



ИНСТИТУТ ПО ФИЛОСОФИЯ И СОЦИОЛОГИЯ- БАН
Проект с Договор № BGo5SFP001-3.004-0006-Co1 „Обучение на докторанти
за изследване и трансформиране на пазара на труда съобразно нуждите на
Индустрия 5.0. и двойния преход“,
по процедурата за предоставяне на безвъзмездна финансова помощ
BGo5SFP001-3.004 „Подкрепа за развитие на проектна докторантура“ по
Програма „Образование“ 2021 – 2027

Формиране на извадката за представителни емпирични изследвания. Основни подходи, принципи и решения

Член кореспондент Атанас Атанасов

02.10.2025 г.



Кратко въведение

- **Необходимост от използването на извадковите изследвания** за проучване на различни проблеми в социалната практика и всекидневния живот
- **Извадковите изследвания** - ефективен инструмент за набиране и анализ на разнообразна емпирична информация за различни явления и процеси в икономиката, социалната сфера, образованието, културата, политическото управление на страната и много други области на обществения живот
- **Изтъкнат български статистик: проф. Венец Цонев** „Основи на репрезентативното изучаване“ (1958)

Институт по социология при БАН

Секция „Методика и организация на ЕСИ“

- Извадковите изследвания – това е занимание за цял живот
- Не само знанията, но и натрупаният опит по планирането, организирането и провеждането на извадкови изследвания има много голямо значение за постигането на успехи и резултати в изследователската работа
- Извадковите изследвания и гарантирането на представителност на набраната емпирична информация

**Подход и основни принципи при
формирането на представителни извадки за
емпирични изследвания.
Изисквания за представителна извадка**

Каква е целта на емпиричното изследване?

- Да се получи представителна информация за изследвания обект и на основата на нейния анализ да се направят обобщения и заключения за неговото състояние и развитие
- Възниква въпросът: ще се използва ли извадковият метод на изследване или може да се направи изчерпателно изследване?
- Генерални съвкупности
- Извадка
- Кога извадката е представителна?
- Що е то „правилен подбор“?

Методиката на извадката - основен елемент на извадковото изследване

- **Кога започва** изграждането на методиката на извадката при подготовка на изследването?
- **Методиката на извадката е функция** от решаването на основните методологически и методически въпроси на изследването, т.е. по-скоро тя зависи от тях, а не обратното
- **Два етапа в изграждането на методиката на извадката:**
 - 1) **планиране на извадката**
 - 2) **формиране на извадката**
- Представителна извадка
- Непредставителна извадка
- Квотната извадка?

Първи стъпки на реализация на извадката

Ясно дефиниране на генералната съвкупност

Разграничаваме **различни основи за формиране на извадката**, които могат да се класифицират по следния начин:

- списъци на отделни лица, в които се обхваща цялото население или определена част от него, изготвени за потребностите на едни или други административни органи;
- съвкупността от въпросниците за лицата, попълвани при преброяване на населението;
- списъци на домакинства или жилища на дадена територия;
- териториално-устройствени планове на градски населени места;
- географски карти на селски райони;
- списъци на градове, села и административно-териториални единици, включващи разнообразни данни за тях"
(Иейтс 1965:97)

Подходи при формиране на извадката за теренното изследване

Когато обект на изследване е населението, независимо дали става дума за цялото население или на различни части от него, тогава като основа за формиране на извадката могат да се използват следните източници:

- списъци на отделни лица;
- списъци на групи лица, като семейства, домакинства, жилища;
- списъци на териториални единици, като преброителни (статистически) участъци, избирателни секции, населени места;
- планове (план-указатели) на населени места, като градове, села, квартали, райони, улици.



Формиране на извадката - традиционен подход

Традиционно формирането на извадката се извършва на основата на списъци, изготвени за други административни цели

Недостатъци на списъците (Владов 1964:175-176):

- Неточност
- Непълнота
- Повторения
- Несъответствия
- Остарялост
- Неадекватност

Идеалната генерална съвкупност – възможна ли е тя?

Какво се случва в реалността – три възможни ситуации

Проблеми с точността на обема на генералните съвкупности, използващи изборни списъци

Псевдотериториални структури на генералната съвкупност

Формиране на извадката - търсене на нов подход

Практиката на теренните изследвания подсказва, че е по-добре да се откажем от използването на предварително подготвени към някакъв предходен момент списъци, а формирането на извадката да се прави от съществуващите (наличните) на място единици – най-вече това са жилища или адреси, където тези единици могат да се намерят по време на изследването

Едно добро решение също е изборът на дадена улица и последователното или чрез определена стъпка обхождане на жилищата или адресите по нея. Ако улицата не стига, за да се набере необходимият брой жилища или адреси, се продължава по съседната улица

Как да се прави този подбор? Две възможности най-често се прилагат в практиката на теренните изследвания у нас:

- а) метод на най-близката (предстояща или минала) рождена дата;
- б) схема на Лесли Киш (Leslie Kish)

Практически решения

Troldal and Carter (1964) и Brayant (1975) предлагат **подобрения** в схемата на Киш, а именно след кратко въведение в целта на изследването, анкетьорът задава **два въпроса на лицето, с което разговаря:**

а) Колко души на 18 и повече навършени години живеят в това жилище (вкл. лицето, с което разговаря)?

б) Колко от тях са мъже/жени?

На тази основа са разработени **четири варианта за подбор**, като всеки от тях стартира със задаването на горните два въпроса

При работата на терена четирите варианта на схемата за подбор **се ротират** така, че да се избегне диспропорционален брой на мъже и жени, особено при домакинствата с по-малък брой

ПЪРВИ ВАРИАНТ НА СХЕМАТА ЗА ПОДБОР НА ЛИЦА ЗА ИЗСЛЕДВАНЕ В ДАДЕНО ЖИЛИЩЕ

		Брой на лицата на 18 и повече години в жилището			
		1	2	3	4 и повече
Брой на жените на 18 и повече години в жилището	0	мъжът	по-възрастният мъж	най-младият мъж	най-младият мъж
	1	жената	жената	жената	най-възрастният мъж
	2	•	по-възрастната жена	най-младата жена	най-младата жена
	3	•	•	най-младата жена	най-възрастната жена
	4 и повече	•	•	•	най-възрастната жена

Заключение върху избора на подход при формиране на извадката на последната степен

Методическата ефективност може да се търси в следните две направления:

Първото – минимизиране на недообхвата при избора на единиците за формиране на извадката поради ненамиране на терена определените за изследване единици, когато се осъществява техният подбор за включване в извадката

Второто – максимално улесняване на организацията на изследването на терена при формиране на извадката

Предимствата на подхода на Лесли-Киш

Как стоят нещата при интернет изследванията?

Основният въпрос тук е има ясна дефиниция на генералната съвкупност

- Знае ли се при едно конкретно интернет изследване **какъв е обемът** (колко единици наброява генералната съвкупност) и **каква е структурата** по различни признаци като **пол , възраст, образование** и т.н. ?
- Дали обемът и структурата на интернет съвкупността отговаря на дадена реална съвкупност - например, на пълнолетното население или някаква подсъвкупност от него?

В търсене на решение

- Каква процедура за рекрутиране на потенциални респонденти за интернет изследване може да се избере, така че тя да бъде съвместима с дефиницията за представителност на извадката?
- Да разгледаме една идея.

Процедурата е следната

Първоначално е необходимо да се проведе **мощно по обем интернет изследване с 15-20 хиляди респонденти** (по сегашната процедура - на отзоваване), обаче с много ограничена програма, включваща две части: едната част с въпроси за пола , възрастта, образованието и местоживееенето на респондентите, а другата част с от три до пет въпроса със съдържателен характер - например за COVID пандемията, за войната в Украйна или някакви други. Така ще се получи информация за една интернет съвкупност (извадкова по същество). За нея ще знаем какви са базисните характеристики (полова структура, възрастовия състав, образователна структура, състав по местоживееене).

Следващата стъпка:

- Структурата на извадковата интернет съвкупност по 4-те признака се сравнява с някаква реална генерална съвкупност, за която се разполага с данни от други източници, най-вече от националната статистика - преброявания, изчерпателни статистически наблюдения и други подобни
- Към тази съвкупност бихме могли впоследствие да генерализираме резултати от бъдещи интернет изследвания. Анализът от сравнението ще покаже доколко е изкривена, орязана извадковата интернет съвкупност по отношение на 4-те признака. Тук могат да се използват съответни статистически методи за проверка на хипотези за различие

Какво следва: две вероятни ситуации

Първата вероятност:

Ако различията са **статистически незначими**, това ще означава, че тези 15-20 хиляди интернет адреси или респонденти са случайна извадка от сравняваната реална съвкупност

От тази голяма майка-извадка по-нататък по случаен път могат да се избират по-малки извадки от интернет адреси (потенциални респонденти) за съответно емпирично изследване - например, извадки с обем 1000, 1500 или 2000 респонденти, с каквито се работи в обичайната практика. **Така респондентите няма да се набират на принципа на отзоваването", а това ще става по правилата на случайната извадка**

Втората вероятност:

- Ако различията са **статистически значими**, това ще означава, че съвкупността от 15-20 хиляди интернет адреса е повече или по-малко изкривена, орязана спрямо сравняваната реална съвкупност
- С други думи, търсената генерална съвкупност за интернет изследванията остава неясна и въпросът за представителността на интернет извадката е без решение
- В такива случаи единственият коректен изход е да се примирим с изкривяванията и при обявяването на резултатите от изследването те почтено да бъдат отбелязани
- **И най-важното - трябва да се откажем от генерализацията върху проверяваната генерална съвкупност или някаква друга генерална съвкупност**
- Или другояче казано, заявяваме, че **резултатите от изследването с такава извадка важат само за нея**. Така потребителите на резултатите от интернет изследването няма да бъдат подвеждани, респективно ще бъдат предпазени от изкривени представи за тези резултати

Претеглянето на извадковите данни - опит да се преодолее тяхното изкривяване в резултат на неслучайния подбор на респондентите

- Ще разгледаме един пример при класическите изследвания, след което ще направим сравнение с интернет изследванията, за да видим какви са възможностите за евентуално приложение (примерът е условен и се дава с демонстративна цел)
- Проведено е извадково изследване с 1400 лица (планирана извадка), но на терена са изследвани 1200 лица (200 лица са отпаднали поради различни причини - отказ, ненамиране на адреса, неотговорили на повечето въпроси и т.н.). Предполагаме (това е необходимо условие), че се разполага с данни за генералната съвкупност за признаците пол и възраст (от проведено преди време изчерпателно изследване). Самата изчислителна процедура изглежда така:

Таблица 1

Пол, възраст	Извадка (брой лица)	Относителен дял в извадката	Относителен дял в генералната съвкупност	Тегла / генерална съвкупност към извадката
Мъже до 30 год.	204	0,170	0,190	1,118
Мъже 31-50 год.	240	0,200	0,240	1,200
Мъже 51 и повече год.	120	0,100	0,150	1,500
Жени до 30 год.	156	0,130	0,160	1,231

Жени 31-50 год.	360	0,270	0,150	0,556
Жени 51 и повече год.	120	0,130	0,110	0,846
Всичко	1 200	1 000	1 000	

Какъв е ефектът от претеглянето на извадковите данни?

Сравняването на една двумерна таблица по два признака, различни по пол и възраст, преди и след претеглянето, ще покаже, че общият брой на изследваните лица остава същият (1 200 лица), **но броят на лицата в клетките на претеглената таблица ще се различава от непретеглената**, в смисъл че след претеглянето в новата таблица разпределението на лицата по двата признака ще отчита отстраненото изкривяване по признаците пол и възраст

Претеглянето на извадковите данни се прави при определени предположения. Те са:

- а) **Ако има висока корелация** между претеглящите признаци (в случая пол и възраст) и другите признаци, включени в програмата на изследването
- б) **Ако няма висока корелация, процедурата по претеглянето е излишна**
- в) **Ако са налице значителни отклонения** в разпределенията по претеглящите признаци между изследваните и неизследваните лица
- г) **Ако няма значителни отклонения** по претеглящите признаци, тогава процедурата по претеглянето е излишна

Сега да видим как стоят нещата при интернет изследванията

- Очевидно усилията трябва да бъдат насочени в друга посока. Например, да се направи предложение до Националния статистически институт при бъдещи големи статистически наблюдения (такива от рода на преброяванията на населението) да се включат специални въпроси, чрез които да се получи информация за подходящи генерални съвкупности, от които да се излъчват извадки за интернет изследвания
- Такова предложение до НСИ може да се инициира от институти на БАН , катедри на университети, Българската социологическа асоциация и други
- **Решаването на въпроса за генералната съвкупност при интернет изследванията ще изисква време и изследователски усилия, но при постигането на успех процедурите за формирането на извадките и тяхната представителност при тези изследвания ще се изравнят с процедурите при класическите емпирични изследвания**

Заклучителни думи

Интернет изследванията са в началото на своето разгръщане и не бива да се отказваме от търсенето на адекватни процедури за формиране на представителни извадки и рекрутиране на потенциални респонденти

Това ще зависи от изобретателността и методическите въображения на изследователите и методиците, занимаващи се в настоящето и в бъдеще с тези изследвания

- **Цитирана литература:**
- **Цонев, В. Ст.** (1958), Основи на репрезентативното изучаване, София : ДИ „Наука и Изкуство“.
- **Cochran, W. G.** (1963), Sampling Techniques, 2nd ed., N.Y.: John Wiley and Sons; Методы въборочного исследования (1976), Москва, Статистика.
- Deming, W. E. (1950), Some Theory of Sampling, N.Y.: John Wiley and Sons.
- **Hansen, M., Hurwitz, N., Madow, W. G.** (1953), Sample Survey Methods and Theory, N.Y.: John Wiley and Sons, Vols 1 & II.
- **Kish, L.** (1965), Survey Sampling, N.Y.: John Wiley and Sons.
- **Yates, F.** (1960), Sampling Methods for Censuses and Surveys, 3rd ed., London: Chas. Griffin and Company.
- **Zarkovich, S. S.** (1961): Sampling Methods and Censuses; vol. I, Collecting Data and Tabulation (Draft), Rome: FAO; vol. II, Quality of Statistical Data (Draft), Rome: FAO.