

$$\bar{y} = \frac{1}{n}(n_1 \bar{y}_1 + n_2 \bar{y}_2),$$

където:

\bar{y} е средната величина на изучаваната променлива за цялата съвкупност;

\bar{y}_1 - средната стойност на променливата в съвкупността на отговорилите;

\bar{y}_2 - неизвестната средна в съвкупността на попадналите единици в изследването, но неотговорили.

Ако за оценка на \bar{y} се използва наличната информация от \bar{y}_1 , то изместването на действителната величина може да се представи чрез:

$$\text{изм}(\bar{y}_{\bar{y}_1}) = \bar{y} - \bar{y}_1 = \frac{n_2}{n}(\bar{y}_2 - \bar{y}_1).$$

Общото изместване, или, както някои автори го наричат, общата грешка, дължаща се на липсата на отговори на респондентите, е функция едновременно на относителния дял на неотговорилите и различието в техните характеристики по изследвания признак. Очевидно е, че при по-голям дял на неотговорилите малките различия в средните величини се мултиплицират и се отразяват фатално върху крайния резултат от изследването.

Обект на поява на липсващи стойности, разбира се, са не само извадковите изследвания, а също и изчерпателните. В голяма степен може да се предполага, че подобен проблем е много по-значим при мащабните изследвания като преброяванията на населението поради факта, че неотговорилите могат да представляват специфична част от изследваната съвкупност.

В следващите страници се представят 1) основните причини за поява на липсващи стойности, 2) теоретичната основа за справяне с проблема на липсващи данни, 3) решенията, които се предприемат от различни национални статистически институции в Америка и Европа, 4) тенденциите към внедряване на единни софтуерни системи за третиране на проблема и сравненията на някои от тях.

ПРИЧИНИ ЗА ПОЯВА НА ЛИПСВАЩИ СТОЙНОСТИ

Появата на които и да е от трите класа липсващи данни е свързана със специфични причини. Обикновено те са пряко свързани с подготовката и провеждането на изследването и неговата техническа осигуреност.

1. Липса на обхват на съвкупността, от която се избира извадката. Част от единиците, обект на изследване, не са представени в списъците, от които се избира извадката или по които се правят наблюденията. Това е пряко свързано с две основни причини - непълнота на списъците или не-