но в крайна сметка решението трябва да се вземе на основата на потребностите.

Независимо от това обаче има софтуерни продукти за редактиране и импутация, които са разработени за SAS или SPSS. Добри примери са CANCEIS и GEIS (разработени в Статистическата служба на Канада) и CONCORD/SCIA (разработени в Статистическата служба на Италия). Всички те работят в пакета SAS.

Софтуерният продукт SAS/STAT съдържа множествена процедура за импутация, която може да бъде използвана и за единична импутация, ако се наложи. Следващите методи за импутация са пряко достъпни: регресия, прогнозно средно съчетаване, методи за тенденцията на резултатите и методи Markov-Chain Monte Carlo (MCMC) за непрекъснати променливи, както и методи за дискриминантна функция или логистична регресия за категорийни променливи.

Софтуерният продукт SPSS v. 14 съдържа два отделни модула: Потвърждаване (контрол) на данни (Data Validation) и Анализ на липсващи стой-