*Глава 7*

# ОПИСАНИЕ НА КОЛИЧЕСТВЕНИ ПРОМЕНЛИВИ. ИЗМЕРВАНЕ НА ЦЕНТРАЛНА ТЕНДЕНЦИЯ

## 1. Две основни свойства на количествените променливи

Принципният проблем при работа с живи организми е присъщата им ***вариабилност.*** Например, ако измерим диастолното налягане на 56 мъже силни пушачи на възраст 50-59 г., макар групата да е привидно еднородна по признаците пол, възраст и навици за пушене, то почти със сигурност можем да очакваме 56 различни стойности на диастолното налягане, простиращи се най-вероятно в интервала от 60 до 120 мм Hg. Подобни примери могат да се посочат и за много други количествени променливи – антропометрични признаци за физическо развитие (ръст, тегло, гръдна обиколка и др.), функционални признаци за дейността на организма (жизнена вместимост на белите дробове, пулс, ниво на хемоглобин, еритроцити, левкоцити) и т. н.

В същото време, въпреки индивидуалните различия, при разглеждане на стойностите на количествените променливи в определена съвкупност се установява ***"стремеж" към определено средно ниво*****,** около което се разполагат индивидуалните стойности на променливата, т.е. установява се определена ***централна тенденция.***

***Средното ниво на количествените променливи*** се явява едно от най-характерните групови свойства на статистическата съвкупност, за изучаването на което се използват няколко описателни статистически критерии.

Преди да се пристъпи към изчисляване на описателните характеристики на количествените променливи, изходните данни

трябва да бъдат представени в подходящ вид – честотно разпределение (групиран или интервален вариационен ред) или графическо изображение (хистограма или полигон). Това позволява на изследователя да се ориентира и да определи вида на разпределението на количествените променливи и да подбере подходящи статистически методи за обобщаване на тези променливи.

***Честотното разпределение (вариационният ред) представлява ред от числени стойности, характеризиращи дадена количествена променлива при всеки отделен случай, подредени най-често във възходящ ред. Всеки вариационен ред има следните основни елементи:***

- ***стойности на променливата,*** означавана с ***x*;**
- ***честота*** – означава се с ***f*** и показва колко пъти се повтаря дадена стойност на променливата в конкретния вариационен ред;
- ***лимит (размах*)** – разликата между най-ниската и най-високата стойност на променливата величина.

## 2. Измерване на централна тенденция

Стремежът на количествените променливи към ***централна тенденция (средно ниво)*** е основно групово свойство на всяка статистическа съвкупност. Тази тенденция се наблюдава поради това, че всяко явление се развива под влиянието на ***определящи, закономерни фактори и причини***, които са налице при всички индивиди от дадена популация.

За описване на централната тенденция се използват ***два основни вида средни величини***:

- *алгебрични средни величини – средна аритметична, средна геометрична, претеглена средна и др.);*
- *позиционни средни величини – медиана, мода, квартили, персентили.*





При изчисляване на алгебричните средни величини се включват всички стойности на изучаваната променлива в емпиричното разпределение, докато позиционните средни зависят само от броя на стойностите и позициите (местата), които те заемат в дадено честотно разпределение при подреждането на измерените стойности във възходящ или низходящ ред.

### 2.1. Средна аритметична величина

Това е най-често използваната описателна числова характеристика на количествените променливи величини. Означава се с $\bar{x}$ **за извадка** и с μ **за популация**. Тъй като в медицината проучванията се базират най-вече на извадки, от които се правят изводи за популацията, в по-нататъшното изложение използваме $\bar{x}$ като символ на средната аритметична.

Подходите за изчисляване на средната аритметична зависят от ***начина на представяне на изходните данни и броя на наблюдаваните случаи.***

***При малък брой случаи*** (под 30) стойностите на количествената променлива се представят като негрупиран ред от числа, т.е. ***непретеглен вариационен ред***. В такъв случай ***средната аритметична величина*** ($\bar{x}$) е равна на сумата от измерените стойности ***(Σx)*** върху броя на случаите ***(n)***:

$$(7.1)\quad \bar{x} = \frac{\sum x}{n}$$

**Пример:** Стойностите на възрастта при 10 първораждащи майки са: 18, 21, 23, 23, 25, 27, 27, 28, 30, 33.

Средната възраст в тази извадка е:

$$(7.2)\quad \bar{x} = \frac{18+23+23+\ldots 33}{10} = \frac{255}{10} = 25{,}5$$

***При степенен вариационен ред*** $\bar{x}$ се изчислява по формулата:

$$(7.3)\quad \bar{x} = \frac{\sum x.f}{\sum f},$$

където:

$\Sigma$ ***x.f*** е сумата от произведенията на стойностите на променливата ***x*** $(x_1, x_2, x_3, ..... x_n)$ и честотата ***f*** за всяка стойност, а $\Sigma f$ е брой наблюдаваните случаи.

При изчисляването на средната аритметична величина в степенен вариационен ред се преминава през следния алгоритъм:

- всяка стойност на променливата се умножава по нейната честота и произведението ***x.f*** се записва на съответния ред;
- получените произведения от стойностите и тяхната честота се сумират и се записват като $\Sigma x.f$;
- сумата $\Sigma$ ***x.f*** се разделя на броя на случаите ($\Sigma f = n$) и се получава $\bar{x}$.

Този метод е твърде бавен при дълъг степенен вариационен ред и голям брой наблюдавани случаи. Тогава се препоръчва преобразуване на степенния ред в интервален, за предпочитане с еднаква ширина на интервалите.

***При интервален вариационен ред*** с еднаква ширина на интервалите средната аритметична величина се изчислява по формулата:

(7.4) $\bar{x} = \frac{\sum c.f}{\sum f}$, където

***c*** е средата на всеки интервал (полусума от долната и горната граница на интервала;

$\Sigma c.f$ – сумата от произведенията на средите на интервалите и честотата във всеки интервал);

$\Sigma f$ – общият брой на случаите.

Изчисляването на средната аритметична при интервален вариационен ред преминава през следния алгоритъм:

- определя се ширината на интервалите;
- данните се прегрупират в равностоящи интервали и се сумират честотите в интервалите;
- определя се средата на всеки интервал;





- честотата за всеки интервал се умножава по средата му;
- произведението ***c.f*** се записва в съответния ред;
- сумират се произведенията ***c.f*** и $\Sigma c.f$ се записва в последния ред;
- сумата $\Sigma c.f$ се разделя на броя на случаите $\Sigma f$ и се получава $\bar{x}$.

Посочените три начина са изключително полезни за изчисляване на средната аритметична величина ръчно или с помощта на обикновен калкулатор. В съвременните условия обаче изчислителната процедура е максимално облекчена от персоналните компютри и съответни програмни продукти. Достатъчно е да бъдат въведени правилно и точно съответните първични данни, след което могат да се приложат разнообразни статистически методи, включващи и изчисляване на различни видове средни величини.

***Основни характеристики и свойства на средната аритметична величина***

1. Средната аритметична е най-често използваната мярка за централна тенденция и нейното ***предимство*** е в това, че ***посредством едно число тя замества множество индивидуални различаващи се стойности и*** описва ***типичното ниво*** на количествената променлива в изучаваната съвкупност.

2. Трябва да се има предвид, обаче, че ***средната аритметична може да характеризира добре типичното ниво само когато емпиричното разпределение е нормално*** (симетрично, Гаус-Лапласово) ***или близко до нормалното***. При асиметрични разпределения тя е абстрактна стойност без реално значение и не може да бъде мярка за централната тенденция.

3. ***Съществен недостатък на средната аритметична*** е това, че тя може да бъде ***повлияна от наличие на рязко отличаващи се стойности.***

Например, 10 HIV-положителни индивиди при интервюиране са посочили следния брой сексуални контакти за последните 6 месеца: 2, 4, 4, 6, 7, 8, 10, 12, 15, 93

Средната аритметична за този елементарен вариационен ред е равна на **16.1** ***(Σ x = 161; n=10***). Както се вижда, тя е по-висока от 9 от посочените стойности сред всички 10 изследвани лица и се различава твърде силно от 10-та стойност. Следователно, средната аритметична по никакъв начин не може да се приеме за типично ниво в този пример.

Разработени са подходи, които позволяват да се отстраняват рязко отличаващите се стойности и чрез повторни изчисления да се определи нова средна аритметична, която да е по-типична за съответното емпирично разпределение.

Такъв подход ни предоставя ***критерият на Шовене***, който се изчислява на базата на средната $\bar{o}$ и стандартното отклонение ***s*** по следната формула:

(7.5) $U = \frac{x_i - \bar{x}}{s}$, където

$x_i$ е стойността. за която трябва да се даде преценка;

***s*** – стандартното отклонение

Изчисленият критерий на Шовене се сравнява с табличен коефициент $u_t$ и ако $u \geq u_t$ рязко отклоняващата се стойност $x_i$ се отстранява като необичайна. Освен това, се прави и логичен анализ на условията, при които е протекло наблюдението и се търсят възможни източници на грешка.

4. ***Средната аритметична не винаги е реална стойност*** и това затруднява нейното възприемане и интерпретация, когато променливата величина има дискретен характер. Например, средният брой деца в група изследвани семейства е 2.1 – не е ясно дали болшинството семейства имат по 2 деца и само малък брой са имали по 3 деца или пък някои семейства са имали 4-5 деца, а други – само по 1.

5. ***Ако към всяка стойност в честотното разпределение се прибави или извади едно и също число, то средната аритметична нараства или намалява със същото число.***





6. ***Сумата от отклоненията на стойностите на променливата от средната аритметична винаги е равна на нула.*** Това е така, защото средната аритметична е фактически математическият център на данните.

7. ***Сумата от квадратите на отклоненията около средната аритметична е по-малка от сумата от квадратите около която и да е друга стойност в честотното разпределение.*** Това свойство лежи в основата на изчисляването на "най-малките квадрати", което се използва в регресионния анализ и при анализ на динамични промени.

## 2.2. Медиана

***Медианата (Ме)*** представлява стойността, която дели вариационния ред на две равни части, т. е. това е срединната точка в поредица от ранжирани стойности. Това означава, че 50% от наблюдаваните случаи се разполагат под тази стойност и 50% – над нея.

***При нечетен брой*** случаи медианата е равна на стойността, разположена точно в средата на реда, т.е. за да се определи правилно медианата първо трябва да се подредят всички стойности на променливата във възходящ ред.

***При четен брой*** случаи медианата е равна на полусумата от двете стойности, разположени в средата на вариационния ред.

В посочения пример с 10-те първораждащи медианата е 26 години – тя се намира между 5-та и 6-та стойности на вариационния ред (между 25 и 27 години), т. е. тук средната аритметична и медианата почти съвпадат, защото в този вариационен ред няма рязко отклоняващи се стойности на количествената променлива.

***Медианата има следните основни характеристики:***

***1. Медианата е по-реална стойност*** от средната аритметична и се измерва чрез цяло число или в най-лошия случай като половинно число. В примера с HIV+ лица медианата е 7.5 (между 5-та и 6-та стойности на вариационния ред) и тя доста по-добре характеризира

типичното сексуално поведение на изследваните лица в сравнение с изчислената по-горе средна аритметична величина.

2. ***Медианата е по-устойчива на влиянието на рязко отклоняващи се стойности,*** тъй като тя не зависи от екстремалните стойности, а само от стойностите, разположени в средата на реда. Например, стойността на медианата няма да се промени, ако последното HIV+ лице посочи 120 или 50 контакта.

3. Единственият ***недостатък на медианата*** е в това, че тя не включва всички индивидуални стойности на променливата, а отразява само една стойност при нечетен брой случаи (напр. 31-та стойност при 61 случая) или две стойности при четен брой (напр. 30-та и 31-та стойности при 60 случая).

4. Медианата е предпочитана характеристика на типичното ниво, когато:

- крайните стойности на количествената променлива са доста отдалечени от останалите стойности;
- има съмнение в някои от стойностите;
- не може да се установи точният вид на разпределението или пък е налице силно изразено асиметрично разпределение;
- обемът на изучаваната съвкупност е малък.

### 2.3. Мода

***Модата (Мо)*** е третата мярка на централната тенденция. ***Модата е тази стойност от вариационния ред, която се среща с най-голяма честота*** (както е с модата в обществения живот). В горния пример, възрастта 23 и 27 години се среща по два пъти, т. е. в извадката има две модални стойности.

***Модата притежава следните основни характеристики:***

***1. Тя е най-лесно оценимата средна величина.***

***2.*** При голям брой случаи и ***нормално разпределение модата е една***. Но, както е в горния пример, в един вариационен ред може





да има две или даже повече от две моди – ***двумодално или полимодално разпределение.*** Последното може да е знак за неоднородност на изучаваната съвкупност.

***3.Модата е единствената мярка, приложима и по отношение на категорийни данни.***

***4. Модата*** се използва по-рядко от средната аритметична и медианата, но тя ***има реален смисъл*** и в медицината това е твърде важно. Например, много по-важно е да определим коя е най-рисковата група лица за дадено заболяване, т.е. да видим коя е модата във възрастовото разпределение на заболелите лица, отколкото да изчисляваме средната възраст на заболелите.

### 2.4. Сравнение на средната аритметична, медианата и модата

***Изборът на метод за описване на централна тенденция*** зависи от скалата на измерване на променливата величина.

Ако данните са ***номинални***, може да се използва само ***модата***.

При ***ординални*** данни често се използват ***модата и медианата***.

Когато данните представляват ***интервална или пропорционална скала***, може да се използва ***всяка една от трите мерки за централна тенденция***.

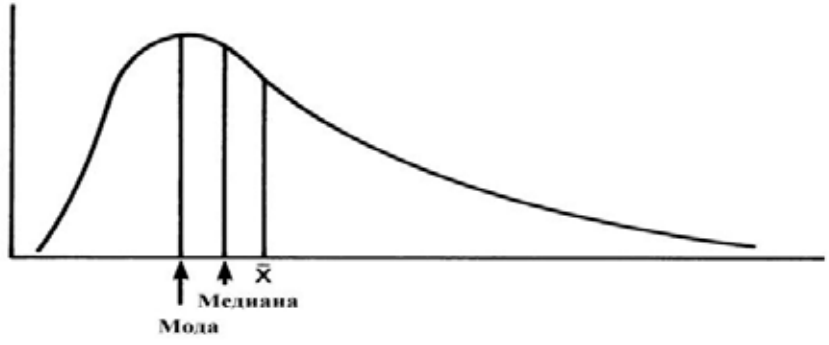
***Средната аритметична величина, медианата и модата се намират в различно съотношение при отделните видове разпределение:***

*1. При идеално нормално (симетрично, Гаус–Лапласово) разпределение средната аритметична, медианата и модата имат еднакви стойности.*

*2. При дясно изтеглено (положително) разпределение* ***(фиг. 7.1)*** *средната аритметична има най-висока стойност, следвана от медианата и модата.*

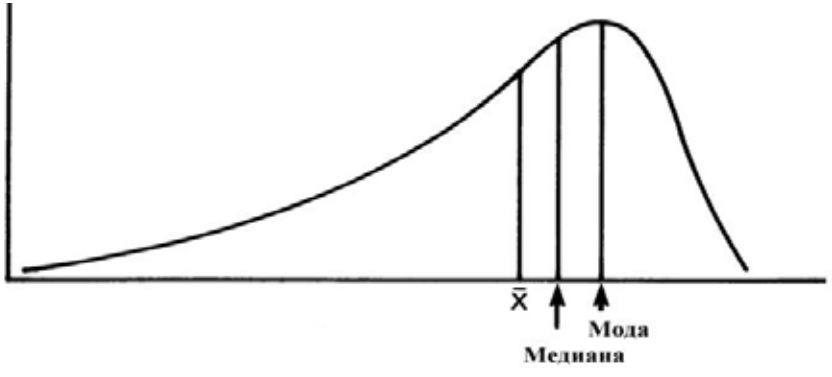
*3. При ляво изтеглено (отрицателно) разпределение* ***(фиг. 7.2)*** *най-висока стойност има модата, следвана от медианата и средната аритметична.*

$Mo < Me < \bar{x}$




**Фиг. 7.1. Съотношение между $\bar{x}$, *Mo* и *Me* при дясно изтеглено разпределение**

$Mo > Me > \bar{x}$




**Фиг. 7.2. Съотношение между $\bar{x}$, *Mo* и *Me* при дясно изтеглено разпределение**

## 5. Други позиционни средни величини – персентили и квартили

В случаите, когато разпределението на изучаваните променливи величини е асиметрично, полезна информация ни предоставят такива измерители на местоположението като ***персентили и квартили.***





**Персентилите** представляват стойности в един вариационен ред, които **делят разпределението на 100 равни части.**

Следователно, има 99 персентила, които се означават с $P_1$, $P_2$, ... $P_{25}$,...$P_{50}$,...., $P_{75}$, ....$P_{99}$.

Персентилите дават информация за относителното място на даден резултат в определен масив данни.