Глава 10

сТАТИСТИЧЕСКО ОЦЕНЯВАНЕ: ОТ ИЗВАДКА КЪМ ПОПУЛАЦИЯ

*Г. Грънчарова*

***В тази глава:***

*10.1. Защо е необходимо да изучаваме извадки?*

*10.2. Същност на статистическото оценяване*

*10.3. Основни понятия при статистическото оценяване*

*10.3.1. Стандартна (средна стохастична) грешка*

*10.3.2. Гаранционна вероятност (доверителност)*

*10.3.3. Гаранционен (доверителен) коефициент*

*10.3.4. Максимална (стохастична) грешка*

*10.3.5. Интервал на доверителност (доверителни граници)*

*10.4. Практически стъпки при статическото оценяване*

*10.4.1. Оценка на средни величини*

*10.4.2. Оценка на коефициенти и пропорции*

*10.5. Определяне на минималния размер на извадката за оценка на параметрите в популацията*

*10.5.1. Определяне размера на извадката при количествени променливи*

*10.5.2. Определяне размера на извадката при качествени променливи*

*10.6. Въпроси за самоподготовка*

# 10.1. За­що е не­об­хо­ди­мо да изучаваме из­ва­д­ки?

 Про­у­ч­ва­ни­я­та в медицината и здравеопазването оби­к­но­ве­но имат за цел да оце­нят ти­пи­ч­но­то ни­во на ня­кои по­ка­за­те­ли или че­с­то­та­та на раз­про­с­т­ра­не­ние на да­де­но яв­ле­ние сред оп­ре­де­ле­на по­пу­ла­ция. При проучване на масови явления доста често се оказва невъзможно да бъдат обхванати вси­ч­ки чле­но­ве на съ­о­т­ве­т­на­та по­пу­ла­ция. За­то­ва по не­об­хо­ди­мост из­с­ле­до­ва­те­ли­те често са при­ну­де­ни да на­б­лю­да­ват извадки и на та­зи ос­но­ва да оценяват сре­д­ни­те ни­ва или честотата на съответни явления в по­пу­ла­ция.

Основните причини за проучване на извадки се свеждат най-често до:

* ограничени финансови, времеви, технологични и други ресурси;
* липса на достъп до цялата популация;
* наблюдението на извадки може да е единствен възможен метод за набиране на информация.

Процесът на извличане на заключения за характеристиките на популацията от фактите, получени при наблюдение на извадка, се нарича статистическо заключение или генерализация (обобщаване).

Статистическите заключения си поставят две основни задачи:

* ***оценка на резултатите от наблюдение на извадка и извличане на изводи за параметрите на съответната популация;***
* ***сравняване на резултати от наблюдение на две или повече извадки, т.е. проверка (тестуване) на хипотези за установяване на статистическа значимост***

Използваната при статистическите заключения информация от наблюдение на извадки има известни ограничения по отношение на надеждността, точността и валидността, но въпреки това тя представлява основата за изграждане на медицинското познание, което имаме за човешките популации.

Вси­ч­ки ста­ти­с­ти­че­с­ки за­к­лю­че­ния из­хо­ж­дат от пре­д­по­с­та­в­ка­та, че ***една добра из­ва­д­ка трябва да бъде:***

* **подбрана на основата на случаен непреднамерен подбор, за да се намали вероятността за систематична грешка;**
* ***представителна, за да се подобри валидността й;***
* ***достатъчно голяма по обем, за да се повиши точността на изчисляваните описателни характеристики.***

**10.2. Същност на статистическото оценяване**

 ***Оценката*** се заключава в ***из­по­л­з­ва­не на ре­зул­та­ти­те от про­у­ч­ва­не вър­ху сра­в­ни­тел­но мал­ка из­ва­д­ка ка­то мяр­ка или ин­ди­ка­ция за ни­ва­та на съ­о­т­ве­т­ни­те по­ка­за­те­ли в мно­го по-ши­ро­ка по­пу­ла­ция***.

 Например, количествените променливи в извадката най-често се описват чрез средна величина ****** и стандартно отклонение **s**. По по­до­бен на­чин мо­гат да бъ­дат опи­са­ни и по­пу­ла­ци­и­те като се използват символите: средна величина **µ** и стандартно отклонение **σ**.

Сре­д­на­та ве­ли­чи­на и стан­дар­т­но­то от­к­ло­не­ние за из­ва­д­ка­та се на­ри­чат оце­нъ­ч­ни ин­ди­ка­то­ри (статистики) на из­ва­д­ка­та за не­из­ве­с­т­ни па­ра­ме­т­ри от по­пу­ла­ци­я­та.

|  |  |  |
| --- | --- | --- |
|  | ***Извадка*** | ***Популация*** |
|  | Известна | ***Неизвестна*** |
|  | Статистики | ***Параметри*** |
| ***Средна величина*** |  | **µ** |
| ***Стандартно отклонение*** | **s** | **σ** |

 Имен­но по­ра­ди то­ва, че стой­но­с­тите на параметрите в популацията са неиз­ве­с­т­ни, най-на­пред се по­д­би­рат из­ва­д­ки, установяват се стойностите на извадковите статистики и на тази основа се извличат статистически заключения за па­ра­ме­т­рите в популацията. На­при­мер, при измерване на ди­а­с­то­л­но­то на­ля­га­не при 56 мъ­же пу­ша­чи на въз­раст 40-59 г. са получени ****** = 86 мм Hg и **s** = 14 мм Hg. Би­х­ме ис­ка­ли да зна­ем ка­к­во е средното ни­во на ди­а­с­то­л­но­то кръ­в­но на­ля­га­не при вси­ч­ки мъ­же пушачи на тази възраст.

 Оценката на даден параметър в популацията може да се извърши по два начина: ***точково и интервално оценяване.***

 ***Точковото оценяване*** предоставя оценка на даден популационен параметър чрез една единствена стойност, която той най-вероятно може да приеме и най-често се изразява чрез извадковите статистики (напр. средна аритметична, пропорция и др.). Точковата оценка игнорира допусканата при извадковите проучвания репрезентативна грешка. Поради това тя не може да се разглежда като точна стойност на популационния параметър и няма самостоятелно приложение, а служи за основа на интервалната оценка.

 ***Интервалното оценяване*** се опира на множество стойности, съсредоточени около точковата оценка, които формират определен интервал, в границите на който при определено ниво на гаранционна вероятност се предполага, че се намира истинската стойност на параметъра в популацията.

# 10.3. Основни понятия при статистическото оценяване

Пре­ди да бъ­де на­п­ра­ве­но за­к­лю­че­ние за па­ра­ме­т­ри­те на по­пу­ла­ци­я­та на ба­за­та на статистиките от из­ва­д­ка­та тря­б­ва да изя­с­ним съ­щ­но­ст­та на следните основни по­ня­тия:

* *стандартна (средна стохастична ) грешка;*
* *гаранционна вероятност (доверителност);*
* *гаранционен (доверителен) коефициент;*
* *максимална стохастична грешка*
* *интервал на доверителност.*

**10.3.1. Стандартна (средна стохастична ) грешка**

***Стандартната грешка*** измерва стандартното отклонение на разпределението на дадена извадкова статистика.Най-често се изчислява стандартната грешка на средната аритметична величина **Ако от една и съща популация са извлечени няколко извадки и е изчислена средната аритметична величина за всяка извадка, то стандартното отклонение на получените средни величини се нарича *стандартна грешка на средната величина.* Като такава тя характеризира точността** наизчислените описателни статистики в из­ва­д­ка­та при из­по­л­з­ва­не­то им ка­то оце­нъ­чни индикатори за параметрите в по­пу­ла­ци­я­та. На практика, стандартната грешка се изчислява по данни само от една репрезентативна извадка и размерът на грешката може да се контролира чрез повлияване на определящите я фактори.

 ***От какво зависи величината на стандартната грешка***?

 ***На пър­во мя­с­то***, ва­ри­а­бил­но­ст­та на да­де­на статистика в из­ва­д­ка­та (напр. за средната величина) за­ви­си ***от ва­ри­рането на ин­ди­ви­ду­ал­ни­те на­б­лю­де­ния*,** от ко­и­то е ком­по­зи­ра­на из­ва­д­ка­та. Кол­ко­то по­ве­че кри­те­рии и изи­с­к­ва­ния за еднородност на извадката са за­ло­же­ни при нейното сфор­ми­ра­не­, тол­ко­ва по-мал­ка е ва­ри­а­бил­но­ст­та, т. е. сре­д­на­та величина е съ­п­ро­во­де­на с по-мал­ко стан­дар­т­но от­к­ло­не­ние.

***На вто­ро мя­с­то***, ва­ри­а­бил­но­ст­та на сре­д­на­та величина в из­ва­д­ка­та за­ви­си ***от раз­ме­ра на из­ва­д­ка­та*.** На­при­мер, сре­д­ни­ят ръст на мъ­же сту­ден­ти- ме­ди­ци, из­чи­с­лен вър­ху 100 на­б­лю­де­ния ще бъ­де мно­го по-то­чен от то­зи, из­чи­с­лен вър­ху 15 на­б­лю­да­ва­ни слу­чая.

 Сле­до­ва­тел­но, ***ва­ри­а­бил­но­ст­та на сре­д­ната величина в из­ва­д­ка­та на­ра­с­т­ва с ин­ди­ви­ду­ал­ното вариране и на­ма­ля­ва с уве­ли­ча­ва­не на раз­ме­ра на из­ва­д­ка­та.*** То­ва се вгра­ж­да във фор­му­лата за стандартната гре­ш­ка по следния начин:

 s

 = , където

 

 - стандартна грешка на средната величина в извадката

 ***s*** - стандартно отклонение на средната величина в извадката

 ***n*** - брой наблюдавани случай в извадката

 Следователно, стандартната грешка на извадковата статистика е право пропорционална на стандартното отклонение и обратно пропорционална на размера на извадката. Стандартната грешка винаги ще бъде по-малка от самото стандартно отклонение, тъй като тя се получава като отношение на стандартното отклонение и квадратен корен от размера на извадката, а размерът на извадката винаги е по-голям от единица, т.е. квадратният корен винаги ще е по-голям от единица.

**10.3.2. Гаранционна вероятност (доверителност)**

Под ***гаранционна вероятност (доверителност)*** в широк смисъл на понятието се разбира ***вероятността, с която се подкрепя дадено твърдение.***

***Статистическото заключение за параметрите в популацията винаги има вероятностен характер.*** Следователно, ***в контекста на интервалното оценяване, гаранционната вероятност е вероятността, с която се подкрепя твърдението, че оценяваният параметър от популацията попада в съответни доверителни граници.***

***Ве­ро­я­т­но­ст­та мо­же да се раз­г­ле­ж­да ка­то ска­ла за из­мер­ва­не сте­пен­та на шан­са ме­ж­ду 0 и 1.*** По-че­с­то тя се пре­д­с­та­вя ка­то про­цент ме­ж­ду 0% и 100% със зна­че­ния на край­ни­те то­ч­ки - не­въз­мо­ж­но и си­гур­но. Означава се най-често с Р.

 ***невъзможно сигурно***

 **0% 5% 50% 95% 100%**

**/ / / / / 0.05 0.5 0.95 1.0**

***малък шанс равно вероятно голям шанс***

При оцен­ка­та на ре­зул­та­ти­те от ре­п­ре­зен­та­ти­в­ни­те про­у­ч­ва­ния не е ну­ж­но да зна­ем как се из­чи­с­ля­ват ве­ро­я­т­но­с­ти­те, но е по­ле­з­но да зна­ем връ­з­ка­та ме­ж­ду че­с­то­т­ни­те раз­пре­де­ле­ния и ве­ро­я­т­но­ст­та.

 Има ре­ди­ца стан­дар­т­ни раз­пре­де­ле­ния, от ко­и­то най-че­с­то сре­ща­но е нор­мал­но­то раз­пре­де­ле­ние. При него 68,2% от слу­ча­и­те прилягат най-бли­з­ко до сре­д­на­та стой­ност - в интервала ***± s,*** т.е. има 68,2% ве­ро­я­т­ност за те­зи стой­но­с­ти да се раз­по­ла­гат око­ло сре­д­но­то ни­во. По същия на­чин раз­съ­ж­да­ва­ме за слу­ча­и­те, намиращи се на две стан­дар­т­ни от­к­ло­не­ния вля­во или вдя­с­но от сре­д­на­та стой­ност, за ко­и­то ве­ро­я­т­но­ст­та е 95% и т. н.

**10.3.3. Гаранционен (доверителен) коефициент**

Ни­во­то на ве­ро­я­т­ност (в %) при нор­мал­но­то раз­пре­де­ле­ние съответства на точно определена чи­с­ле­на стой­ност на ***t-кри­те­рия на Стю­дент***, поради което той се нарича ***гаранционен*** ***(доверителен) коефициент***. При го­лям брой слу­чаи (над 120) съотношението е сле­д­но­то:

|  |  |
| --- | --- |
| **Стойност на t-критерий** | **Вероятност в %** |
| 0.5 | 38.2% |
| 1.00 | 68.2% |
| 1.64 | 90.0% |
| 1.96 | 95.0% |
| 2.58 | 99.0% |
| 3.29 | 99.999% |

 В медицината и здравеопазването статистическите заключения следва да бъдат подкрепяни с ***ви­со­ко ниво на гаранционна ве­ро­я­т­ност*** - не по-мал­ка от 95%, а в ня­кои слу­чаи, ко­га­то ста­ва ду­ма за гра­ни­ч­ни об­ла­с­ти ме­ж­ду жи­во­та и смърт­та - дори над 99%. Стойността на доверителния коефициент t в тези случаи при извадка над 120 случая ще бъде приблизително равна на 2, а при 99% - над 2.5.

При размер на извадката под 120 случая стойността на доверителния коефициент се определя по специална таблица за критичните стойности на t-критерия (***Приложение 1***). При едно и също ниво на гаранционна вероятност стойността на t намалява с увеличаване на размера на извадката, а при един и същ размер на извадката t нараства с увеличаване на нивото на гаранционната вероятност.

**10.3.4. Максимална стохастична грешка**

 ***Максималната стохастична грешка*** характеризира максималното отклонение на стандартната грешка от истинската стойност на параметъра и се определя като произведение на стандартната грешка и доверителния коефициент. Най-често се означава със символа **∆** и за средната аритметична величина **∆ = **

 Практически максималната стохастична грешка се използва при изчисляване интервала на доверителност, тъй като коефициентът t е свързван пряко с възприетото ниво на гаранционна вероятност.

10.3.5. Интервал на доверителност (доверителни граници)

Както подчертахме, при статистическото оценяване на параметрите в популацията се използва вероятностен подход, който се свежда до определяне на ***интервал на доверителност (доверителни граници).***

***Доверителният интервал*** (означава се с **ДИ или CI** – от Confidence Interval) ***представлява интервал, в гра­ни­ци­те на ко­й­то при възприетата от изследователя га­ран­ци­он­на ве­ро­я­т­ност се намира истинската стойност на параметъра за популацията.***

***Доверителният интервал*** се построява като към точковата оценка на извадковата статистика се прибави и извади максималната стохастична грешка. Последната включва в себе си гаранционната вероятност чрез стойността на доверителния коефициент **t** и по такъв начин се оформят долната и горната граница на интервала. Например, за средната аритметична величина в популацията доверителният интервал ще има следния вид:

**CI = μ1 ÷ μ2 = ** ± ∆ = ** ± t.****,** където **μ1 =** - ∆** и **μ2 = ** + ∆**

Следователно, истинската стойност на съответния параметър за по­пу­ла­цията ще се на­ми­ра в пре­де­ли­те на стойността на оценъчния индикатор за из­ва­д­ка­та плюс ми­нус максималната стохастична гре­ш­ка, т.е. **μ** щебъде не по-малка от **μ1** и непо-голяма от **μ2.**

Гаранционната вероятност на статистическия извод не зависи от размера на извадката – тя се задава предварително от изследователя, но при едно и също ниво на вероятност интервалната оценка ще бъде по-точна (интервалът ще бъде по-тесен), ако размерът на извадката е по-голям.

При един и същ масив от данни, ширината на доверителния интервал зависи от стойността на t-критерия. При Р=95% стойността на t-критерия е по-малка, отколкото при 99%, а оттук и по-тесен доверителен интервал. Обратно, при по-висока гаранционна вероятност интервалът на доверителност ще бъде по-широк, тъй като стойността на t нараства (при равни други условия - напр. един и същ размер на извадката и стандартното отклонение).

Степен на свобода

**Терминът *“степен на свобода”* (означава се с *“к”* или *df* – от degree of freedom) за дадена променлива величина се използва за характеристика на броя на независимите части от информацията, съдържаща се в дадена статистика, т.е. df представлява брой резултати, които могат свободно да варират при изчисляването на дадена статистика, така че да не се промени крайния резултат. Например, ако в случайна извадка от n наблюдения сме изчислили средната аритметична, то за да получим същия резултат** n-1 от измерванията могат да варират, но последното измерване трябва да има такава стойност, че да не се промени сумата от направените измервания.

Определянето на степените на свобода е задължително условие при работа с таблиците за критичните стойности на t-критерия при интервално оценяване, както и при проверка на хипотези чрез параметрични и непараметрични методи и при други статистически методи на анализ.

10.4. Пра­к­ти­че­с­ки стъ­п­ки при статистическото оценяване

## 10.4.1. Оценка на средни величини

 Обо­б­ща­ва­не­то на ре­зул­та­ти­те за по­пу­ла­ци­я­та на основата на наблюдения на извадка пре­ми­на­ва през сле­д­ни­те ета­пи:

 ***I етап - Определяне на стандартната грешка***

В примера от ***табл. 5.6*** (стр.67) диастолното налягане в извадка от 56 мъже силни пушачи на възраст 40-59 г. ****** = 86 мм Hg и s = 14 мм. Стандартната грешка се определя по формулата:

 **s** 14 14

 = = = = ± 1.8 mm Hg

   7.42

 ***II етап - Из­бор на стой­ност на t-кри­те­рия*** съ­о­б­ра­з­но броя на слу­ча­и­те в из­ва­д­ка­та и ни­во­то на га­ран­ци­он­на ве­ро­я­т­ност, с ко­я­то же­ла­ем да бъ­дат по­д­к­ре­пе­ни на­ши­те из­во­ди.

 При 56 на­б­лю­де­ния (сте­пен на сво­бо­да df = 56 – 1 = 55) и га­ран­ци­он­на ве­ро­я­т­ност P = 95% от та­б­ли­ца­та за кри­ти­ч­ни­те стой­но­с­ти на t-кри­те­рия (***При­ло­же­ние 1***) на­ми­ра­ме t = 2.00 (в реда съответстващ на df = 60).

 ***III етап - Определяне на CI за средната величина в популацията***

**CI = μ1 ÷ μ2 = ** ± ∆ = ** ± t.****,** където **μ1 =** - ∆** и **μ2 = ** + ∆**

Изчисляваме **μ1 =**86 - 2.00 х 1.8 = 82.4 mm Hg и **μ2** = 86 + 2.00 х 1.8 = 89.6mm Hg

Построяваме доверителния интервал **CI = μ1 ÷ μ2** = 82.4 ÷ 89.6.

 ***IV етап - Фор­му­ли­ра­не на за­к­лю­че­ни­е за стой­но­ст­та на параметъра в по­пу­ла­ци­я­та***

 На основата на посочения пример мо­же да се на­п­ра­ви извод, че сре­д­на­та стой­ност на ди­а­с­то­л­но­то на­ля­га­не за ця­ла­та по­пу­ла­ция (мъ­же, сил­ни пу­ша­чи, на въз­раст 40-59 г.) се оча­к­ва да бъ­де не по-ни­с­ка от 82.4 и не по-ви­со­ка от 89.6 мм Hg; то­ва твър­де­ние е по­д­к­ре­пе­но с га­ран­ци­он­на ве­ро­я­т­ност 95%.

 Уве­ли­ча­ва­не­то на га­ран­ци­он­на­та ве­ро­я­т­ност на из­во­ди­те за по­пу­ла­ци­я­та изи­с­к­ва при­е­ма­не на по-ви­со­ка стой­ност на t-кри­те­рия, а сле­до­ва­тел­но и ши­ри­на­та на ин­тер­ва­ла на до­ве­ри­тел­ност ще се уве­ли­чи. По­с­ле­д­но­то е на­пъл­но ло­ги­ч­но, тъй ка­то ще на­ма­лее ве­ро­я­т­но­ст­та за не­то­ч­ност.

## 10.4.2. Оценка на коефициенти и пропорции

 Оцен­ката на коефициенти за честота и про­по­р­ции се различава само по фор­му­лата за стан­дар­т­ната гре­ш­ка. Използват се следните символи:

|  |  |  |
| --- | --- | --- |
|  | ***Извадка*** | ***Популация*** |
|  | Известна | ***Неизвестна*** |
|  | Статистики | ***Параметри*** |
| ***Коефициенти/пропорции*** | ***р*** | ***π*** |
| ***Стандартно отклонение*** | ***s*** | ***σ*** |

Преминава се през същите етапи на работа:

 ***I етап -*** ***Оп­ре­де­ля­не на стан­дар­т­на­та гре­ш­ка*** по фор­му­ла­та:

 **sp = ,** където:

**sp** e стан­дар­т­на­та гре­ш­ка на ко­е­фи­ци­ен­та (ин­тен­зи­в­ния по­ка­за­тел) или на про­по­р­ци­я­та (ек­с­тен­зи­в­ния по­ка­за­тел);

**р** е из­чи­с­ле­ни­ят от из­ва­д­ка­та по­ка­за­тел;

**q** e про­ти­во­по­ло­ж­но­то до 1 или 100%, **q = 1 – p** или **100 - р**

 ***II етап - Оп­ре­де­ля­не на ве­ли­чи­на­та на t-кри­те­рия*** според желаното от изследователя ни­во на га­ран­ци­он­на ве­ро­я­т­ност на из­во­ди­те и сте­пен­та на сво­бо­да за из­ва­д­ка­та (по та­б­ли­ца­та за кри­ти­ч­ни­те стой­но­с­ти на t-критерия)

 ***III етап - Оп­ре­де­ля­не на ин­тер­ва­ла на до­ве­ри­телност*** за параметъра в по­пу­ла­ци­я­та – **CI=*π*1÷*π*2 = p ± t.sp ,** където ***π*1 = p - t.sp** и ***π*2 = p + t.sp**

# *IV етап - Фор­му­ли­ра­не на за­к­лю­че­ни­е­то за стой­но­ст­та на параметъра в по­пу­ла­ци­я­та*

# 10.5. Оп­ре­де­ля­не на минималния раз­мер на из­ва­д­ка­та за оценка на па­ра­ме­т­ри­те в по­пу­ла­ци­я­та

Минималният размер на извадката зависи от:

* ***целта и постановката на проучването;***
* ***плана за статистически анализ*** (методи за статистическа обработка);
* ***точността на измерванията***, които трябва да бъдат направени;
* ***степента на точност при обобщаване на данните за популацията,*** т.е. допусканата стандартна грешка;
* ***гаранционната вероятност (доверителност)*** на заключенията

Практически, изследователите за­да­ват пре­д­ва­ри­тел­но же­ла­но­то ни­во на га­ран­ци­он­на ве­ро­я­т­ност на статистическото заключение за популацията. Оп­ре­де­ля се чрез пи­ло­т­но про­у­ч­ва­не или по ли­те­ра­тур­ни дан­ни оча­к­ва­но­то ни­во на сре­д­ната ве­ли­чи­на или про­по­р­цията в извадката и стойността на до­пу­с­ти­мата стан­дар­т­на гре­ш­ка. Чрез пре­об­ра­зу­ва­не на фор­му­ла­та за максимална стохастична гре­ш­ка се оп­ре­де­ля не­об­хо­ди­ми­ят раз­мер на из­ва­д­ка­та (n).

**10.5.1. Определяне размера на извадка при количествени променливи**

При подбор на извадка за изучаване на количествена характеристика, изследователят трябва да посочи:

* допустимата максимална грешка (в абсолютни или относителни единици);
* стандартното отклонение на променливата в популацията;
* възприетата гаранционна вероятност на изводите – обикновено 95%.

**Пример:** Изследовател иска да оцени средното ниво на хемоглобина в дадена общност. Предварителната информация, с която разполага е, че средната е около 150mg/l със стандартно отклонение 32mg/l. Ако допустимата максимална грешка е до 5mg/l, колко лица трябва да бъдат включени в проучването?

От формулите **∆** = t. и = s/ намираме, че **∆** = t. s/

Оттук n = t2 . s2/**∆**2 = [(1,96)2 . (32)2]/(5)2=157.4 лица

Ако размерът на популацията, от която ще се подбере извадката, е известен (напр. 3000 души), то необходимия размер на извадката ще бъде:

n = (1,96)2 . (32)2 / [(5)2 + (1,96)2 . (32)2 / 3000] = 149.5 лица; т.е. извадката трябва да включва поне 150 лица.

**10.5.2. Определяне на размера на извадка при качествени променливи**

При подбор на извадка за изучаване на качествена характеристика, изследователят трябва да посочи:

* приблизителната стойност на пропорцията (р);
* допустимата максимална грешка (в абсолютни или относителни единици);
* стандартното отклонение на променливата в популацията;
* възприетата гаранционна вероятност на изводите.

От формулите **∆** = t.  и = намираме, че **∆** = t .

Оттук n = t2 . p . q/**∆**2

**Пример**: р = 26% (0,26); **∆**= 3% (0,03); Р = 95%; t = 1.96

Тогава: n = t2 . p . q/**∆**2 = (1.96)2 . 0,26 . 0,74/(0,03)2 = 821,2 лица

Ако размерът на популацията, от която ще се подбере извадката, е известен (напр. 3000 души), то необходимия размер на извадката ще се коригира по следния начин:

n = 821,2 / (1 + 821,2/3000) = 644,7 лица; т.е. извадката трябва да включва поне 645 лица.

Ако стойността на р е неизвестна, най-добре да се приеме р=50%. Тогава и q ще бъде 50%. Произведението ***p.q*** ще бъде максимално и по такъв начин и стандартната грешка ще има максимална стойност. Определеният на тази основа обем на извадката няма опасност да е твърде малък или прекалено голям.

#### 10.6. Въпроси за самоподготовка

1. Оценката (статистическото заключение) се използва с цел да се опишат специфичните характеристики на данните в извадката.

А. вярно Б. невярно

2. Статистическото заключение включва оценка на параметрите на популацията на основата на фактите, получени от извадка.

А. вярно Б. невярно

3. Използването на средната аритметична величина от извадката като оценъчен индикатор за средната в популацията е пример за статистическо заключение (извод).

А. вярно Б. невярно

4. С ка­к­ва цел се пра­ви оцен­ка на данни от ре­п­ре­зен­та­ти­в­ни извадки?

А. за да се ус­та­но­ви съ­ще­с­т­ве­но ли е раз­ли­чи­е­то ме­ж­ду тях

##### Б. за да се обо­б­щят дан­ни­те от извадката за по­пу­ла­цията

В. за да се уе­д­на­к­ви стру­к­ту­ра­та на сре­да­та, от ко­я­то са из­чи­с­ле­ни данните

5. Целта на използване на извадка и изчисляване на средна аритметична е:

А. да се намери средната за извадката

Б. да се определи разсейването на извадката

##### В. да се оцени средната аритметична за популацията

Г. да се оцени размера на извадката

6. Всяка описателна характеристика, изчислена от извадка, се нарича статистика (оценъчен индикатор или показател) на извадката.

А. вярно Б. невярно

7. Всяка характеристика на разпределението на дадена популация се нарича:

А. стандартно отклонение

Б. стандартна грешка

В. параметър

8. Ако подборът на извадката е случаен непреднамерен, грешката на извадката е нула.

А. вярно Б. невярно

9. Грешката на средната аритметична величина в извадката:

А. се дължи на неподходящи техники на подбор на извадката

Б. намалява с увеличаване на размера на извадката

В. не зависи от стандартното отклонение

Г. винаги е равна на единица

10. Ка­к се про­ме­ня ре­п­ре­зен­та­ти­в­на­та гре­ш­ка при че­ти­ри­к­ра­т­но уве­ли­ча­ва­не на броя на на­б­лю­да­ва­ни­те слу­чаи?

А. ре­п­ре­зен­та­ти­в­на­та гре­ш­ка на­ма­ля­ва два пъ­ти

Б. ре­п­ре­зен­та­ти­в­на­та гре­ш­ка на­ра­с­т­ва

В. ре­п­ре­зен­та­ти­в­на­та гре­ш­ка на­ма­ля­ва че­ти­ри пъ­ти

11. Ка­к­во тря­б­ва да на­п­ра­вим, ако же­ла­ем на на­ма­лим три пъ­ти ре­п­ре­зен­та­ти­в­на­та гре­ш­ка?

А. да на­ма­лим девет пъ­ти броя на на­б­лю­да­ва­ни­те слу­чаи

Б. да уве­ли­чим броя на на­б­лю­да­ва­ни­те слу­чаи девет пъти

В. да уве­ли­чим три пъ­ти броя на на­б­лю­да­ва­ни­те слу­чаи

12. С нарастване на размера на извадката средните стойности в извадките се приближават до средната за популацията .

А. вярно Б. невярно

13. Колкото е по-голяма репрезентативната грешка на извадката, толкова интервалът на доверителност е по-малък.

А. вярно Б. невярно

14. За големи по размер извадки, грешката на средната аритметична е по-голяма от стандартното отклонение.

А. вярно Б. невярно

15. Репрезентативната грешка на средната аритметична:

А. се определя теоретично

Б. наподобява точно стандартното отклонение

В. се влияе от размера на извадката

##### Г. се характеризира с всичко посочено по-горе

16. С увеличаване на размера на извадката, репрезентативната грешка нараства.

А. вярно Б. невярно

17. Извадка от 3600 случая е подбрана случайно от безкрайно голяма популация. Стандартното отклонение е равно на 10. Грешката на средната аритметична е:

А. 2/15 Б. 1/6 В. 4/5 Г. 10

18. Коефициентът на доверителност представлява вероятността, че даден неизвестен параметър има определена стойност.

А. вярно Б. невярно

19. Коефициентът на доверителност представлява вероятността, че даден неизвестен параметър ще се намира в рамките на даден интервал.

А. вярно Б. невярно

20. По­ня­ти­е­то “ин­тер­вал на до­ве­ри­тел­ност” се из­по­л­з­ва за:

А. из­г­ра­ж­да­не на нор­ма­ти­ви

Б. за генерализиране на данни от извадки за популации

В. за про­вер­ка на хи­по­те­зи от­но­с­но вли­я­ни­е­то на да­де­ни фа­к­то­ри

21. Как се оп­ре­де­ля до­ве­ри­тел­ния ин­тер­вал на сре­д­ната ве­ли­чи­на в популацията?

А. към сре­д­на­та от извадката се при­ба­вя и из­ва­ж­да 0.5 пъ­ти стан­дар­т­но от­к­ло­не­ние

Б. към сре­д­на­та от извадката се при­ба­вя и из­ва­ж­да 2 пъ­ти репрезентативната гре­ш­ка

В. към сре­д­на­та се при­ба­вя и из­ва­ж­да t пъ­ти ре­п­ре­зен­та­ти­в­на­та гре­ш­ка

22. 95%-ят интервал на доверителност, сравнен с 99%-я, е:

А. по-широк

Б. по-тесен

В. с по-голяма вероятност да съдържа средната за популацията

23. 95%-ят интервал на доверителност в даден експеримент е 72÷79. Следователно:

А. с вероятност 0.05 може да се твърди, че μ попада в интервала 72÷79.

Б. с вероятност 0.95 може да се твърди, че интервалът 72÷79 съдържа ******

В. с вероятност 0.95 може да се твърди, че интервалът 72÷79 съдържа μ

24. При извадка от 16 случая от популация с нормално разпределение, каква е стойността на t-критерия за 99% интервал на доверителност?

А. 2.602 Б. 2.326 В. 2.576 Г. 2.947

25. Подбрана е извадка от 25 наблюдения от нормална популация. Границите на 90% доверителен интервал при ****** = 30 и **s2 = 9** ще бъдат:

А. 30 **±** 9.00 Б. 30 ± 0.79 В. 30 ± 1.03Г. 30 ± 0.47

26. Подбрана е извадка от 16 наблюдения от популация с нормално разпределение. Долната граница на 98% доверителен интервал при **** = 3.2** и **s = 4** ще бъде:

###### А. 0.6 Б. 2.555 В. 0.62 Г. 5.8

27. При извадка от 26 наблюдения каква стойност на t-критерия е нужна за определяне на 95% доверителен интервал за средната в популацията?

А. 2.064Б. 2.060 В. 1.711 Г. 1.96

28. Девет мъже с генетично състояние, причиняващо затлъстяване, са включени в програма за намаляване на теглото. След 4 месеца данните за снижението на теглото показват **** =** 11.2кги **s =** 9kг**.** Границите на 95%доверителен интервал за средната в популацията, от която е извлечена извадката (допускайки нормално разпределение и непреднамерен подбор) ще бъдат:

А. 11.2 ±1.96 х 3Б. 11.2 ±1.86х3 **В**. 11.2 ±2.306 х 3

29. Социологическо проучване върху извадка от 400 души показва, че 56% от изследваните лица ще гласуват в подкрепа на кандидат А. в предстоящи избори. 95%-я интервал на доверителност за относителния дял на избирателите от популацията, които ще подкрепят кандидат А. ще бъде:

А. 0.56 **±** 1.96 х **** Б. 0.44 **±** 1.96 х ****

В. 0.56 ± 1.96 х **** Г. 0.56 **±** 1.64 х ****

30. При проучване на мнението на 100 студенти по медицина, 50 от тях поддържат въвеждането на проблемно базирано обучение (ПБО). Ако р е относителният дял на студентите “за въвеждане на ПБО”, то 99% доверителен интервал се представя чрез:

А. 0.5 **±** 0.098 Б. 0.5 **±** 0.112 В. 0.5 ± 0.13Г. 0.5 **±** 0.135

31. В извадка от 400 жени на възраст 40-59 г. непушачите са 80%. За цялата популация жени на същата възраст при Р=95% делът този дял е:

А. 0.768 **÷** 0.832 Б. 0.761 ÷ 0.839В. 0.775 ÷ 0.825 Г. 0.790 ÷ 0.810

32. За случайна извадка от 16 болници е установено, че средната цена aа лечение и възстановяване след инфаркт на миокарда е 3000 лева и s = 500 лева. Доверителният 99% интервал за всички болници ще бъде (в цели числа):

А. 3100 **÷** 3350Б. 2850 **÷** 3330 В. 2632 ÷ 3368 Г. 2890 **÷** 3230

33. При положение, че 99% доверителен интервал за μ е 42 **÷** 58, то кои от следните числа биха били граници на 95% доверителен интервал?

А. 43 и 57 Б. 41 и 59 В. 41 и 57

34. Психолог иска да определи средния коефициент на интелигентност (IQ) на учениците от последния клас на средните училища. Той избира извадка от 25 ученика и установява средна аритметична за IQ = 103 и стандартно отклонение = 12. Кой от следните интервали е 90% доверителен интервал за средния IQ на всички ученици от последния клас на средните училища?

А. 98.9 ÷ 107.1 Б. 97.6 ÷ 108.3 В. 98.0 ÷ 108.0 Г. 94.5 ÷ 111.5

35. Случайна извадка от 100 семейства показва, че по годишни доходи 20 семейства попадат в “крайно бедни”. Оценете относителния дял (в %) на “крайно бедните семейства”, използвайки 95% доверителен интервал.

А. 15.6 **÷** 23.2 Б. 14.8 ÷ 24.4 В. 12.1 ÷ 27.9 Г. 10.6 ÷ 21.8

36. Доверителният интервал за **µ** ще бъде по-малък, ако коефициентът на доверителност е 0.90, отколкото ако той е равен на 0.95.

А. вярно Б. невярно

37. 95%-ят доверителен интервал е два пъти по-широк от 90%-я.

А. вярно Б. невярно

38. При всеки размер на извадка може да се извлече високо ниво на доверителност на интервалната оценка.

А. вярно Б. невярно

39. При интервална оценка на средна величина за дадена популация, ширината на интервала може да бъде стеснена чрез:

А. увеличаване на размера на извадката

Б. намаляване на коефициента на доверителност

В. намаляване на стандартното отклонение

##### Г. всичко посочено е вярно

40. Кое от посоченото води до най-тесен интервал на доверителност:

### А. голяма извадка и интервал на доверителност 0.95

Б. голяма извадка и интервал на доверителност 0.99

В. малка извадка и интервал на доверителност 0.95

41. Колкото е по-широк интервалът на доверителност, толкова е по-ниска степента на доверителност в него.

А. вярно Б. невярно

42. При дадена ситуация, колкото е по-широк интервалът на доверителност, толкова е по-ниска степента на доверителност в него.

А. вярно Б. невярно

43. При по-голяма извадката има по-широк доверителен интервал.

А. вярно Б. невярно

44. При доверителен интервал за средната в нормално разпределение, използвайки t- критерий, студент А. работи с гаранционна вероятност 0.95, а студент Б. – с 0.99. Кое от следните твърдения е вярно?

А. Интервалът на студент Б. винаги ще бъде по-тесен от интервала на студент А.

Б. Интервалите на студент Б. и на студент А. са еднакви

В. Интервалът на студент Б. е по-широк от интервала на студент А.

**При­мер:** За да се ус­та­но­ви сре­д­но­то те­г­ло на но­во­ро­де­ни­те мом­че­та е по­д­б­ра­на слу­чай­на из­ва­д­ка от 100 но­во­ро­де­ни мом­че­та в ед­на АГ бол­ни­ца. Тя­х­но­то сре­д­но те­г­ло е би­ло 3000 гра­ма със стан­дар­т­но от­к­ло­не­ние 100 гра­ма.

**Въ­п­ро­си 45-47 се от­на­сят за те­зи дан­ни:**

45. Ка­к­ва е ре­п­ре­зен­та­ти­в­на­та гре­ш­ка на сре­д­на­та в то­зи при­мер?

А. 25 Б. 10 В. 2

46. За да из­чи­с­лим 95% до­ве­ри­те­лен ин­тер­вал, ка­к­ва стой­ност на t-кри­те­рия тря­б­ва да при­е­мем (по та­б­ли­ца­та за кри­ти­ч­ни­те стой­но­с­ти на t)?

А. 1.980 Б. 1.960 В. 2.358

47. Ме­ж­ду кои два въз­мо­ж­ни ре­зул­та­та мо­жем да бъ­дем 95% си­гур­ни, че ще се на­ми­ра ис­тин­с­ко­то сре­д­но те­г­ло за вси­ч­ки но­во­ро­де­ни мом­че­та?

А. 2980 ÷ 3020 Б. 2900 ÷ 3100 В. 2800 ÷ 3200

**При­мер:** По­д­б­ра­на е ре­п­ре­зен­та­ти­в­на из­ва­д­ка от 25 па­ци­ен­ти и е из­ме­ре­но си­с­то­л­ното на­ля­га­не. Из­чи­с­ле­на е сре­д­на ари­т­ме­ти­ч­на ******= 115 mmHg и стан­дар­т­но от­к­ло­не­ние s = 10.

**Въ­п­ро­си 48-50 се от­на­сят за те­зи дан­ни:**

48. Ка­к­ва е репрезентативната гре­ш­ка на сре­д­на­та величина за та­зи из­ва­д­ка?

А. 10 Б. 20 В. 2

49. За да из­чи­с­лим 99% до­ве­ри­те­лен ин­тер­вал, ка­к­ва стой­ност на t - кри­те­рия тря­б­ва да при­е­мем по та­б­ли­ца­та?

А. 2.492 Б. 2.797 В. 2.787

50. Ка­къв е 99% до­ве­ри­те­лен ин­тер­вал на сре­д­на­та величина в то­зи при­мер?

А. 110.0 ÷ 120.0 Б. 109.4 ÷ 120.6 В. 113.0 ÷ 117.0

**Отговори на въпросите в глава 10:**

1Б; 2А; 3А; 4Б; 5В; 6А; 7В; 8Б; 9Б; 10А; 11Б; 12А; 13Б; 14Б; 15Г; 16Б; 17Б; 18Б; 19А; 20Б; 21В; 22Б; 23В; 24Г; 25В; 26А; 27Б; 28В; 29Б; 30В; 31Б; 32В; 33А; 34А; 35В; 36А; 37Б; 38А; 39Г; 40А; 41Б; 42Б; 43Б; 44В; 45Б; 46А; 45А; 48В; 49Б; 50Б