

Обемът на извадката (n) се определя по формулата:

$$n = \frac{Z^2 \frac{q}{p} 100^2}{(\Delta_p \%)^2 + \frac{Z^2 \frac{q}{p} 100^2}{N}} [1 + \delta (\bar{n} - 1)],$$

където:

p е относителният дял на показателя, за който се изчислява обемът на извадката;

$$q = 1 - p;$$

$\Delta_p \%$ - максималната относителна грешка, с която се отклонява оценяваният относителен дял от този в генералната съвкупност;

Z - множителят, съответстващ на приетата гаранционна вероятност, при която $Z=1.96$;

δ - величината на вътрешногнездовата корелация;

\bar{n} - средният брой единици (лица) в едно гнездо (жилище);

N - обемът на генералната съвкупност (преброеното население).

За относителен дял $p = 0.004$ при $\Delta_p \% = 10\%$, $Z = 1.96$, $\delta = 0.3$, $\bar{n} = 2.7$ и $N = 8000000$ обемът на извадката е 142734 лица. Този обем на извадката е 1.784% от преброеното население.

От практически съображения при формирането на извадката се предлага обемът на извадката да бъде 160 хил. лица, т. е. 2.0% от генералната съвкупност. Такъв беше дялът на извадката и при репрезентативната разработка на данните от преброяването на населението към 4 декември 1992 г., като обемът на извадката бе 169466 лица. Намаляването сега на обема на извадката с 5.4% ще доведе до увеличаване на максималните стохастични грешки на $p(\Delta_p)$ с 2.9%, но максималните абсолютни грешки на случаите в генералната съвкупност (Δ_A) ще намалее с 3.0%, тъй като множителят N към 1 март 2001 г. е с 5.7% по-малък от този към 4 декември 1992 г.

При съставяне на програмата на разработката (таблици с разпределение на броя на населението по разновидностите на един и повече признака) следва да се вземат предвид стохастичните грешки, с които ще бъдат обременени показателите (подсъвкупностите на населението по определена разновидност на даден признак или комбинациите от разновидности на два и повече признака), получени от извадката, като оценки на съответните показатели в генералната съвкупност. За тази цел в табл. 1 са представени показатели, характеризиращи стохастичната точност на различни относителни дялове, установени от извадката, като оценки на генералната съвкупност.

Стандартната грешка S_p на p е изчислена по формулата: