

## **Сравнение на различни модели на извадката според разходите за тяхното реализиране**

**Доц. д-р Калоян Харалампиев**

*Резюме: В текста се сравняват два модела на извадки, посредством сравнението на две реално проведени изследвания с различни модели на извадката. Предмет на сравнението са разходите да достигане до населените места в извадката (дневни, пътни и квартирни пари за анкетьорите). За целите на сравнителния анализ се преизчислява броят на населените места, в които извадката е разпръсната, при равни други условия, т.е. при условие, че общият обем на двете извадки е един и същ. Получените резултати показват, че броят на градовете в извадката е приблизително еднакъв при двата сравнявани модела на извадката. Драматичната разлика е при броя на селата в извадката – при модела на простия случаен подбор в извадката попадат много повече села, като в по-голямата част от тях трябва да се анкетира само по един респондент.*

## **Comparison of different sample models according to the cost of their realization**

**Assoc. Prof. Kaloyan Haralampiev, PhD**

*Abstract: In the paper two sample models are compared by comparing two real studies with different sample models. The subject of the comparison is the cost for the reaching to the settlements in the sample (daily, travel and accommodation money for the interviewers). For the purpose of the comparative analysis, the number of the settlements in the sample is recalculated under the equal other conditions, i.e. provided that the sizes of the two samples are the same. The obtained results show that the number of the cities in the sample is approximately the same for the two compared sampling models. The drastic difference is the*

*number of the villages in the sample – when the sample model is random choice then the sample includes many more villages, with only one respondent in the majority of them.*

Ние с Милена Стефанова имахме уникалния шанс да участваме в две изследвания, които се проведеха по едно и също време, от един и същи изследователски екип и от един и същи анкетъорски екип. Но с различни модели на извадката. Това ни позволява да сравним различните модели на извадката според тяхната цена и според логистичните трудности при организирането и управлението на теренната работа.

Първото изследване беше по поръчка на Световната банка. Възложителят имаше три изрични изисквания – извадката да бъде с обем от 1000 респондента, да обхваща клиенти на всички ВиК оператори в страната и да бъде пропорционална по етнически групи. Последните две изисквания по естествен път определиха стратифицирания подбор като модел на извадката. За да организираме и да управляваме по-лесно теренната работа, както и за да спестим ресурси, комбинирахме стратифицирания подбор с гнездови. По този начин планираната извадка се оказа разпръсната в 124 населени места.

Второто изследване беше по поръчка на Агенцията по заетостта. Ние получихме от Агенцията по заетостта списъци на лицата, попадащи в генералната съвкупност, но в тези списъци липсваше информация за населените места. Нашето искане да получим информация за населеното място на всяко лице получи отказ. Вместо това Агенцията по заетостта предложи ние първо да формираме извадката, а след това те да ни дадат адресите и телефоните, но само на тези лица, които са попаднали в извадката. В този случай единственият възможен модел на извадката стана простият случаен подбор. Планираната извадка беше с обем 2340 респондента и се оказа разпръсната в 752 населени места.

Логистичните проблеми при осъществяването на модела на простата случайна извадка са отдавна известни в статистическата теория и практика. Съйкова и Чакалов посочват, че едно „неудобство е свързано с факта, че при формирането на извадката единиците в нея са възможно най-силно разпръснати по територията, на която се намира генералната съвкупност. От това се пораждат допълнителни организационно-практически проблеми, както и оскъпяване на изследването.“ (Съйкова, Чакалов 1977: 47). За подобни проблеми говори и Ивайло Парчев, като посочва и конкретни числови стойности: „На практика съвсем не е невъзможно да се състави проста случайна извадка, представителна за

пълнолетното население на България. Ако имаме достъп до масива от данни на системата ЕСГРАОН, извадката може да се осъществи за секунди. Трудностите идват при анкетирането. Простата случайна извадка е много силно разпръсната по цялата територия на страната. Например една проста случайна извадка от 1212 души, съставена от автора специално за този пример, се оказа разпиляна в общо 496 населени места. В 384 от тях би трябвало да се анкетира само един човек, в 48 – по двама, в 19 – по трима, и така нататък. От научна гледна точка това е прекрасно, но на практика би довело до неприемливо високи пътни разходи за анкетьорите.“ (Парчев 1998: 112).

Като средство за решаването на логистичните проблеми обикновено се посочва моделът на гнездовата извадка. „Най-големите предимства на гнездовите извадки са от организационно-практическо естество, които непосредствено рефлектират върху финансовите, трудовите, разходи на време и други.“ (Съйкова, Чакалов 1977: 48)

Нашите две изследвания от една страна недвусмислено потвърждават тези изводи, а от друга страна могат да обогатят статистическата теория и практика с допълнителни аргументи в полза на гнездовите извадки.

Видно е, че броят на населените места при изследването, реализирано чрез проста случайна извадка, е по-голям в сравнение със стратифицираната гнездова извадка, но пък и извадката също е по-голяма. За да можем да сравняваме двете извадки, се налагат известни преизчисления, така че сравнението да бъде направено „при равни други условия“.

Най-напред изчисляваме средния брой на респондентите в населено място. За моделът на стратифицираната гнездова извадка този среден брой е  $\frac{1000}{124} = 8,1$ , а за моделът на простата случайна извадка е  $\frac{2340}{752} = 3,1$ . Още отгук се вижда, че с едни и същи разходи за достигане до едно конкретно населено място (дневни, пътни и квартирни пари за анкетьорите), при простия случаен подбор ще се реализират 2,6 пъти по-малко анкети.

Също така можем да преизчислим в колко населени места би се провела простата случайна извадка, ако нейният обем би бил също 1000 респондента –  $\frac{1000}{3,1} = 321$ . Това е със 197 населени места повече спрямо модела на стратифицираната гнездова извадка или отново 2,6 пъти повече. А това означава, че, при равни други условия, разходите за достигане до съответните населени места също биха били 2,6 пъти повече.

Но когато говорим за разходите за достигане до съответните населени места, трябва да държим сметка и за типа населено място. Като правило провеждането на анкети в селата е по-скъпо, отколкото в градовете. Това е така, защото обикновено анкетъорската мрежа е разположена в (големите) градове, а до селата трябва да се пътува специално за целите на анкетирането. При това публичният транспорт до по-отдалечените села е по на рядко, а и в тях липсват хотели, в които анкетъорът да пренощува при нужда. Затова много често се налага до едно и също село да се пътува няколко пъти, в няколко различни дни, най-често с автомобил, докато се реализира съответната анкета, а понякога всички пътувания се оказват напразни, защото анкетата така или иначе не може да се реализира.

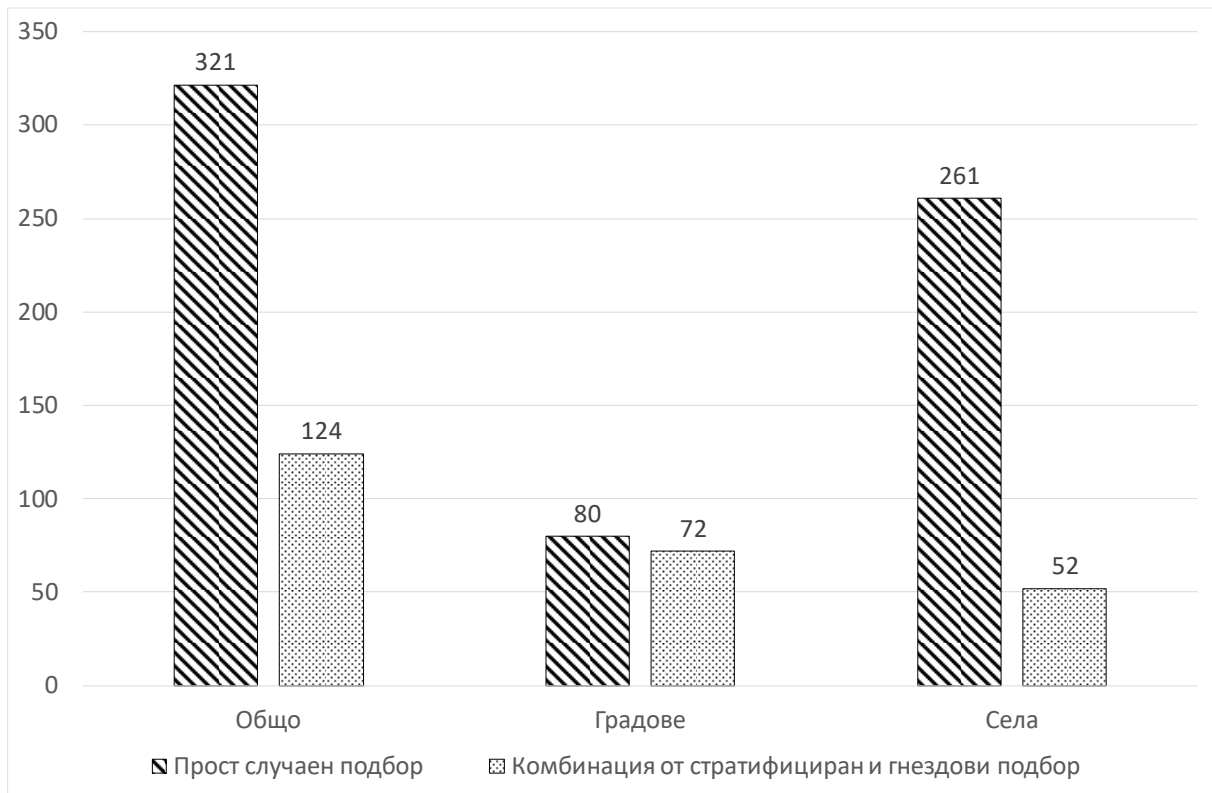
Преизчислението на броя на населените места в градовете и селата е систематизирано в Таблица 1 и Фиг. 1.

Таблица 1. Разпределение на извадките по тип населено място

Тип населено място	Град		Село	
	Прост случаен подбор	Комбинация от стратифициран и гнездови подбор	Прост случаен подбор	Комбинация от стратифициран и гнездови подбор
Брой респонденти	1611	658	729	342
Брой населени места	196	72	556	52
Среден брой респонденти в населено място	8,2	9,1	1,3	6,6
Преизчислен брой населени места	80	72	261	52

Източник: собствени изчисления

Фиг. 1. Преизчислен брой на населените места по тип населено място



Източник: собствени изчисления.

Резултатите от Таблица 1 и Фиг. 1 показват, че разходите за достигане до градовете в извадката са приблизително еднакви и при двата модела на извадката. При равни други условия, при модела на простия случаен подбор градовете в извадката са само 1,1 пъти повече отколкото при модела на стратифицираната гнездова извадка (съответно 80 спрямо 72 града), следователно и разходите за достигане до градовете ще са също 1,1 пъти по-високи.

Не така стоят нещата обаче при разходите за достигане до селата в извадката. При равни други условия, при модела на простия случаен подбор селата в извадката са 5,0 пъти повече отколкото при модела на стратифицираната гнездова извадка (съответно 261 спрямо 52 села), следователно и разходите за достигане до селата ще са също 5,0 пъти по-високи.

Основната причина за тази драстична разлика е в средния брой респонденти в населено място. При модела на стратифицираната гнездова извадка в селата трябва да се анкетираят средно по 6,6 респондента, докато при модела на простия случаен подбор в селата трябва

да се анкетира средно по 1,3 респондента. На практика това означава, че при модела на простия случаен подбор в повечето села трябва да бъде анкетиран само един респондент, т.е. всички разходи за достигането до тези села (дневни, пътни и квартирни пари) се правят само за получаването на една попълнена анкетна карта.

В резултат на всичко казано до тук съвсем не е неочаквано, че при изследването, поръчано от Световната банка, реализираната извадка покри всички планирани 124 населени места, докато при изследването, поръчано от Агенцията по заетостта, реализираната извадка покри 158 от 196 града и 320 от 556 села. Тоест стратифицираната гнездова извадка покри 100,0% от планираните населени места, докато простата случайна извадка покри 80,6% от планираните градове и 57,6% от планираните села. Това не означава, че в останалите 19,4% от планираните градове и 42,4% от планираните села не са изпращани анкетъори, а означава, че анкетъорите са посетили тези населени места, някои от тях и по повече от един път, но направените разходи за достигането до съответните населени места не са довели до получаване на нито една попълнена анкетна карта.

Можем да наречем разходите за достигане до населеното място „постоянни разходи“, тъй като те се правят така или иначе, независимо от броя на реализираните анкети. А парите, които анкетъорите получават за реализираните анкети, можем да наречем „променливи разходи“, тъй като те зависят от броя на реализираните анкети.

В такъв случай, при равни други условия, моделът на простия случаен подбор повишава постоянните разходи на изследването, без да влияе върху променливите разходи. Това повишаване на постоянните разходи обаче е за сметка основно на селата, където, при равни други условия, при модела на простия случаен подбор постоянните разходи са много по-големи, тъй като разходите се правят за много по-малко реализирани анкети. При равни други условия, в градовете постоянните разходи на двата модела на извадката са приблизително еднакви.

## **Библиография**

Парчев, И. 1998. „Избор на партия, избор на президент. Осем етюда върху една таблица“. София: Статистическо издателство и печатница при НСИ

Съйкова, И., Б. Чакалов. 1977. „Методология и методика на социологическите изследвания“. София: Наука и изкуство

Калоян Харалампиев е доктор по икономика, специалност „Статистика и демография”. Доцент в катедра „Социология” на Философски факултет на СУ „Св. Климент Охридски”, преподавател по Статистика и по SPSS на студенти от специалности „Социология“, „Психология“ и „Публична администрация“. Дисертация „Трудов потенциал на населението на България – минало, настояще и бъдеще” (2003). Научни интереси основно в областта на теорията на вероятностите, бейсовската статистика, демографската статистика и статистическия софтуер.

Основни публикации: „Трудов потенциал на населението на България (1992-2001)“ (2003), „Нетрадиционен поглед върху традиционни статистически проблеми“ (2004), „Рискови криминогенни и виктимогенни фактори при децата“ (в съавторство) (2010), „Политики за наука в Софийския университет „Св. Климент Охридски““ (в съавторство) (2012), „Доклад за резултатите от националното изследване на условията на труд в България“ (в съавторство) (2012), „Въведение в основните статистически методи за анализ“ (учебник) (2003, 2012), „IBM SPSS – Статистически решения на приложни изследователски задачи“ (учебно помагало) (2007, 2012), „Работа с данни с SPSS“ (учебно помагало) (2009).