**Глава 4**

**РЕПРЕЗЕНТАТИВНИ ПРОУЧВАНИЯ**

# *Г. Грънчарова*

# *В тази глава:*

*4.1. Същност на репрезентативните проучвания*

*4.2. Основни понятия и принципи при подбор на извадки*

*4.3. Видове извадки*

*4.4. Системни грешки при репрезентативните проучвания*

*4.5. Групови свойства на статистическите съвкупности*

*4.6. Въпроси за самоподготовка*

# 4.1. Съ­щ­ност на ре­п­ре­зен­та­ти­в­ни­те про­у­ч­ва­ния

Както бе подчертано в предходния раздел, двете основни понятия в статистиката са “***популация” и “извадка”.*** Когато правим заключение за характеристиките на дадена популация на базата на данни, получени при изучаване на определена част от популацията, ние всъщност използваме информация от извадки, за да направим индуктивни изводи за популацията. Такава информация има известни ограничения по отношение на надеждността, точността и валидността. Но независимо от това, проучванията, опиращи се на наблюдения на извадки, представляват най-честата практика в научните изследвания.

***Основните доводи за предпочитане на репрезентативните пред цялостните проучвания*** се свеждат най-често до следното:

* ограничените финансови и материални ресурси;
* ограничения във времето;
* липсата на достъп до цялата популация (тъй като понякога тя е безкрайна) и наблюдението на извадка може да е единственият възможен метод за събиране на информация.

Всичко това важи особено за медицината и здравеопазването, където поради масовия характер на явленията и ограничените ресурси много рядко могат да се обхванат всички единици на дадена популация. Затова медицинското познание, което имаме днес за човешките популации, се опира преди всичко на ***репрезентативни проучвания,*** на основата на които се правят изводи и заключения за съответните популации, от които са извлечени наблюдаваните извадки.

***Ре­п­ре­зен­та­ти­в­ни­те про­у­ч­ва­ния*** имат ре­ди­ца ва­ж­ни ***пре­дим­с­т­ва*** пред ця­ло­с­т­ни­те про­у­ч­ва­ния:

* ***намаляват потребностите от финансови, човешки и материални ресурси, т.е. те са по-икономични;***
* ***резултатите се извличат по-бързо***;
* ***постига се по-голяма точност на събраните данни***, тъй като по-малките извадки позволяват да се вложат повече усилия за намаляване на грешките, несвързани с подбора и намаляване на систематичните грешки поради неотзоваване на изследвани лица;
* ***по-то­ч­ни са от ця­ло­с­т­ни­те про­у­ч­ва­ния,*** кол­ко­то и учу­д­ва­що да е то­ва на пръв по­г­лед. При тях се до­пу­с­ка ***ре­п­ре­зен­та­ти­в­на гре­ш­ка*** (наричана още ***случайна или стохастична***), но тя мо­же да се оп­ре­де­ли сравнително точно (да се изчисли ма­те­ма­ти­че­с­ки) и да се пре­д­ви­ди в край­ни­те из­во­ди за по­пу­ла­ци­я­та. Тя зависи от броя на наблюдаваните случаи и начина на подбор на извадката и може да бъде сведена до минимум чрез ***наблюдение на достатъчен брой непреднамерено подбрани случаи***, докато при изчерпателните проучвания се допуска грешка на регистрацията, която може да нараства с броя на наблюдаваните случаи и трудно се контролира.

***Репрезентативните проучвания*** имат и някои ***недостатъци:***

* винаги се наблюдава определена ***репрезентативна грешка;***
* работата с извадки може да създаде ***чувство за дискриминация*** в рамките на популацията;
* работата с извадки не е за препоръчване, когато има правно изискване за всяка единица от популацията да се води съответен запис (регистрация);
* при някои ***редки явления*** малките извадки може да не предоставят достатъчно случаи за изучаване.

При репрезентативните проучвания с помощта на статистическите методи могат да се постигнат ***три главни цели:***

* да се на­п­ра­вят из­во­ди за ця­ла­та по­пу­ла­ция въз ос­но­ва на ана­ли­за на дан­ни от съ­о­т­ве­т­ни из­ва­д­ки;
* да се на­п­ра­ви оцен­ка на сте­пен­та на си­гур­ност, т. е. да се оп­ре­де­ли га­ран­ци­он­на­та ве­ро­я­т­ност на из­во­ди­те за по­пу­ла­ци­я­та;
* да се оп­ре­де­ли на­чи­нът на сфор­ми­ра­не на из­ва­д­ка­та и не­й­ния раз­мер, та­ка че да се оси­гу­рят то­ч­ни и на­де­ж­д­ни из­во­ди и за­к­лю­че­ния.

Изключително важно място в организацията на репрезентативните проучвания принадлежи на ***подбора на извадката***.

**4.2. Основни понятия и принципи при подбор на извадки**

Преди да разгледаме конкретните схеми за подбор на извадките, нека изясним няколко основни понятия:

* ***Извадка (Sample)*** – подмножество, подклас от дадена популация, чиито свойства са или предстои да бъдат генерализирани за популацията.
* ***Подбор на извадка (Sampling)*** – това е самата процедура, самият метод за избиране на индивиди от една или повече популации.
* ***Извадкова единица (Sampling unit)*** – това е най-малката единица, използвана в процеса на подбор (лице, домакинство, район).
* ***Единица на изследване или наблюдение (Unit of inquiry)*** – най-малката единица, върху която се събират данните (извършва се измерване или запис) при едно извадково проучване.
* ***Рамка на извадката (Sampling frame)*** – това е наборът от извадкови единици, от които се подбира дадена извадка (напр., списък с имена, места или други неща, които се използват като единици на извадката).
* ***Извадкова фракция (Sampling fraction)*** – това е пропорцията от извадкови единици, които трябва да се подберат от определена извадкова рамка за включване в извадката.

Съществуват два основни подхода за подбор:

* + - ***непреднамерен подбор***
    - ***преднамерен подбор***

Една ***добра извадка*** трябва да притежава следните ***характеристики***:

* + ***да бъде подбрана случайно*** с цел да се намали систематичната грешка;
  + ***да бъде репрезентативна*** по отношение на популацията с цел да се подобри нейната валидност;
  + ***да бъде достатъчно голяма*** с цел да се повиши нейната точност.

Явно е, че на тези условия отговаря единствено ***непреднамереният случаен по­д­бор*** (probality sampling scheme - схе­ма за ве­ро­я­т­но­с­тен по­д­бор), при който ***всеки индивид (елемент) от популацията има еднакъв шанс да бъде включен в извадката.***

Най-ва­ж­но­то изи­с­к­ва­не, на ко­е­то тря­б­ва да от­го­ва­ря вся­ка из­ва­д­ка, е тя да бъ­де ***ре­п­ре­зен­та­ти­в­на (пре­д­с­та­ви­тел­на*)** по отношение на популацията. Са­мо при на­ли­чи­е­то на та­ко­ва съ­о­т­ве­т­с­т­вие ме­ж­ду дан­ни­те от из­ва­д­ка­та и ця­ла­та по­пу­ла­ция е въз­мо­ж­но обо­б­ща­ва­не на дан­ни­те за по­пу­ла­ци­я­та.

Мяр­ка за ре­п­ре­зен­та­ти­в­но­ст­та на вся­ка из­ва­д­ка е раз­ли­ка­та ме­ж­ду сре­д­ни­те и от­но­си­тел­ни ве­ли­чи­ни в из­ва­д­ка­та и ця­ла­та по­пу­ла­ция, ко­я­то се из­мер­ва с ***ре­п­ре­зен­та­ти­в­на­та гре­ш­ка.*** Най-голямото предимство на извадка, подбрана на основата на непреднамерен подбор в сравнение с преднамерен подбор е в това, че може да се оцени размера на репрезентативната грешка и да се вземе предвид при обобщаване на данните. В за­ви­си­мост от ха­ра­к­те­ра на изу­ча­ва­ни­те при­з­на­ци (ка­че­с­т­ве­ни или ко­ли­че­с­т­ве­ни), ве­ли­чи­на­та на ре­п­ре­зен­та­ти­в­на­та гре­ш­ка се оп­ре­де­ля по раз­ли­чен на­чин.

***Репрезентативността зависи от:***

* ***числеността на изучаваната извадка;***
* ***от начина на сформиране на извадката*.**

Съ­г­ла­с­но ***за­ко­на за го­ле­ми­те чи­с­ла,*** с уве­ли­ча­ва­не на броя на на­б­лю­да­ва­ни­те слу­чаи на­ма­ля­ва вли­я­ни­е­то на слу­чай­но­ст­та и все по­ве­че се про­я­вя­ва за­ко­но­мер­но­стта. С дру­ги ду­ми, с уве­ли­ча­ва­не на броя на на­б­лю­да­ва­ни­те слу­чаи на­ма­ля­ва ве­ли­чи­на­та на ре­п­ре­зен­та­ти­в­на­та гре­ш­ка и ре­зул­та­ти­те от на­б­лю­де­ни­е­то на из­ва­д­ка­та все по­ве­че се при­б­ли­жа­ват до те­зи за ця­ла­та по­пу­ла­ция. Има спе­ци­ал­ни статистически ме­то­ди, ко­и­то по­з­во­ля­ват да се оп­ре­де­ли ***не­об­хо­ди­ми­ят и до­с­та­тъ­чен брой слу­чаи за на­б­лю­де­ние***, та­ка че ве­ли­чи­на­та на ре­п­ре­зен­та­ти­в­на­та гре­ш­ка да бъ­де сведена до ми­ни­мум и до­с­то­вер­но­ст­та на из­во­ди­те за по­пу­ла­ци­я­та да бъ­де до­с­та­тъ­ч­но ви­со­ка.

**4.3. Видове извадки**

***Про­с­та слу­чай­на из­ва­д­ка (simple random sample)***

То­ва е най-про­с­ти­ят слу­чай на подбор на извадки, по­д­хо­дящ за ог­ра­ни­че­ни по раз­мер про­у­ч­ва­ния. Характерните особености на този подход са:

* всяка единица в извадковата рамка има равен шанс да бъде избрана;
* случайният подбор от извадковата рамка може да бъде направен чрез теглене на жребий, използване на таблица за случайните числа или чрез компютърни програми, генериращи случайни числа;
* необходим е списък на единиците в извадковата рамка.

Напр., за да по­д­бе­рем про­с­та из­ва­д­ка от 50 слу­чая от ма­сив от 800 ин­ва­ли­ди­зи­ра­ни ли­ца в гр. А (оп­ре­де­лен чрез се­рий­ни но­ме­ра от 1 до 800 в до­ку­мен­та­ци­я­та на Териториалните ек­с­пер­т­ни­ ле­кар­с­ки ко­ми­сии), би­х­ме мо­г­ли да про­с­ле­дим по­с­ле­до­ва­тел­но три­ци­ф­ре­ни чи­с­ла от та­б­ли­ца­та за слу­чай­ни­те чи­с­ла. Та­ка на­при­мер, ако та­б­ли­ч­ни­те оз­на­че­ния са 12454, 45730, 07944, 73506, 81149,..........., то то­га­ва из­би­ра­ме чи­с­ла­та 124, 544, 573, 007, 944, 735, 068 и т. н. до­ка­то на­бе­рем 50 слу­чая. При то­ва чи­с­ла­та над 800 се пре­не­б­ре­г­ват (тъй ка­то ця­ла­та по­пу­ла­ция е ма­к­си­мум от 800 слу­чая и ня­ма слу­чай с по­ре­ден но­мер 944 напр.), а съ­що та­ка тря­б­ва да се вни­ма­ва с ну­ли­те. Ако те не се взе­мат пре­д­вид, има опа­с­ност да не вклю­чим слу­чаи с по­ре­д­ни но­ме­ра под 100, ко­е­то ще бъ­де съ­в­сем по­г­ре­ш­но.

|  |  |
| --- | --- |
| **Предимства** | Недостатъци |
| * Тъй като всяка единица в популацията има равен шанс за включване в извадката, репрезентативността е гарантирана и извадката е изложена само на репрезентативна грешка. * Оценъчните характеристики на извадката се изчисляват лесно. | * Ако извадковата рамка е голяма, този метод може да се окаже непрактичен, поради трудността и разходите по съставянето и актуализирането на списъка на извадковите единици. |

***Си­с­те­ма­ти­ч­на из­ва­д­ка (systematic sample)***

На­ри­ча се още ***ме­ха­ни­ч­на (пропорционална)*** из­ва­д­ка. Характеризира се със следните особености:

* включва избор на всяка n-та единица в популацията или извадковата рамка, където 1/n представлява извадковата фракция (пропорцията от извадкови единици, които трябва да се подберат);
* първата единица (стартовото число) се избира случайно сред първите к единици в зависимост от извадковата фракция (напр. сред първите 10, ако извадковата фракция е 10%);
* след определяне на стартовото число започва подбор със стъпка равна на извадковата фракция;
* необходим е списък на единиците в извадковата рамка.

Например, за да по­д­бе­рем напр. 10% из­ва­д­ка от по­пу­ла­ция с чи­с­ле­ност N, пър­во из­би­ра­ме слу­чай­но на­чал­на­та то­ч­ка ме­ж­ду чи­с­ла­та 1 и 10 и по-на­та­тък в из­ва­д­ка­та вклю­ч­ва­ме все­ки 10-ти слу­чай от по­пу­ла­ци­я­та. Ако сме из­б­ра­ли слу­чай­но за старт чи­с­ло­то 6, по-на­та­тък вклю­ч­ва­ме в из­ва­д­ка­та 6-я, 16-я, 26-я, 36-я и т. н.

То­зи по­д­ход е осо­бе­но по­ле­зен, ко­га­то слу­ча­и­те се по­д­ре­ж­дат ав­то­ма­ти­ч­но с те­че­ние на вре­ме­то, ка­к­то е при по­с­тъ­п­ва­не и из­пи­с­ва­не на хо­с­пи­та­ли­зи­ра­ни­те бол­ни.

Се­ри­о­зен про­б­лем при си­с­те­ма­ти­ч­ния по­д­бор е въз­мо­ж­но­ст­та за до­пу­с­ка­не на си­с­те­ма­ти­ч­на гре­ш­ка (bias), ако има ня­ка­к­ва оп­ре­де­ле­на тен­ден­ция в по­пу­ла­ци­я­та, от ко­я­то се сфор­ми­ра из­ва­д­ка­та. На­при­мер, при изу­ча­ва­не на ха­ра­к­те­ри­с­ти­ки­те на па­ци­ен­ти­, при­е­ти в от­де­ле­ни­я­та за спешна по­мощ, не­ра­зум­но е да по­д­би­ра­ме са­мо при­е­ти­те в съ­бо­та през но­щ­та, тъй ка­то па­ци­ен­ти­те, по­с­тъ­пи­ли в на­ча­ло­то и в сре­да­та на се­д­ми­ца­та мо­гат да се раз­ли­ча­ват до­с­та съ­ще­с­т­ве­но по ре­ди­ца ос­но­в­ни ха­ра­к­те­ри­с­ти­ки.

|  |  |
| --- | --- |
| **Предимства** | Недостатъци |
| * Извадката се подбира лесно. * Лесно може да се определи подходяща извадкова рамка. * Извадката е равномерно разпределена върху цялата референтна популация. | * Извадката може да бъде повлияна от систематична грешка, ако някаква скрита периодичност съвпадне със стартовото число и извадковата фракция. * Трудно е да се определи точността на оценката от едно проучване. |

***Стра­ти­фи­ци­ра­на из­ва­д­ка (stratified sample)***

Нарича се още ***послойна извадка*** (от strata – слой). Характеризира се със следните особености:

* популацията първо се разделя на групи или слоеве според характеристиките, от които се интересуваме (напр. пол, възраст и др.);
* след това се подбира проста случайна извадка от всеки слой, използвайки една и съща извадкова фракция, освен ако не се препоръчва нещо друго по специални причини.

На­при­мер, дан­ни­те от пре­б­ро­я­ва­не­то на на­се­ле­ни­е­то по­ка­з­ват, че в да­ден ре­ги­он въз­ра­с­то­во­то раз­пре­де­ле­ние е сле­д­но­то: 0 -14 г. - 15%; 15-34 г.- 20%; 35-49 г. - 27%; 50-64 г. - 22%; 65 г. и повече - 16%.

Ако се нуждаем от из­ва­д­ка с чи­с­ле­ност 200 ду­ши, то то­га­ва ум­но­жа­ва­ме про­цен­ти­те за все­ки слой по 200 и ги раз­де­ля­ме на 100 като по та­къв на­чин оп­ре­де­ля­ме по кол­ко слу­чая тря­б­ва да се вклю­ч­ат в из­ва­д­ка­та от все­ки слой (стра­та) на по­пу­ла­ци­я­та. В на­шия при­мер то­ва оз­на­ча­ва, че от раз­ли­ч­ни­те въз­ра­с­то­ви ка­те­го­рии тря­б­ва да вклю­чим съ­о­т­ве­т­но 30, 40, 54, 44 и 32 слу­чая. От­дел­ни­те слу­чаи във все­ки слой се по­д­би­рат по ме­то­да на слу­чай­ни­те чи­с­ла или си­с­те­ма­ти­ч­но (напр. вся­ко 30-то ли­це на въз­раст до 14 г.).

|  |  |
| --- | --- |
| **Предимства** | Недостатъци |
| * Всяка единица в даден слой има равен шанс да бъде избрана. * Използването на една и съща извадкова фракция за всички слоеве гарантира пропорционална представителност в извадката на характеристиките, според които популацията е стратифицирана. | * Извадковата рамка от цялата популация трябва да бъде подготвена отделно за всеки слой. |

Гне­з­до­ва (кластерна) из­ва­д­ка

Характеризира се със следните особености:

* + - популацията първо се разделя на кластери (гнезда) от хомогенни единици, обикновено опиращи се на географска близост;
    - подбира се извадка от всички гнезда;
    - вътре в избраните гнезда се изследват или изучават всички единици.

|  |  |
| --- | --- |
| **Предимства** | Недостатъци |
| * Намалява разходите за изготвяне на извадкова рамка * Намалява разходите за пътуване между избраните единици. | * Репрезентативната грешка е обикновено по-висока, отколкото при проста случайна извадка от същия размер. |

***Ком­би­ни­ра­на или мно­го­с­те­пен­на из­ва­д­ка (multistage sample)***

То­зи по­д­ход се прилага, ко­га­то же­ла­ем да сфор­ми­ра­ме из­ва­д­ка за ши­ро­ко­ма­ща­б­но про­у­ч­ва­не с го­лям ге­о­г­ра­ф­с­ки об­х­ват. Ка­к­то оз­на­ча­ва име­то й, подборът се осъществява на етапи докато не се достигне до крайните единици на извадката:

* + - в първия етап се изготвя списък на големи извадкови единици (градове, села, училища и др.);
    - от този списък се подбира случайна извадка с вероятност на подбора пропорционална на размера;
    - за всяка от избраните в първия етап единици се съставя списък на по-малки извадкови единици (напр., ако в първия етап единиците са били градове, тогава във втория етап единиците могат да бъдат къщи или домакинства);
    - след това се подбира случайна извадка от единиците на втория етап, които се изучават;
    - процедурата може да съдържа три или повече етапа.

|  |  |
| --- | --- |
| **Предимства** | Недостатъци |
| * Намалява разходите за изготвяне на извадкова рамка | * Репрезентативната грешка нараства в сравнение с проста случайна извадка от същия размер. |

Съ­ще­с­т­ву­ват и дру­ги ме­то­ди, но те са по-мал­ко на­де­ж­д­ни. Например:

*•* ***Из­ва­д­ки по удо­б­с­т­во*** (convenience samples) – включват избиране на всички лица, които желаят да участват в дадено проучване, при условие, че тези лица отговарят на установени критерии. С други думи, извадките вклю­ч­ват ин­ди­ви­ди, ко­и­то най-ле­с­но мо­гат да бъ­дат раз­пи­та­ни или из­с­ле­д­ва­ни, ко­е­то е пре­дим­с­т­во от една страна, но те не са до­с­та­тъ­ч­но представителни за ця­ла­та по­пу­ла­ция.

*•* ***Са­мо­фор­ми­ра­ли се из­ва­д­ки*** (self-selected samples) – например, при по­щен­с­ки­те ан­ке­ти лицата, ко­и­то от­го­ва­рят на из­п­ра­те­ния по по­ща­та или пу­б­ли­ку­ван във ве­с­т­ни­ци или спи­са­ния въ­п­ро­с­ник, ве­ро­я­т­но се раз­ли­ча­ват от не­от­го­во­ри­ли­те ли­ца. Оби­к­но­ве­но те­зи, ко­и­то са по-удо­в­ле­т­во­ре­ни или по-не­у­до­в­ле­т­во­ре­ни, от­го­ва­рят по-че­с­то от ос­та­на­ли­те. Не­от­го­во­ри­ли­те, от дру­га стра­на, мо­гат да ока­жат съ­ще­с­т­ве­но вли­я­ние вър­ху до­с­то­вер­но­ст­та на ре­зул­та­ти­те, ко­е­то тру­д­но мо­же да се пре­це­ни.

Въ­п­ре­ки то­ва, по­со­че­ни­те два по­д­хо­да по­ня­ко­га са не­об­хо­ди­ми - на­при­мер, при про­ве­ж­да­не на пи­ло­т­ни про­у­ч­ва­ния за те­с­ту­ва­не на въ­п­ро­с­ни­ци­те мо­же да се из­по­л­з­ва из­ва­д­ка по удо­б­с­т­во.

# 4.4. Систематични грешки при репрезентативните проучвания

***Си­с­те­ма­ти­ч­на гре­ш­ка (bias)******е вся­ка тен­ден­ция на да­де­на из­ва­д­ка да се от­к­ло­ня­ва из­вън слу­чай­но­то ва­ри­ра­не от съответната по­пу­ла­ци­я. Тя се при­чи­ня­ва от стра­ни­ч­ни, объркващи фактори (confounding effects).***

Цел­та на сфор­ми­ра­не на пре­д­с­та­ви­тел­на из­ва­д­ка е имен­но на­ма­ля­ва­не на си­с­те­ма­ти­ч­на­та гре­ш­ка или в иде­ал­ния слу­чай - све­ж­да­не на не­й­но­то вли­я­ние до не­з­на­чи­тел­но.

Ако е на­ли­це се­ри­о­з­на си­с­те­ма­ти­ч­на гре­ш­ка, то уве­ли­ча­ва­не­то на обе­ма на из­ва­д­ка­та може да за­дъл­бо­чи про­б­ле­ма.

Сле­до­ва­тел­но, още на ета­па на пла­ни­ра­не на про­у­ч­ва­не­то тря­б­ва да бъ­дем мно­го вни­ма­тел­ни по от­но­ше­ние на въз­мо­ж­ни­те из­то­ч­ни­ци на си­с­те­ма­ти­ч­ни гре­ш­ки и да се опи­т­ва­ме да ги от­с­т­ра­ним, ко­га­то то­ва е въз­мо­ж­но.

Някои източници на такива грешки са посочени по-горе при разглеждане на проблема за подбора на извадката:

•***си­с­те­ма­ти­ч­ни гре­ш­ки при са­мо­фор­ми­ра­ли се из­ва­д­ки*** (при по­щен­с­ки ан­ке­ти);

•***си­с­те­ма­ти­ч­ни гре­ш­ки по­ра­ди не­от­зо­ва­ва­не*** или не­об­х­ва­ща­не на вси­ч­ки пре­д­ви­де­ни ли­ца в из­ва­д­ка­та;

•***не­б­ре­ж­ност при прилагане на ме­то­ди­те на си­с­те­ма­ти­чен по­д­бор****;*

•***вклю­ч­ва­не на ре­зул­та­ти от пи­ло­т­но про­у­ч­ва­не вър­ху из­ва­д­ка по удо­б­с­т­во при анализа на данните от основното проучване****.*

Сле­до­ва­тел­но, ***систематичните гре­ш­ки, свър­за­ни с по­д­бо­ра на из­ва­д­ка­та***, са най-се­ри­о­з­ния про­б­лем.

Други видове систематични грешки са свързани с:

•***гре­ш­ки на при­по­м­ня­не­то*** *(recall bias)* – те са свързани със селективната способност на паметта и са характерни за ретроспективните проучвания, когато се събират данни за ми­на­ли събития, *история на заболяването и др.*

*•* ***си­с­те­ма­ти­ч­ни гре­ш­ки, свър­за­ни с из­с­ле­до­ва­те­ля*** (*interviewer bias*) - ен­ту­си­а­зъм, свен­ли­вост, страх, ча­рът на про­ти­во­по­ло­ж­ния пол и без­к­рай­но мно­го дру­ги фа­к­то­ри се ком­би­ни­рат, по­ня­ко­га по­д­съ­з­на­тел­но, в съ­з­на­ни­е­то на ин­тер­вю­и­ра­щия и мо­гат да до­ве­дат до пре­д­на­ме­ре­ност в по­д­бо­ра на из­с­ле­д­ва­ни­те ли­ца.

•***си­с­те­ма­ти­ч­ни гре­ш­ки, свър­за­ни с оце­ня­ва­щия*** *(assessor bias)* - оцен­ка­та на из­с­ле­до­ва­те­ля мо­же да бъ­де по­в­ли­я­на от предварителната информация, с която той разполага за из­с­ле­д­ва­но­то ли­це (или ле­че­ние ).

•***си­с­те­ма­ти­ч­ни гре­ш­ки по­ра­ди ли­п­с­ва­щи дан­ни*** *(missing data bias*) – това се наблюдава често, когато се проучват документи за минал период, съставени по друг повод, а не за целите на дадено проучване.

# • *си­с­те­ма­ти­ч­ни гре­ш­ки, свързани с въ­п­ро­с­ни­ка* (*questionnaire bias*) - ка­к­ви въ­п­ро­си се за­да­ват, как са фор­му­ли­ра­ни, ка­къв е ре­дът на за­да­ва­не­то им и т. н. - вси­ч­ко то­ва вли­яе вър­ху от­го­во­ри­те.

**4.5. Групови свойства на статистическите съвкупности**

***Ста­ти­с­ти­че­с­ката съ­в­ку­п­ност*** представлява ***мно­же­с­т­во от­но­си­тел­но ед­но­ро­д­ни еле­мен­ти, взе­ти за­е­д­но в из­ве­с­т­ни гра­ни­ци на вре­ме­то и про­с­т­ран­с­т­во­то*** *-* на­пр., на­се­ле­ни­е­то на да­ден ра­йон, гру­па­та на жи­во­ро­де­ни­те или мър­т­во­ро­де­ни­те през да­де­на го­ди­на в да­де­на те­ри­то­рия и т.н.

Пър­ви­ч­ни­те еле­мен­ти, от ко­и­то се из­г­ра­ж­да вся­ка ед­на ста­ти­с­ти­че­с­ка съ­в­ку­п­ност, се на­ри­чат ***еди­ни­ци (случаи) на на­б­лю­де­ние****.* Всеки отделен случай на наблюдение при­те­жа­ва ***при­з­на­ци на схо­д­с­т­во,*** ко­ито по­з­во­ля­ват включването му в дадена статистическа съ­в­ку­п­ност, но ед­но­в­ре­мен­но с то­ва има и ***при­з­на­ци на раз­ли­чие***, ко­и­то го пра­вят не­по­в­то­ри­м и ин­ди­ви­ду­але­н.

Обе­ди­ня­ва­не­то на мно­же­с­т­во единици на наблюдение, с при­съ­щи­те им еле­мен­ти на схо­д­с­т­во и раз­ли­чие, в ед­на ста­ти­с­ти­че­с­ка съ­в­ку­п­ност до­ве­ж­да до фор­ми­ра­не на нови характеристики, които ли­п­с­ват при от­дел­ни­те случаи и се на­ри­чат ***гру­по­ви свой­с­т­ва на ста­ти­с­ти­че­с­ка­та съ­в­ку­п­ност.***

Прилагането на разнообразни ста­ти­с­ти­че­с­ки ме­то­ди и по­ка­за­те­ли (кри­те­рии) е насочено именно към измерване и оценка на груповите свойства на изучаваните статистически съвкупности.

Основните групови свойства на всяка статистическа съвкупност са:

* ***разпределение на изучаваните променливи величини;***
* ***средно ниво (централна тенденция);***
* ***разнообразие (вариране);***
* ***репрезентативност;***
* ***взаимовръзка между променливите.***

# *Разпределение на променливите величини*

То­ва е ед­но от най-ва­ж­ни­те свой­с­т­ва на ста­ти­с­ти­че­с­ка­та съ­в­ку­п­ност. От­дел­ни­те еле­мен­ти на съ­в­ку­п­но­ст­та се раз­пре­де­лят не­е­д­на­к­во спо­ред стойностите на изу­ча­ва­ните променливи величини и по та­къв на­чин се образува оп­ре­де­ле­на въ­т­ре­ш­на стру­к­ту­ра на съ­в­ку­п­но­ст­та по отношение на всяка променлива. На­при­мер, при наблюдение на репрезентативна извадка от 100 жи­во­ро­де­ни мо­ми­че­та са регистрирани стойности на ръ­с­та, които варират в границите от 46 см до 54 см, но отделните числени значения на променливата “ръст” се проявяват с различна честота в извадката. Както се вижда от ***табл. 4.1****,* тридесет от наблюдаваните 100 живородени мо­ми­че­та имат ръст 50 см, по двадесет са съответно с ръст 49 и 51 см, а останалите 30 случая се разполагат почти симетрично около тези стойности и силно намаляват към крайните стойности.

Видът на раз­пре­де­ле­нието характеризира въ­т­ре­ш­ната стру­к­ту­ра на съ­в­ку­п­но­ст­та по отношение на дадена променлива и очертава в най-общ план зако­но­мер­но­с­тите на изу­ча­ва­но­то яв­ле­ние. Раз­пре­де­ле­ни­е­то мо­же да се пре­д­с­та­ви също така гра­фи­чно и то­ва спо­ма­га за още по-бър­зо ори­ен­ти­ра­не във фор­ма­та на раз­пре­де­ле­ние.

## *Табл.4.1. Стойности на ръста при 100 новородени момичета*

|  |  |
| --- | --- |
| **Ръст в см – х** | **Честота f** |
| 46 | 2 |
| 47 | 6 |
| 48 | 7 |
| 49 | 20 |
| 50 | 30 |
| 51 | 20 |
| 52 | 8 |
| 53 | 5 |
| 54 | 2 |
|  | **Σf = N = 100** |

Различаваме следните по-често срещани форми на разпределение:

• ***ал­тер­на­ти­в­но раз­пре­де­ле­ние*** - то се на­б­лю­да­ва при ка­че­с­т­ве­ни при­з­на­ци с две раз­но­ви­д­но­с­ти;

• ***нор­мал­но или си­ме­т­ри­ч­но раз­пре­де­ле­ние*** (на­ри­ча­но още Га­ус-Ла­п­ла­со­во) - то­ва е най-че­с­та­та фор­ма на раз­пре­де­ле­ние на ко­ли­че­с­т­ве­ни не­пре­къ­с­на­ти про­мен­ли­ви. При не­го слу­ча­и­те с раз­ли­ч­на стой­ност на при­з­на­ка се раз­по­ла­гат си­ме­т­ри­ч­но по от­но­ше­ние на сре­да­та на ре­да, при то­ва най-го­лям брой слу­чаи се стру­п­ват око­ло сре­д­но­то ни­во *(****фиг.4.1****).*

• ***аси­ме­т­ри­ч­но раз­пре­де­ле­ние*** - при не­го най-го­лям брой слу­чаи се стру­п­ват не в сре­да­та на ре­да, а се из­ме­с­т­ват към по-мал­ки­те или към по-го­ле­ми­те зна­че­ния на да­ден изу­ча­ван при­з­нак. Съ­о­т­ве­т­но го­во­рим за дя­с­но­изтеглено или лявоизтеглено разпределение. Такива фор­ми на раз­пре­де­ле­ние са по-чести при ди­с­к­ре­т­ни­те про­мен­ли­ви. Например, раз­пре­де­ле­ни­е­то на се­мей­с­т­вата у нас по брой де­ца е аси­ме­т­ри­ч­но - най-го­лям брой се­мей­с­т­ва имат 1-2 де­ца, а броят на се­мей­с­т­ва­та с по­ве­че от 2 де­ца рязко намалява.

Пример за несиметрично лявоизтеглено (отрицателно) разпределение са резултатите от тест при 50 студенти (***фиг.4.2***).

*Фиг. 4.1. Крива на нормално разпределение*

*Фиг.4.2. Крива на несиметрично отрицателно изтеглено разпределение*

В дру­ги си­ту­а­ции мо­же да се на­б­лю­да­ва стру­п­ва­не на повече слу­чаи в двата края или в други две точки на измерителната скала. Такива разпределения са ***дву­върхови или би­мо­дал­ни***. Възможни са и ***полимодални разпределения*** – с повече от 2 върха. Подобни случаи са се­ри­о­зен си­г­нал за не­е­д­но­ро­д­ност на изу­ча­ва­ната съ­в­ку­п­ност, което може да доведе до неверни изводи.

Формата на разпределението е свързана с по­д­бора на най-по­д­хо­дя­щите ста­ти­с­ти­че­с­ки ме­то­ди за об­ра­бо­т­ка на дан­ни­те. На­при­мер, ***па­ра­ме­т­ри­ч­ни­те ме­то­ди се при­ла­гат са­мо при нор­мал­но раз­пре­де­ле­ние, а не­па­ра­ме­т­ри­ч­ни­те методи се използват******при всякакви форми на разпределение.***

# *Средно ниво (централна тенденция)*

При бол­шин­с­т­во­то ко­ли­че­с­т­ве­ни променливи величини, не­за­ви­си­мо от ва­ри­ра­не­то на стойностите им при от­дел­ни­те на­б­лю­да­ва­ни слу­чаи, е на­ли­це тен­ден­ция към фор­ми­ра­не на оп­ре­де­ле­но сре­д­но ни­во, т. е. на­ли­це е оп­ре­де­ле­на ***цен­т­рал­на тен­ден­ция***.

Сре­д­но­то ни­во ха­ра­к­те­ри­зи­ра ти­пи­ч­но­то проявление на ко­ли­че­с­т­ве­ни­те променливи, ко­е­то се фор­ми­ра под вли­я­ние на ***оп­ре­де­ля­щи, за­ко­но­мер­ни фактори и при­чи­ни.*** Това сатакивафактори и причини, които се проявяват при всички единици на наблюдение, а не са присъщи само на отделни случаи.На­при­мер, сре­д­но­то ни­во на показателите за физическо развитие (тегло, ръ­с­т и др.) се фор­ми­ра под вли­я­ние на та­ки­ва за­ко­но­мер­ни фа­к­то­ри ка­то пол, въз­раст и др. Затова средните нива на тези показатели се различават съществено при лица на различна възраст или от различен пол.

За из­мер­ва­не и ха­ра­к­те­ри­с­ти­ка на сре­д­но­то ни­во се из­по­л­з­ват раз­ли­ч­ни ви­до­ве ***сре­д­ни ве­ли­чи­ни (сре­д­на ари­т­ме­ти­ч­на, мо­да, ме­ди­а­на).***

# *Разнообразие (вариране, разсейване)*

Въ­п­ре­ки стре­ме­жа на все­ки из­с­ле­до­ва­тел да по­с­ти­г­не ­е­д­но­ро­д­ност при формирането на изу­ча­ва­на­та съ­в­ку­п­ност, стойностите на променливите вели­чи­ни при отделните случаи се различават в по-голяма или по-малка степен. То­ва се дъл­жи на стремежа на ко­ли­че­с­т­ве­ни­те променливи към раз­но­о­б­ра­зие (ва­ри­ра­не), ко­е­то се про­я­вя­ва при всички живи организми и е ре­зул­тат от вли­я­ни­е­то на ***слу­чай­ни, не­о­п­ре­де­ля­щи фа­к­то­ри и при­чи­ни.*** Различните стойности на ръста и теглото на децата, например, се свързват с влиянието на такива фактори като наследственост, недохранване, хронични заболявания и др. Тези фактори не са еднакви при всички деца - при някои те са налице и предизвикват отклонения от типичното ниво на ръста и теглото за съответния пол и възраст, докато при други деца тези фактори отсъстват и стойностите на променливите са много по-близки до средното ниво.

За из­мер­ва­не на варирането се из­по­л­з­ват ***ли­мит на ва­ри­а­ци­он­ния ред, интерквартилен обхват, ди­с­пер­сия, стан­дар­т­но от­к­ло­не­ние, сре­д­но от­к­ло­не­ние, ко­е­фи­ци­ент на ва­ри­а­ция и др.***

# *Репрезентативност*

То­ва е ед­но от най-ва­ж­ни­те групови свой­с­т­ва и оз­на­ча­ва ***спо­со­б­но­ст­та на из­ва­д­ка­та да от­ра­зя­ва свой­с­т­ва­та на ге­не­рал­на­та съ­в­ку­п­ност (по­пу­ла­ци­я­та).*** За обе­з­пе­ча­ва­не на ре­п­ре­зен­та­ти­в­но­ст­та е не­об­хо­ди­мо да се спа­з­ват стро­го изи­с­к­ва­ни­я­та за не­пре­д­на­ме­рен по­д­бор на случаите в из­ва­д­ка­та и обезпечаване на до­с­та­тъ­чен обем на извадката. Ре­п­ре­зен­та­ти­в­ността се измерва т. нар. ***ре­п­ре­зен­та­ти­в­на (стандартна) гре­ш­ка,*** която по­ка­з­ва до­кол­ко ре­зул­та­ти­те от наблюдение на извадката се от­ли­ча­ват от те­зи, ко­и­то би­х­ме по­лу­чи­ли при на­б­лю­де­ние на популацията, от която е извлечена съответната извадка. Определянето на стандартната грешка е задължителен елемент при формулирането на заключения за параметрите на популацията на базата на данните от наблюдения на извадки.

# *Взаимовръзка между променливите велични*

Сте­пен­та на за­ви­си­мост ме­ж­ду променливите, ха­ра­к­те­ри­зи­ра­щи изучаваните яв­ле­ния, се из­мер­ва чрез раз­ли­ч­ни по­ка­за­те­ли, от ко­и­то най-ши­ро­ко се из­по­л­з­ват *ко­е­фи­ци­ен­ти­те на ко­ре­ла­ция.*

## 4.6. Въпроси за самоподготовка

1. Проста случайна извадка е тази, при която:

А. определяш размера на извадката и правиш избор пропорционален на популацията

Б. подбираш всеки случай без да отчиташ предишните избори

В. всеки случай в популацията има еднакъв шанс да бъде избран в извадката

2. Кое е необходимо условие, за да бъде една извадка стратифицирана:

А. всяко лице в популацията има една и съща вероятност да бъде включено в извадката

Б. изборът на метод за подбор на индивидите от популацията се определя изцяло от шанса

В. пропорциите на различните подгрупи в извадката да съответстват на тези в популацията

3. Проста случайна извадка се определя от:

### А. метода на подбор

Б. резултата от подбора

В. нейната степен на сходство с популацията

4. Кое от следните твърдения **не е вярно** за проста случайна извадка?

А. всеки елемент от популацията трябва да има равен шанс за включване в извадката

### Б. извадката трябва да бъде доста голяма, за да се счита случайна

В. нито едно от двете

5. Изследовател иска да подбере случайна извадка от 50 студенти. Той решава да подбере първите 50 студенти, влезли в студентския стол за вечеря. Тази извадка случайна ли е?

А. да Б. не

6. Изследовател иска да подбере случайна извадка от 50 студенти. Той решава да подбере първите 50 студенти, влезли в студентския стол за вечеря. Какъв вид е тази извадка?

А. проста случайна

Б. стратифицирана

В. извадка по удобство

Г. пропорционална (систематична)

7. Изследовател иска да подбере случайна извадка от 50 студенти. Той избира първите 50 студенти, влезли в студентския стол за вечеря. Тази извадка репрезентативна ли е?

А. да Б. не

8. Ако искаме да подберем случайна извадка от 10 лица, трябва да внимаваме да заместим избраните индивиди преди на подберем следващия.

А. вярно Б. невярно

9. Студентите, присъстващи на лекция по дадена дисциплина, са:

А. репрезентативна извадка от студентската общност във ВУЗ-а

### Б. не са репрезентативна извадка от студентската общност във ВУЗ-а

В. нито едно от двете

10. Репрезентативна грешка се среща, защото:

А. повечето интервюиращи не представят точни доклади

### Б. се наблюдава определена извадка вместо съответната популация

В. изследователят налага мнението си при избора на извадката

Г. всичко посочено по-горе

11. Ос­но­в­но­то пра­ви­ло при фор­ми­ра­не на из­ва­д­ка за про­у­ч­ва­не мнението на населението относно работата на семейните лекари сле­д­ва да бъ­де:

А. из­ва­д­ка­та да вклю­ч­ва пре­дим­но ли­ца с хро­ни­ч­ни за­бо­ля­ва­ния

Б. да се обе­з­пе­чи не­пре­д­на­ме­рен по­д­бор на слу­ча­и­те

В. да се по­д­бе­рат мла­ди и об­ра­зо­ва­ни ли­ца

12. Та­б­ли­ца­та за слу­чай­ни­те чи­с­ла слу­жи за:

А. оп­ре­де­ля­не на ве­ро­я­т­но­ст­та на ну­ле­ва­та хи­по­те­за

Б. обо­б­ща­ва­не на ре­зул­та­ти за по­пу­ла­ци­я­та ка­то ця­ло

В. за фор­ми­ра­не на из­ва­д­ка чрез не­пре­д­на­ме­рен по­д­бор

13. Ако ис­хе­ми­ч­на­та бо­лест на сър­це­то се сре­ща 2 пъ­ти по-че­с­то сред мъ­же­те, от­кол­ко­то сред же­ни­те и 3 пъ­ти по-че­с­то сред ли­ца­та над 50 г. в сра­в­не­ние с те­зи под 50 г., кол­ко же­ни над 50 г. би трябвало да има в стра­ти­фи­ци­ра­на из­ва­д­ка от 120 ли­ца?

А. 40 Б. 10 В. 30 Г. 80

14. Ако ис­хе­ми­ч­на­та бо­лест на сър­це­то се сре­ща 2 пъ­ти по-че­с­то сред мъ­же­те, от­кол­ко­то сред же­ни­те и 3 пъ­ти по-че­с­то сред ли­ца­та над 50 г. в сра­в­не­ние с те­зи под 50 г., кол­ко мъже под 50 г. би трябвало да има в стра­ти­фи­ци­ра­на из­ва­д­ка от 120 ли­ца?

А. 40 Б. 20 В. 30 Г. 80

15. Основната идея при извадковите проучвания е да се подбере репрезентативна извадка, от която изследователят да може да направи изводи за популацията.

А. вярно Б. невярно

16. Дадена извадка се счита за случайна, когато тя не е репрезентативна по отношение на популацията.

А. вярно Б. невярно

17. Ако дадена популация се състои от 50% мъже и 50% жени, а подбраната от изследователя извадка има 10% мъже и 90% жени, такава извадка е повлияна от систематична грешка.

А. вярно Б. невярно

18. Стратифицираната извадка включва подбор на най-лесно достъпните елементи от популацията.

А. вярно Б. невярно

19. Случайна извадка е тази, в която 50% от елементите в дадена популация имат еднакъв шанс да попаднат в извадката.

А. вярно Б. невярно

20. Колкото е по-голям размерът на извадката, толкова по-голяма е репрезентативната грешка.

А. вярно Б. невярно

21. Когато една извадка е репрезентативна, тогава тя предоставя валидни данни за извличане на заключения за популацията, от която тази извадка е подбрана.

А. вярно Б. невярно

22. Репрезентативната грешка намалява при нарастване на размера на извадката.

А. вярно Б. невярно

23. Ако размерът на извадката се намали два пъти, репрезентативната грешка ще се увеличи два пъти.

А. вярно Б. невярно

24. Ако една извадка е голяма (напр. повече от 500 случая), то тя е репрезентативна.

А. вярно Б. невярно

25. С нарастване на размера на извадката:

А. намалява репрезентативната грешка

Б. популацията става по-достъпна

В. извадката се повлиява в по-голяма степен от систематична грешка

Г. нито едно от посоченото

26. Репрезентативната извадка:

А. трябва да се състои поне от 500 случая

Б. е обратно пропорционална на корен квадратен от размера на извадката

В. отразява точно основните измерения на популацията

Г. нито едно от посоченото

27. Ако артериалното налягане на пациента се колебае значително в течение на деня, то тогава единично измерване на налягането ще представлява неподходяща извадка.

А. вярно Б. невярно

28. Да предположим, че са ни известни основните подгрупи в дадена популация и непреднамерено са подбрани лица пропорционално на тези подгрупи. Това е:

А. гнездова извадка

Б. проста случайна извадка

В. стратифицирана извадка

Г. многостепенна извадка

29. Гнездовата извадка включва избиране на даден район случайно и обхващане на всички населяващи този район.

А. вярно Б. невярно

30. Случайна извадка е тази, в която:

А. всички елементи са имали равен шанс да попаднат в извадката

Б. е използван непреднамерен подбор на елементите, включени в извадката

В. верни са и двете

# Отговори на въпросите от глава 4:

# 1В; 2В; 3А; 4Б; 5Б; 6В; 7Б; 8Б; 9Б; 10Б; 11Б; 12В; 13В; 14Б; 15А; 16Б; 17А; 18Б; 19Б; 20Б; 21АЯ; 22А; 23Б; 24Б; 25А; 26Б; 27А; 28В; 29А; 30В