

ИЗВАДКИ

Изучаването на определена съвкупност от единици може да се осъществи по два начина:

- *изчерпателно* – когато в изследването се включат всички единици на изучаваната съвкупност, наричана генерална съвкупност.
- *извадково* – когато изследването се провежда с ограничен брой от единици на съвкупността, подборът на които се осъществява с различни видове извадки.

Предимства на извадковите изследвания пред изчерпателните:

- по-евтини
- по-бързи резултати
- възможности за по-добра организация, което намалява грешките от регистрацията и обработката (систематичната грешка)
- приложимост в области, където е невъзможно провеждането на изчерпателни изследвания (много големи генерални съвкупности)
- при представителни извадки изводите могат да се обобщават за цялата съвкупност

Извадките биват:

- *Представителни* (случайни, репрезентативни, стохастични) при които се наблюдават само единиците от извадката, а получените резултати се обобщават за цялата съвкупност.
- *Непредставителни* (неслучайни) при които изводите не могат да се генерализират за съвкупността.

I. ПРЕДСТАВИТЕЛНИ ИЗВАДКИ

1. Условия за представителност на извадката:

- единиците в извадката да бъдат избрани чрез случаен подбор, т.е. на всяка единица от генералната съвкупност да се осигури еднаква вероятност, еднакъв шанс, да попадне в извадката;
- извадката да съдържа достатъчен брой единици, което ще осигури достатъчно малка случайна грешка.

Ако извадката може да бъде представителна, но не достатъчно точна, то точността е немислима без представителността.

2. Етапи на формиране на случайната извадка:

- определяне на генералната (целевата) съвкупност, за което е необходимо множество от същностни, времеви и пространствени характеристики, които разграничават интересуващите ни единици от онези, които имат определена прилика с тях;
- определяне модела на извадката – съществува голямо разнообразие от модели случайни извадки. В някои случаи е възможно да се приложи само един от тях, в други имаме възможност за избор, който се определя от много практически съображения;
- съставяне основата на извадката – „списъка” от единици, от който със съответната процедура трябва да се излъчат единиците на извадката. Проблемите при списъците са: непълни, неточни, с повторения, с недостъпни части, с липсващи части, с чужди единици. Списъкът може да бъде на първични или съставни единици;

- определяне на интересуващите ни параметри – средни аритметични, стандартни отклонения, относителни дялове;
- определяне (изчисляване) обема на извадката, забравяйки че не относителният дял, а абсолютният брой на единиците в извадката е величината свързана с представителността.

3. Методи за случаен подбор

3.1. Лотарийен подбор

Подборната процедура има няколко варианта:

- чрез разиграване на лотария – номерата на единиците се поставят в урна и се разбъркват добре, изтеглят се толкова номера колкото е обема на извадката, като веднъж изтеглен даден номер не се връща в урната. Метода е подходящ при относително малки съвкупности, където не се създават технически трудности. Лотарийният подбор може да се реализира с SPSS с последователното изпълнение от главното меню на функциите: Data -> Select Cases -> Random sample of cases -> Sample Size, където имаме възможност да зададем обема на извадката като процент от цялата съвкупност или с абсолютен брой от всички единици в съвкупността.
- с *таблица на случайните числа*, които всъщност са числа подбрани по лотарийен път.

3.2. Систематичен случаен подбор включващ:

- определяне на подборната стъпка (крачка) като отношение на обема на генералната съвкупност (N) и обема на извадката (n) $\Rightarrow k = N/n$;
- определяне на стартовото число n_1 в рамките на стъпката ($n_1 < k$) по случаен начин. Това число е номера на първия попаднал в извадката;
- номерът на втория се определя като към номера на първия се прибави подборната стъпка, т.е. $n_2 = n_1 + k$, номерът на третия е $n_3 = n_2 + k$ и т.н.

Пример: Формиране на извадка с обем 300 от генералната съвкупност съдържаща 1500 единици.

- определя се подборната стъпка $k=1500/300=5$
- изтегля се лотарийно едно число от 1 до 5, например 2
- номерата на единиците от списъка попадащи в извадката са 2, 7, 12, 17, 22, 27 и т.н.

Систематичният подбор може да бъде по ефективен, когато единиците са подредени последователно по някакъв критерий.

Проблем при систематичния подбор може да има при цикличност при подреждането на единиците. Преодоляването му става чрез рандомизиране на списъка (случайно присвояване на номерата).

4. Модели представителни извадки

4.1. Проста извадка – формира се направо от списъка на първичните единици на съвкупността чрез някой от методите за случаен подбор.

Недостатъци:

- необходимо е осигуряване на списъка на генералната съвкупност – скъпо, изисква време, бързо остарява
- единиците в извадката са максимално разпръснати по цялата територия

Принципите на простата извадка са залегнали във всички други извадки.

4.2. Стратифицирана (районирана) извадка – прилагаме при нехомогенни съвкупности, т.е. съвкупността е вътрешно разнородна и може да бъде разгледана като съставена от няколко подсвкупности (страти).

Основанията за стратификация са:

- повишаване на статистическата ефективност – стратите са силно хомогенни, а помежду си хетерогенни, което намалява вътрешногнездовата дисперсия. Така при един и същи обем на извадката може да се получи по-голяма точност или същата точност се получава при по-малко разходи.
- осигуряване на данни за подсвкупностите поотделно

Методът се състои в:

- обособяване на стратите по определен критерий (признак). Като критерии за стратифициране се използват признаци, които силно корелират с изучаваните променливи и осигуряват максимална разлика между стратите и минимална разлика между единиците в стратите. Като критерии за райониране понякога се използва териториален признак .
- излъчване на единиците чрез случаен подбор в рамките на всяка страта.

Броят на стратите се определя от подсвкупностите за които ще се правят статистически изводи.

Големината на стратите зависи от обема на извадката и начина на разпределение между отделните страти.

Съществуват две *разновидности на стратифицирания модел:*

- *пропорционална извадка* при която извадката във всяка от стратите е пропорционална на относителното и тегло в съвкупността. Прилагаме когато не съществуват съществени различия вътре в стратите и извадката в отделните страти са еднакво скъпи;
- *непропорционална извадка* при която извадката не е пропорционална за всяка страта. Съществуват множество схеми на непропорционален подбор. Една от тях е пропорционален подбор на степента на разсейване и обема на района, т.е.

$$n_1 : n_2 : n_3 = N_1\sigma_1 : N_2\sigma_2 : N_3\sigma_3$$

Непропорционалният подбор налага претегляне на районните оценки.

4.3. Гнездова (кълъстерна) извадка

При гнездовата извадка съвкупността се разделя на много на брой подгрупи (гнезда), всяко от което с малък брой единици. Гнездовата единица се определя от критерий, позволяващ улесняване на работата по събиране на информацията (териториални, производствени и др.) и респективно намаляване на разходите. Освен това се цели осигуряване на хетерогенност вътре в гнездата и хомогенност между тях. Тъй като изискването за хетерогенност в гнездата е рядко изпълнимо, това намалява статистическата ефективност на гнездовата извадка. Компенсацията на този недостатък е добрата икономическа ефективност, което позволява увеличаване на обема на извадката.

При гнездовата извадка най-напред се избират случайно определен брой гнезда (съставни единици на съвкупността). След това наблюдението може да се извърши вътре в гнездото изцяло или на част от единиците, определени чрез съответната процедура на случаен подбор.

Следователно не е необходим списъка на генералната съвкупност, а само на гнездата попаднали в извадката.

В зависимост от това на колко стъпки се извършва подбора, гнездовите извадки биват едностепенни, двустепенни, многостепенни.

Едностепенна гнездова извадка – избират се случайно определен брой гнезда и вътре в избраните гнезда се наблюдават всички единици.

Ако гнездата имат различен размер, за да се получи пропорционален подбор от различни по-размер гнезда, последните се групират (районират) по броя на единиците в тях и от всеки район се подбират пропорционален брой гнезда.

Общият брой на гнездата в извадката не трябва да бъде по-малък от 50, което при голям обем на гнездата създава трудности. Поради това в практиката по-често се прилагат двустепенни и многостепенни извадки.

Двустепенна гнездова извадка – на първата стъпка се подбират случайно определен брой гнезда, а на втората в рамките на тези гнезда се подбират определен брой единици.

Първата стъпка има преимуществата на едностепенната гнездова извадка, а втората стъпка осигурява възможност да се компенсира нарушеното изискване за хетерогенност на единиците в гнездото.

Съотношението между извадката на първата и втората степен варира така, че едновременно да се осигури изискването за повече от 50 гнезда и да се намали влиянието на вътрешно гнездовата корелация.

Броят на изследваните единици в гнездото е пропорционален на размера на гнездото.

Частен случай на този извадков модел е **двустепенната гнездова извадка с вероятност пропорционална на размера на гнездото** при която във всяко гнездо се изследват еднакъв брой единици, независимо от размера на гнездото. Това нарушава еднакъв шанс за избор, защото в малките гнезда единиците се избират с по-голям процент, а в големите с по-малък процент. За да се компенсират нарушените шансове, подборът се реализира така, че големите гнезда да имат по-голям шанс да бъдат избрани в извадката от малките.

Подборът на гнездата се извършва по-специална техника, с помощта на кумулативна колона на обема на гнездата (сумата от обемите на всички предходни гнезда), чрез подборна крачка (стъпка) определена като съотношение на обема на генералната съвкупност (N) и броя на гнездата в извадката (m) $\Rightarrow k = N/m$.

Първото гнездо ще бъде това, чийто кумулативен сбор съдържа стартовото число определено случайно, обикновено по-малко от k . Второто гнездо е онова, чийто кумулативен сбор съдържа стартовото число плюс подборната крачка и т.н.

Пример: Обемът на генералната съвкупност е 75000, броят на гнездата в извадката е 50. Определя се $k = N/m = 75000/50 = 1500$ и стартово число 300 (избрано случайно между числата от 1 до 1500).

Построява се кумулативна колона за обемите на гнездата (дадени са първите 20 гнезда)

№ на гнездото	Обем на гнездото	Кумулат. колона		№ на гнездото	Обем на гнездото	Кумулат. колона	
1.	230	230		11.	288	4159	
2.	723	953	300	12.	620	4779	
3.	550	1503		13.	563	5342	4800
4.	250	1753		14.	158	5500	
5.	436	2189	1800	15.	463	5963	
6.	198	2387		16.	258	6221	
7.	131	2518		17.	356	6577	6300

8.	466	2984		18.	298	6875
9.	267	3251		19.	123	6998
10.	620	3871	3300	20.	423	7421

Така номерата на гнездата попаднали в извадката са 2, 5, 10, 13, 17 и т.н.

Във всяко от тези гнезда се изследват между 5 и 20 души, така че броя на гнездата по броя на изследваните във всяко гнездо да даде планирания обем на извадката.

II. НЕПРЕДСТАВИТЕЛНИ ИЗВАДКИ

1. Нецелевни неслучайни извадки:

1.1. *Извадка по „удобство“* – основен принцип е съобразяване с удобството. Прилага се при еднородни съвкупности, при предварителни изследвания имащи за цел изпробване на инструментариума на изследването

1.2. *Извадка според „отзовалите се“* – включва доброволно откликналите. Прилага се при аудиторни изследвания на ТВ, радио, вестници – класация на песни.

2. Целеви неслучайни извадки

2.1. *Типологична извадка* – включва единици, които отговарят на дадени интересувачи ни критерии. За дефиниране на критериите се използват различни методи вкл. експертни оценки.

2.2. *Квотни извадки* – единиците се подбират въз основа на характеристики обвързани с предмета на изследването.

Квотната извадка осигурява структурата на извадката по посочените квотни признаци идентична с тази на изучаваната съвкупност.

Колкото повече квотни признаци бъдат включени, толкова извадката би била по-мощна, но на практика това създава трудности при идентификацията на единиците за изследването.

За осъществяването на квотния подбор се формират квотни матрици, в които са отразени квотните признаци и техните категории. В клетките се фиксира абсолютния брой на лицата подлежащи на изследване

Проблеми при квотната извадка:

- преписват и се несъществуващи черти (поради приликата и със стратифицираната извадка) – незаконно се прилагат методи да определяне на обема на извадката и определяне на размера на грешката, приложими при случайните извадки
- липсва актуална информация за разпределение на съвкупността по квотообразуващите признаци

Квотната извадка се *валидизир*, чрез проверка за съответствие на структурата на извадката и генералната съвкупност по неквотиращи признаци и колкото повече са те, толкова по-добра е квотната извадка.

2.3. *Извадка „снежна топка“* – първата единица се подбира случайно, а следващите на информацията, придобита от първите. Прилага се при:

- липса на списъци на генералната съвкупност
- липса на информация относно структурата, за да може да се приложи квотна извадка
- съвкупността от интересувачите ни единици е твърде малка и със случайна извадка не биха се уловили критичното число случай, върху които да се прави обобщения.

Слабости:

- препоръчаните последващи единици са твърде подобни, което изкривява резултатите, поради слабата вариация

- не може да бъде валидирана, поради липса на информация относно структурата на генералната съвкупност.

ИЗВОДИ:

- **Винаги когато е възможно случайните извадки трябва да бъдат предпочитани пред неслучайните.**
- При невъзможност да бъде използван случаен модел за предпочитане е такъв неслучаен модел, който имитира случайните или се използват процедури типични за случайните.