

МЕТОДЫ ДЛЯ ЗАМЕЩЕНИЯ ОТСУТСТВУЮЩИХ ДАННЫХ

Станислав Георгиев*

РЕЗЮМЕ В статье уделено внимание проблеме с отсутствующими данными в статистических исследованиях, обыкновенно полученные и необработанные первичные данные нуждаются от дополнительного рассмотрения и устранения некоторых основных ошибок и пропусков. Эти ошибки могут появиться в результате от: недостаточных или неясных методологических указаний; плохой подготовки анкетора; ошибочного мнения некоторых вопросов со стороны анкетированных; и соответственно получения неправильных ответов (или неполучения ответов); поради ошибочного введения первичных данных и пр. Рассматриваются различные методы для устранения этих ошибок путем замещения отсутствующих данных. Описаны различные методы для замещения отсутствующих данных в зависимости от:

- Вуда признака, чьи значения замещаются;
 - Признаки с количественными значениями;
 - Признаки с качественными значениями;
- Способ, с помощью которого замещаются отсутствующие значения признака;
- На основании "донора" (выбирается определенным путем подходящая единица которая является "донором" и с ее характеристиками замещаются отсутствующие значения);
 - На основании математической и логической процедуры (например когда находится мода при категориальных признаках) находится характеристика, которая используется для замещения отсутствующей стоимости.

Рассмотрен пример с целью лучшего просмотра всех описанных методов.

METHODS FOR IMPUTING MISSING DATA

Stanislav Georgiev*

SUMMARY The article considers the problem with the missing data in statistical surveys.

Usually the primary data, received and not yet processed, needs additional review and elimination of some basic mistakes and lapses. These mistakes may result from inadequate and obscure methodological instructions, poor preparation of the interviewers, questions misunderstood by the respondents and consequently incorrect answers received (also lack of response), incorrect filling of the primary data, etc. Different methods for eliminating these mistakes through imputing missing data are considered. Different methods for imputing missing data are described, depending on:

- Type of characteristic with imputed value;
 - Characteristics of quantitative values;
 - Characteristics of qualitative values;
 - Ways of imputing the missing values of the characteristic;
 - By a donor (an appropriate unit is selected in a definite way and this unit appears as a donor and its characteristics are imputed in the place of the missing values);
 - Through a mathematical or logical procedure (for example, when the mode of categorical characteristics is found) a characteristic for imputing the missing value is discovered.
- An example is given for better demonstration of all methods described.

* Senior expert at the Department "Statistical Methods and Standards" in the National Statistical Institute, Sofia, 10 "6 September" Str.; e-mail: sgeorgiev@nsi.bg .