

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 статистическото 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 ѝ;*
- * *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 \bar{x} 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.

	<i>Извадка Известна Статистики</i>	<i>Популация Неизвестна Параметри</i>
<i>Средна величина</i>	\bar{x}	μ
<i>Стандартно отклонение</i>	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 $\bar{x} = 86$ 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 на изчислените 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 заложили при нейното сформирание, 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**

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-близо до средната стойност - в интервала $\bar{x} \pm s$, 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 <i>t</i> -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 вероят-

ност 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 Δ и за средната аритметична величина $\Delta = t \cdot s_{\bar{x}}$

Практически максималната стохастична грешка се използва при изчисляване интервала на доверителност, тъй като коефициентът t е свързан 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 **СИ** – от Confidence Interval) **представява интервал, в границите на който при възприетата от изследвателя 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:

$$CI = \mu_1 \div \mu_2 = \bar{x} \pm \Delta = \bar{x} \pm t \cdot s_{\bar{x}}, \text{ където } \mu_1 = \bar{x} - \Delta \text{ и } \mu_2 = \bar{x} + \Delta$$

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 ще бъде не по-малка от μ_1 и не по-голяма от μ_2 .

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 $P=95\%$ 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 “*k*” 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. $\bar{x} = 86$ мм Hg и $s = 14$ мм. Стандартната грешка се определя по формулата:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{14}{\sqrt{56}} = \frac{14}{7.42} = \pm 1.8 \text{ mm Hg}$$

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 I**) намираме $t = 2.00$ (в реда съответстващ на $df = 60$).

III етап - Определяне на CI за средната величина в популацията

$$CI = \mu_1 \div \mu_2 = \bar{x} \pm \Delta = \bar{x} \pm t \cdot s_{\bar{x}}, \text{ където } \mu_1 = \bar{x} - \Delta \text{ и } \mu_2 = \bar{x} + \Delta$$

Изчисляваме $\mu_1 = 86 - 2.00 \times 1.8 = 82.4 \text{ mm Hg}$ и $\mu_2 = 86 + 2.00 \times 1.8 = 89.6 \text{ mm Hg}$

Построяваме доверителния интервал $CI = \mu_1 \div \mu_2 = 82.4 \div 89.6$.

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 mm 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:

	<i>Извадка Известна Статистики</i>	<i>Популация Неизвестна Параметри</i>
<i>Коефициенти/пропорции</i>	<i>p</i>	<i>π</i>
<i>Стандартно отклонение</i>	<i>s</i>	<i>σ</i>

Премахва се през същите етапи на работа:

I етап - Определяне на стандартната грешка по формулата:

$$s_p = \sqrt{p \cdot q / n}, \text{ където:}$$

s_p е стандартната грешка на коефициента (интензивния показател) или на пропорцията (екстензивния показател);

p е изчисленият от извадката показател;

q е противоположното до 1 или 100%, $q = 1 - p$ или $100 - p$

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 – $CI = \pi_1 \div \pi_2 = p \pm t \cdot s_p$, където $\pi_1 = p - t \cdot s_p$ и $\pi_2 = p + t \cdot s_p$

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;
- * **степената 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 $\Delta = t \cdot s_{\bar{x}}$ и $s_{\bar{x}} = s/\sqrt{n}$ намираме, че $\Delta = t \cdot s/\sqrt{n}$

Отгук $n = t^2 \cdot s^2/\Delta^2 = [(1,96)^2 \cdot (32)^2]/(5)^2 = 157,4$ лица

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

$n = (1,96)^2 \cdot (32)^2 / [(5)^2 + (1,96)^2 \cdot (32)^2 / 3000] = 149,5$ 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 (p);
- * допустимата максимална грешка (в абсолютни или относителни единици);
- * стандартното отклонение на променливата в популацията;
- * възприетата гаранционна вероятност на изводите.

От формулите $\Delta = t \cdot s_{\bar{x}}$ и $s_{\bar{x}} = \sqrt{p \cdot q/n}$ намираме, че $\Delta = t \cdot \sqrt{p \cdot q/n}$

Отгук $n = t^2 \cdot p \cdot q/\Delta^2$

Пример: p = 26% (0,26); Δ = 3% (0,03); P = 95%; t = 1.96

Тогава: $n = t^2 \cdot p \cdot q/\Delta^2 = (1,96)^2 \cdot 0,26 \cdot 0,74/(0,03)^2 = 821,2$ 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 p е неизвестна, най-добре да се приеме $p=50\%$. Тогава и q ще бъде 50% . Произведението $p \cdot q$ 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:
- A. стандартно отклонение
 - B. стандартна грешка
 - B. параметър
8. Ако подборът на извадката е случаен непреднамерен, грешката на извадката е нула.
- A. вярно
 - B. невярно
9. Грешката на средната аритметична величина в извадката:
- A. се дължи на неподходящи техники на подбор на извадката
 - B. намалява с увеличаване на размера на извадката
 - B. 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?
- A. репрезентативната грешка намалява два пъти
 - B. репрезентативната грешка нараства
 - B. репрезентативната грешка намалява четири пъти
11. Какво трябва да направим, ако желаем да намалим три пъти репрезентативната грешка?
- A. да намалим девет пъти броя на наблюдаваните случаи
 - B. да увеличим броя на наблюдаваните случаи девет пъти
 - B. 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 .
- A. вярно
 - B. невярно
13. Колкото е по-голяма репрезентативната грешка на извадката, толкова интервалът на доверителност е по-малък.
- A. вярно
 - B. невярно
14. За големи по размер извадки, грешката на средната аритметична е по-голяма от стандартното отклонение.
- A. вярно
 - B. 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 $72 \div 79$. Следователно:
- А. с вероятност 0.05 може да се твърди, че μ попада в интервала $72 \div 79$.
 - Б. с вероятност 0.95 може да се твърди, че интервалът $72 \div 79$ съдържа \bar{x}
 - В. с вероятност 0.95 може да се твърди, че интервалът $72 \div 79$ 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 $\bar{x} = 30$ и $s^2 = 9$ ще бъдат:
- А. 30 ± 9.00 Б. 30 ± 0.79 В. 30 ± 1.03 Г. 30 ± 0.47
26. Подбрана е извадка от 16 наблюдения от популация с нормално разпределение. Долната граница на 98% доверителен интервал при $\bar{x} = 3.2$ 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. 2.064 Б. 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 $\bar{x} = 11.2$ кг и $s = 9$ кг. Границите на 95% доверителен интервал за средната в популацията, от която е извлечена извадката (допускайки нормално разпределение и непреднамерен подбор) ще бъдат:
- А. $11.2 \pm 1.96 \times 3$ Б. $11.2 \pm 1.86 \times 3$ В. $11.2 \pm 2.306 \times 3$
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. $0.56 \pm 1.96 \times \sqrt{(0.44) \cdot (0.56)}$ Б. $0.44 \pm 1.96 \times \sqrt{(0.56) \cdot (0.44) / 400}$
В. $0.56 \pm 1.96 \times \sqrt{(0.44) \cdot (0.56) / 400}$ Г. $0.56 \pm 1.64 \times \sqrt{(0.44) \cdot (0.56) / 400}$

30. При проучване на мнението на 100 студенти по медицина, 50 от тях поддържат въвеждането на проблемно базирано обучение (ПБО). Ако p 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 ± 0.13 Г. 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 $P=95\%$ делът този дял е:

А. $0.768 \div 0.832$ Б. $0.761 \div 0.839$ В. $0.775 \div 0.825$ Г. $0.790 \div 0.810$

32. За случайна извадка от 16 болници е установено, че средната цена за 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. $3100 \div 3350$ Б. $2850 \div 3330$ В. $2632 \div 3368$ Г. $2890 \div 3230$

33. При положение, че 99% доверителен интервал за μ е $42 \div 58$, 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. $98.9 \div 107.1$ Б. $97.6 \div 108.3$ В. $98.0 \div 108.0$ Г. $94.5 \div 111.5$

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. $15.6 \div 23.2$ Б. $14.8 \div 24.4$ В. $12.1 \div 27.9$ Г. $10.6 \div 21.8$

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)?

- A. 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?

- A. 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 $\bar{x} = 115$ 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?

- A. 10 Б. 20 В. 2

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

- A. 2.492 Б. 2.797 В. 2.787

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

- A. 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