

1.4. Всеобхватен среден импутационен метод.

Този метод се прилага за импутиране на липсващи стойности на количествени или категорийни признаци (в зависимост от вида на признака). За импутация се използва средната аритметична стойност при количествени значения на признака или модата на признака при категорийни значения на мястото на липсващата стойност. Нека се опитаме да разгледаме категорийните и количествените значения за липсващите данни на признаците "наличие на шофьорска книжка" и "средномесечен разход" (табл. 1). При признака "средномесечен разход" средното значение се изчислява по формула (3) след изключване на единицата с липсващо значение, или се получава 228.64 лева. Ако се опитаме да припишем стойността на мястото на липсващото значение, ще се получи доста нереално положение (средномесечният доход на единица № 7 е 40 лева).

Модата приема характеристиката на онова значение на признака, което се наблюдава при най-много единици от изучаваната съвкупност (Мишев, Цветков, 1998, стр. 67). Модата на признака „наличие на шофьорска книжка" е "да" и ако се импутира на единица № 1 по-късно, съвкупността ще се нуждае от редактиране (тъй като на 14-годишна възраст лицето не може да притежава шофьорска книжка). Както се вижда, този метод не е подходящ за изследваната съвкупност.

1.5. Класическа средна импутация.

При този метод за разлика от предходния изследваната съвкупност се разделя на страти (табл. 3). Намира се средната аритметична на съответната страта при данни на интервалната скала по формула (3) или модата на съответната страта при данни на слабите скали след изключване на единицата с липсваща стойност на признака. Получените значения се импутират съответно в същите страти. При този метод се получава "не" на наличието на шофьорска книжка в страта № 1 и средна алгебрична стойност 38.33 лв. за средномесечен разход в същата страта, които могат да се припишат съответно на единица № 1 и единица № 7 в тази страта. От разгледаните методи за импутиране за тази съвкупност засега този показва най-добри резултати. Той се използва от отдел "Статистика на промишлеността" на НСИ за импутиране на липсващи данни при изчисляване на индекса на промишленото производство.

1.6. Регресионна импутация.

Чрез регресионен модел можем да изследваме променливата Y_i , т. е. ако имаме 5 стойности на променливата - Y_1, Y_2, Y_3, Y_4 и Y_5 , където