

човек на същата възраст и от същия пол. Обикновено се прилага някой метод, използващ случайния принцип, за да се намерят адекватни донори.

Cold deck често се използва в бизнес статистиката: липсващата към момента стойност се импутира от последната наблюдавана стойност от същата единица.

4. Импутация на базата на модел. Обикновено регресивен модел (или логистична регресия), при който импутираните стойности са предполагаеми.

Дава сравнително добри резултати при положение, че първоначалният модел пасва добре. Следва да се комбинира със случайни грешки за придаване на по-голяма вариация на набора от данни.

5. Множествена импутация. Импутира се набор от данни с помощта на методи, основани на модели (например 5-10 пъти), и всички импутирани набори от данни се анализират. Този метод е в добро съответствие със стандартната статистическа теория, но изисква програми за специален анализ за изчисляване на вариациите.

Не трябва никога да се забравя, че импутацията зависи от наблюдаваните стойности, а по този начин би се омаловажила извадковата вариация, ако формулите не се коригират за нея. Важно е също да се обозначават (flag) всички данни и променливи, които са били редактирани или импутирани. Индикатор за обозначените редактирани и импутирани данни (Flag-indicator) улеснява потребителите при оценяването на данните дали искат да анализират импутираните стойности, или не, за да се изчислят правилно вариациите. Освен това той служи като показател за качеството: пропорцията на импутираните стойности следва да се докладва.

3. Софтуер за редактиране на данни от изследване и импутация

Особено важно е да се поддържа доброто качество на статистическите данни. Всичко започва от оценката на получените данни без оглед на техния източник. Първата фаза обикновено се нарича "проверка на данните". След първия анализ може да има нужда да се коригират някои неправилни или неправдоподобни стойности, да се импутират някои липсващи детайли и т.н. с оглед да се запази достатъчно добра матрица с данни за по-нататъшна употреба - например за табулация и анализ. Нужните на този етап техники се наричат "редактиране" и "импутация". През последните десетина години към тези въпроси се проявява голям интерес и в резултат на това са проучени средства, насочени към предоставящите статистически данни и потребителите, за защита на резултатите. Най-голямото усилие беше направено с научноизследователската програма EUREDIT, която срав-