

Глава 5

ОРГАНИЗАЦИЯ И ПРЕДСТАВЯНЕ НА ДАННИТЕ ОТ НАУЧНИ ПРОУЧВАНИЯ

Г. Грънчарова

В тази глава:

5.1. Въведение

5.2. Таблично представяне на данните

5.3. Честотни разпределения

5.4. Графично представяне на таблични данни и честотни разпределения

5.5. Заключение

5.6. Въпроси за самоподготовка

5.1. Въведение

Преди да се пристъпи към обсъждане на резултатите от дадено проучване, необработените данни трябва да бъдат организирани и представени по ясен и интелигентен начин, обезпечаващ стегнат и компактен вид на масива от данни и на неговите основни характеристики. Редуцирането на данните до управляема и лесна за разбиране и осмисляне форма представлява една от най-важните цели на статистиката.

Съществуват различни подходи (ръчни или чрез използване на компютри) за обединяване в едно цяло на данните от различните рутинни или специални проучвания. Начинът за организиране и представяне на данните зависи от техния вид (номинални или ординални, дискретни или непрекъснати) и използваната измерителна скала.

Особено важни са правилната ***групировка и обобщаване*** на данните.

Групировката има за цел да разпредели единиците на изучаваното масово явление в еднородни статистически групи и спомага да се изучат типичните закономерности. Уточняването на разновидностите за извършване на групировката на данните трябва да стане още на етапа на планиране на проучването и зависи от вида на изучаваните променливи и задачите на изследването. Разпределянето на единиците на изучаваното явление по разновид-

ности на променливите и брой на случаите в отделните разновидности се извършва в етапа на обработка на данните.

Обобщаването на данните представлява многоетапна операция, свързана с изчисляване на групови и крайни резултати, като програмата за това се разработва заедно с плана и програмата на проучването. Трябва да се знае какви резултати се очакват на изхода на проучването, за да се съберат необходимите за това данни. Само така може да се обезпечи добър баланс между целта и задачите, програмата на проучването, събирането на данните, методите на тяхната обработка и крайните резултати от проучването.

Обобщаването на данните може да се извърши в:

- * **статистически таблици**
- * **честотни разпределения**

5.2. Таблично представяне на данните

По същество **таблицата** представлява разнообразна по съдържание мрежа от пресичащи се редове и колони, които са озаглавени и запълнени с числа. Изготвянето на всяка таблица трябва да отговаря на редица **общи изисквания**.

Всяка една правилно изработена таблица следва да съдържа следните формални елементи:

- **Номер на таблицата** – изисква се когато данните са представени в множество таблици. Най-често номерът на таблицата е непосредствено пред заглавието или се разполага на горния ред вдясно или вляво. При анализ на данните авторите цитират номера на таблицата в текста, откъдето се тълкуват данните и се правят съответни изводи.
- **Заглавие на таблицата** - то трябва да отразява в ясна и сбита форма съдържанието на таблицата, т.е. да отразява обекта на изследване, разглежданите характеристики, времето и мястото, измерителните единици (брой, %).
- **Челна колона** – това е първата колона на таблицата, в която обикновено се разполага т. нар. **статистически подлог**, т. е. това, което ще бъде описано в таблицата по съответен начин (например, видове заболявания, основни причини за умирация и др.).

- **Челен ред** – това е най-горният ред на таблицата, в който обикновено се отразява т.нар. *статистическо сказуемо*, т.е. признаците, които характеризират статистическия подлог (например, пол и възраст на заболялите или умрелите лица).
- **Глава на таблицата** – това е най-горната лява клетка, в която се пресичат челния ред и челната колона.
- **Клетки на таблицата** – те се получават от пресичането на колоните и редовете. Техният брой зависи от разновидностите или интервалите, чрез които се описват изучаваните признаци.
- **Сумарен ред** – това е най-долният ред на таблицата, в който се отразяват сумите за всяка колона.
- **Сумарна колона** – последната дясна колона на таблицата, в която се отразяват сумите за всеки ред.
- **Контролна клетка на таблицата** – най-долната клетка вдясно, в която се пресичат сумарният ред и сумарната колона. Тя отразява общия брой на изследваните лица, проби и др. и служи за контрол относно правилното попълване на таблицата.

Табл.5.1 представя *примерен макет* на четирикратна таблица с нейните формални елементи.

Табл.5.1 Примерно заглавие “Разпределение на изследваните лица според резултатите от теста и по пол (в брой и %)

		Глава ↓		Челен ред →			
		Пол→					
		Резултат ↓	Мъже	Жени		Всичко	
← Челна колона		Положителен тест	Клетка 1	Клетка 2		Всичко положителни	← Сумарна колона
		Отрицателен тест	Клетка 3	Клетка 4		Всичко отрицателни	
		Общо	Общо мъже	Общо жени		Всичко изследвани	
		Сумарен ред →		↑ Контролна клетка			

В клетките на таблицата, освен числа, могат да се поставят различни *условни знаци*:

- (.) – липсват данни и не е възможно да се получат въобще
- (...) – не са търсени такива данни
- (x) – в клетката не е логично да има число поради естеството на данните
- (-) – липсват съответни абсолютни числа, въпреки че са търсени
- (?) – посочената цифра е под съмнение
- (*) – числото в клетката подлежи на уточняване или се отнася за различен период спрямо посочения за останалите клетки
- (0) – съответният показател е под $\frac{1}{2}$ от възприетата мерна единица

Видове таблици

Проста (едномерна) таблица – резултатите са групирани само по един признак (напр. разпределение на изследваните лица по професия) или пък представя честотите на една променлива величина (напр. разпределение на група лица по ниво на систолното налягане).

Сложни таблици:

- **Групова таблица** – съдържа повече от един признак, но всеки признак се разглежда поотделно, без да се съчетават помежду си (например, заболяванията могат да се разглеждат отделно по пол и по възраст).
- **Комбинационна таблица (крос-табулация)** – тук е налице съчетание на разглежданите променливи величини.
- В зависимост **от броя на крос-табулираните променливи сложните таблици биват:**
 - **двумерни таблици**
 - **многомерни таблици**
- В зависимост **от броя на разновидностите на комбинираните променливи величини** сложните таблици биват:
 - **четирикратна (четириклетъчна) или таблица 2x2** - когато се съчетават биноминални променливи величини, всяка от които има само по две разновидности (наричани още **алтернативни признаци**). Тогава са възможни само 4 съчетания на разновидностите на променливите.
 - **многократна (многоклетъчна)** – когато поне един от изучаваните признаци има повече от 2 разновидности и възможните съчетания между разновидностите на признаците са повече от четири.

Всеки вид таблица има своето значение. Простите таблици дават начална представа за основните характеристики на изучаваното явление и подсказват в каква насока би могъл да се провежда последващия анализ.

Най-съвършена е комбинационната таблица, която дава представа за взаимоотношенията между признаците на изучаваното явление и за неговата обусловеност от различни фактори. Много важно, обаче, е да се избягва прекомерното усложняване и опити за съчетаване на повече от 3-4 признака, тъй като това утежнява анализа. По-добре е да се съставят няколко комбинационни таблици с умело и логично подбрани признаци, което ще позволи да се разкрият по-бързо съществени закономерности.

Начинът на обобщаване на данните е важно изискване за приложението на различни статистически методи за обработка на данните (напр., за изчисляване на коефициенти на корелация, на критерия χ^2 и др. в зависимост от това дали таблицата е четирикратна или многократна).

5.3. Честотни разпределения

Честотните разпределения показват начина, по който са разпределени резултатите в дадена извадка или популация.

Пример: На изпит 50 студенти отговарят на тест, съдържащ 50 въпроси от типа Multiple Choice (с множество отговори, от които само един верен) като за всеки верен отговор са присъждани по 2 точки. Резултатите варират от 54 до 100 точки и са оценявани по скала:

под 60 точки - слаб

60-64 точки – среден 3.00

66-70 точки – добър 3.50

72-76 точки – добър 4.00

78-82 точки – много добър 4.50

84-88 точки – много добър 5.00

90-94 точки – отличен 5.50

96-100 – отличен 6.00

Ако срещу поредния номер на всеки изпитан студент посочим неговия резултат (x), ще получим ред от 50 числа (**табл.5.2**), които варират в широки граници и не са групирани по някакъв системен начин. Такъв ред се нарича **прост или непретеглен вариационен ред**. От него едва ли бихме могли да направим извод доколко добре са се справили студентите с теста.

Табл. 5.2. Резултати от тестово изпитване на 50 студента (брой точки)

Лице N	Резултат x	Лице N	Резултат x	Лице N	Резултат x	Лице N	Резултат x	Лице N	Резултат x
1	80	11	64	21	66	31	78	41	86
2	64	12	84	22	98	32	88	42	70
3	82	13	68	23	76	33	84	43	90
4	100	14	74	24	98	34	60	44	94
5	70	15	90	25	94	35	92	45	74
6	86	16	82	26	58	36	82	46	96
7	90	17	92	27	92	37	96	47	62
8	72	18	56	28	78	38	72	48	96
9	88	19	98	29	86	39	94	49	100
10	54	20	76	30	68	40	80	50	88

Сравнително лесен и полезен начин за внасяне на по-голям ред и яснота в представянето на количествени данни, измерени при голям брой случаи (над 20-30) е подреждането на стойностите на измерванията (x) във възходящ или низходящ ред, като срещу всяка стойност се посочи колко лица имат такъв резултат, т.е. каква е **честотата** (f – *frequency*) на поява на конкретна стойност на променливата в изучаваната съвкупност (*табл.5.3*). Този прием за групиране на данните се нарича **честотно разпределение или претеглен вариационен ред** (тъй като за всяка стойност на променливата е посочено нейното тегло, т.е. колко пъти се среща).

Табл. 5.3. Честотно разпределение на резултати от тест при 50 студента

x	f	x	f	x	f
54	1	70	2	86	3
56	1	72	2	88	3
58	1	74	2	90	3
60	1	76	2	92	3
62	1	78	2	94	3
64	2	80	2	96	3
66	1	82	3	98	3
68	2	84	2	100	2

След групирането на данните се вижда, че само 3 студенти не са покрили минималните изисквания, а 16 студента са се справили отлично. Вместо в 50 реда данните са групирани в 24 реда, което е доста по-лесно за анализ.

Когато срещу всяка измерена стойност на променливата се посочва честотата ѝ, разпределението е **степенно (степенен вариационен ред)**.

По-нататъшната организация на данните може да доведе до обединяване в интервали и превръщане на разпределението в **интервално (интервален вариационен ред)**. При съставянето на интервално честотно разпределение трябва да се отчитат следните моменти:

1. **Оптимален брой интервали**. Прието е той да не надвишава 9, за да не се затруднява анализът на данните, но не бива да е твърде малък, за да не се прикрият някои важни тенденции в данните.

2. Предпочитат се **интервали с еднаква ширина**, която се избира според изучаваната количествена променлива и опита на изследователя. Когато има затруднения се препоръчва прилагане на **формулата на Стърджес**:

$e = (x_{\max} - x_{\min}) : (1 + 3.322 \cdot \lg n)$, където

e – ширина на интервала

x_{\max} - максимална стойност на променливата

x_{\min} – минимална стойност на променливата

3.322 – константна величина

n – брой на наблюдаваните случаи в извадката

Ако броят на групите е предварително фиксиран, то ширината на интервала се определя по формулата $e = (x_{\max} - x_{\min}) : k$, където k е броят на групите. Ако се получи дробно число, то се закръглява на цяло число.

3. За предпочитане е **ширината на интервала да е нечетно число**, което улеснява изчисляването на средните аритметични и графичното представяне на честотните разпределения, тъй като интервалът се представя чрез неговата среда (при нечетни числа тя е цяло число). Понякога това е невъзможно.

4. **Границите на интервалите** трябва да бъдат **взаимно изключващи се и изчерпателни**, което означава, че всеки резултат може да попадне само в един интервал. Напр., общоприетите интервали за възрастта при ширина 5 години са: 0-4, 5-9, 10-14, 15-19, 20-24 и т.н.; при ширина 10 години – 0-9, 10-19, 20-29, 30-39 и т.н.

5. **Следва да се избягват интервали с отворено начало или отворен край**, но понякога данните налагат това. Напр., класификацията на СЗО за възрастта е с отворено начало и край: млади - до 44 г.; средна възраст - 45-59 г.; възрастни – 60-74 г.; стари хора – 75-89 г.; 90 г. и повече – дълголетници.

Честотните разпределения могат да бъдат:

- * **абсолютни**
- * **относителни**
- * **кумулятивни**

Абсолютното честотно разпределение е представено с действителния брой случаи, които имат конкретна стойност на променливата и попадат в конкретен интервал.

При **относителното разпределение** действителната честота се превръща в относителна (общия брой случаи се приема за 100 или за 1).

Кумулативното честотно разпределение се получава чрез прибавяне на относителната честота за даден интервал към тази от предходните интервали. Това позволява да направим заключение какъв % от всички случаи попадат до съответния интервал.

Трите посочени вида честотно разпределение са представени нагледно в табл. 5.4.

Табл. 5.4. Абсолютно, относително и кумулативно разпределение на резултатите от тест при 50 студенти

<i>x</i>	Абсол. <i>f</i>	Относит. <i>f</i> (в %)	Кумулат. (<i>f</i> в %)	<i>Интервал</i>	Абсол. <i>f</i>	Относит. <i>f</i> (в %)	Кумулат. <i>f</i> (в %)
54	1	2	-	54-58	3	6	-
56	1	2	4	60-64	4	8	14
58	1	2	6	66-70	5	10	24
60	1	2	8	72-76	6	12	36
62	1	2	10	78-82	7	14	50
64	2	4	14	84-88	8	16	66
66	1	2	16	90-94	9	18	84
68	2	4	20	96-100	8	16	100
70	2	4	24				
72	2	4	28				
74	2	4	32				
76	2	4	36				
78	2	4	40	<i>Интервал</i>	Абсол. <i>f</i>	Относ. <i>f</i>	Кумул. <i>f</i>
80	2	4	44	54-60	3	6	-
82	3	6	50	60-64	4	8	14
84	2	4	54	66-76	11	22	36
86	3	6	60	78-88	15	30	66
88	3	6	66	90-100	17	34	100
90	3	6	72				
92	3	6	78				
94	3	6	84				
96	3	6	90				
98	3	6	96				
100	2	4	100				

5.4. Графично представяне на таблични данни и честотни разпределения

Графичното представяне на данните е неделима част от изследователския процес и с него се постигат *две основни цели*:

- *по-лесно възприемане на резултатите*, дори и за неспециалисти;
- *по-добро очертаване на закономерностите* на изучаваното явление.

Ако използваните средства за онагледяване не отговарят на тези цели, те стават безсмислени, дори може да затруднят интерпретацията на резултатите. Ако се онагледяват данни за няколко групи, между които няма съществено различие, онагледяването също има по-малка стойност и бихме могли да се задоволим само с текстово описание.

Графичните изображения имат редица *предимства* пред таблиците:

1. *Нагледност* - добре оформените графически изображения се възприемат зрително по-леко.

2. *Изразителност* - статистическите данни, представени графически, се запомнят много по-бързо.

3. *Атрактивност* - вниманието на читателя или слушателя може да бъде ангажирано много по-пълно от графически представени данни.

4. *Логичност* - вътрешната логика, която е налице при причинно обвързаните явления, проличава изключително добре при тези изображения.

5. *Експресивност* - те дават богати възможности за изразяване на известна идея, за защита на определена концепция и за изява на определена теза.

Основните изисквания към графичните изображения включват:

- за всяко изучавано явление и вид данни да бъде избран *най-подходящият вид графично изображение*;
- да имат *ясно формулирани заглавия и легенди*;
- да *представят предимно основните данни и зависимости*, без излишни детайли;
- да са *добре изпълнени технически* с подбор на подходящи шриховки и означения;
- да са *достатъчно ясни* и да позволяват бързо възприемане и осмисляне на информацията без допълнителни изчисления.

Видове графични изображения:

◆ *По измерение*

- * плоскостни (едномерни)

- * пространствени (двумерни или тримерни)

◆ **По характер**

- * линейни диаграми
- * стълбови диаграми (вертикални и хоризонтални)
- * кръгово-секторни диаграми
- * цилиндрови, конусовидни, пирамидални, радиални
- * площи
- * ХУ диаграми на разсейването
- * фигурни
- * картограми и картодиаграми и т.н.

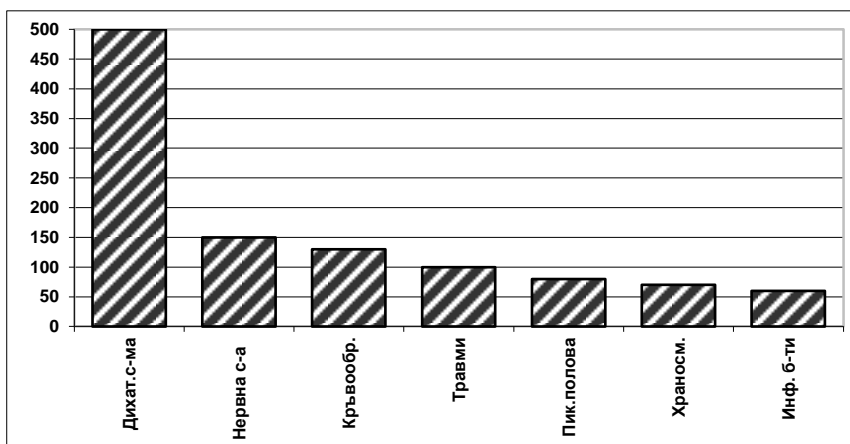
Компютърните технологии днес предоставят изключително богато разнообразие от графични изображения. При **подбора на вида им** трябва да се изхожда преди всичко от характера на данните, т.е. дали те са категорийни (номинални, биноминални, ординални) или количествени (прекъснати и непрекъснати) и от начина на тяхното групиране.

Графично представяне на качествени променливи величини

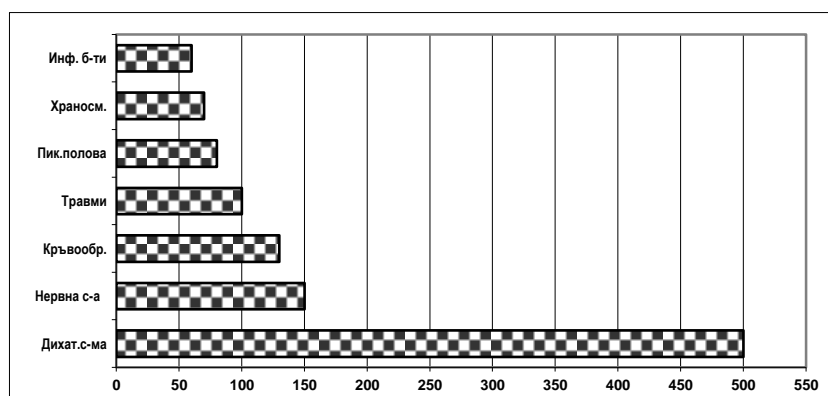
Качествените променливи величини (номинални, биноминални и ординални) се представят чрез два основни вида изображения: **стълбови диаграми (вертикални и хоризонтални) и кръгово-секторни диаграми.**

Стълбовите диаграми представят абсолютната или относителна честота за всяка разновидност на променливата (**фиг. 5.1 и 5.2**). При използването им следва да се спазват следните общоприети **правила**:

- * Оста x отразява разновидностите на категорийната променлива, а оста y – честотата на отделните разновидности на променливата.
- * Ширината на стълбовете е еднаква, а височината (дължината) им отразява честотата в абсолютни числа или в относителни величини (% , ‰ и др.). Стълбовете или лентите са отделени за разлика от хистограмата.
- * Оста y започва от нула, така че височината на стълбовете (или на лентите при хоризонталните диаграми) да е пропорционална на честотите.
- * Цветовете или щриховането трябва да подчертават закономерностите, да облекчават възприемането и да стимулират естетическо удовлетворение.

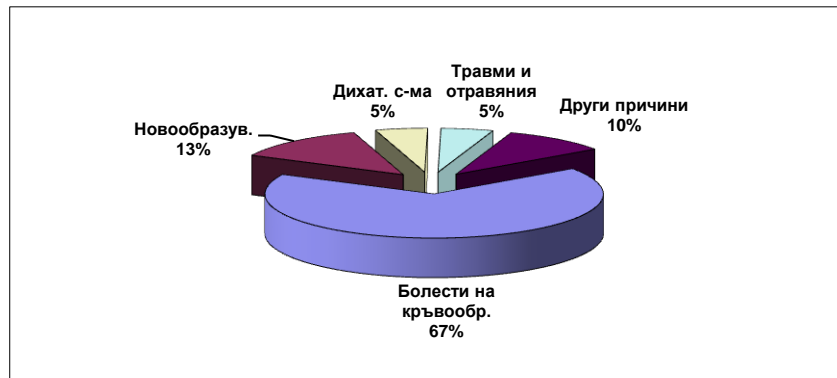


Фиг. 5.1 Регистрирани заболявания в район А.



Фиг. 5.2. Регистрирани заболявания в район А.

Кръгово-секторните диаграми представляват кръгове, разделени на сектори, пропорционално на абсолютните или относителните честоти на разновидностите на променливата величина (фиг.5.3). Всички случаи са 100% ($1\% = 3.6^0$). Понякога за 100% може да се приеме полукръг ($1\% = 1.8^0$).



Фиг. 5.3. Водещи причини за умирация в район А. – 2002 г.

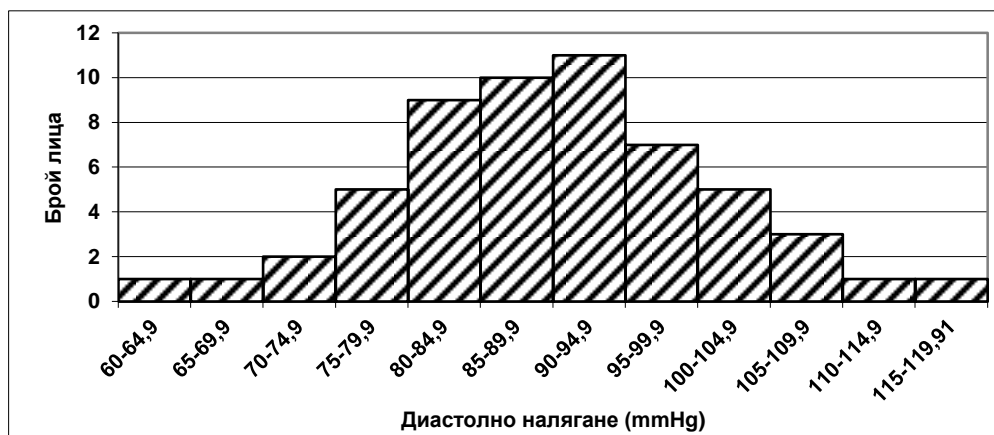
Графично представяне на количествени променливи величини

Количествените данни се измерват най-често чрез интервални или пропорционални скали и се представят чрез абсолютни, относителни и кумулативни честотни разпределения, графичният образ на които е най-често **хистограми и честотни полигони** (обикновени и кумулативни).

Хистограмата наподобява стълбова диаграма, но стълбовете са плътно прилепени, за да подчертаят непрекъснатостта на данните. Височината на стълбовете по оста y съответства на честотата в съответните класове и интервали, които са разположени на оста x . При интервални разпределения средата на всеки интервал е средна точка за съответния стълб, а ширината им съответства на границите на интервалите, които не трябва да се припокриват и трябва да обхващат цялата поредица от данни (**фиг. 5.4**).

Честотният полигон (линейна диаграма) представлява линия, която свързва срединните точки в горната част на всеки стълб на хистограмата, т.е. отделните точки на линията съответстват на абсолютните честоти за всеки интервал. Честотните полигони имат редица предимства:

- облекчават визуалното сравнение между две или повече разпределения, изобразени на една и съща диаграма;
- позволяват да се интерполират непрекъснати променливи величини;
- позволяват да се оцени честотата на стойностите между отделни точки на полигона.



Фиг. 5.4 Хистограма на непрекъсната променлива величина

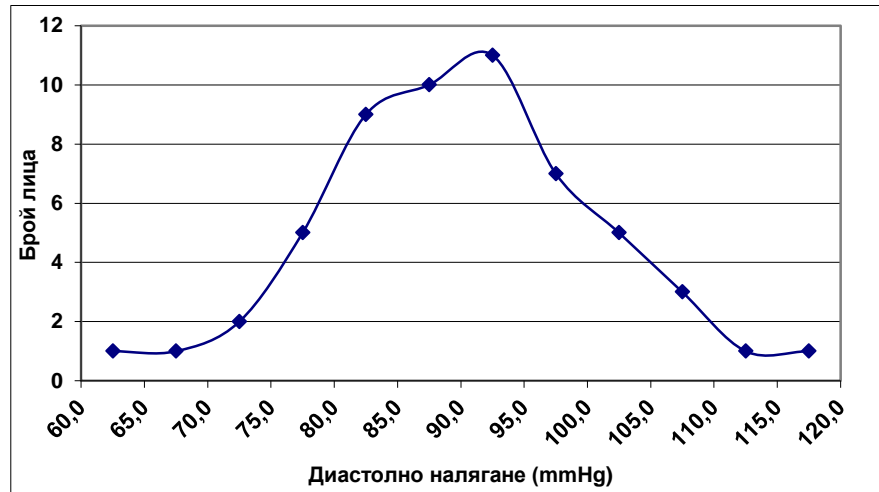
Пример: При измерване на диастолното налягане в извадка от 56 мъже на възраст 50-59 г., които са силни пушачи (пушат по 2 и повече кутии цигари на ден) са получени следните резултати (табл. 5.5.).

Табл. 5.5. Стойности на диастолно налягане при мъже пушачи на възраст 50-59 г.

Диастолно налягане (mmHg)	Абсолютна честота (f)	Относителна честота (в %)	Кумулативна честота (в %)
60-64,9	1	1.8	-
65-69,9	1	1.8	3.6
70-74,9	2	3.6	7.2
75-79,9	5	8.9	16.1
80-84,9	9	16.1	32.2
85-89,9	10	17.8	50.0
90-94,9	11	19.6	69.6
95-99,9	7	12.5	82.1
100-104,9	5	8.9	91.0
105-109,9	3	5.4	96.4
110-114,9	1	1.8	98.2
115-119,9	1	1.8	100.0
	N = 56	100.0	

Честотните полигони могат да приемат различна форма в зависимост от начина на разпределение на стойностите в проучваната съвкупност. Най-честа е симетричната камбановидна крива линия, която характеризира т.нар. *нормално (Гаус-Лапласово) разпределение* и показва, че болшинството от

измерените стойности се разполагат около средното ниво и относително малко попадат в двете “опашки” (фиг. 5.5).



Фиг. 5.5. Честотен полигон на непрекъсната променлива величина

Несиметричните разпределения са по-рядко срещани. Те се характеризират с това, че имат голям брой стойности, разположени в една от страните и изтеглена в противоположна страна “опашка” с малък брой случаи. Посоката на “опашката” дава името на съответното разпределение.

При групиране на болшинството стойности към по-високите и изтегляне на “опашката” към по-ниски стойности разпределението се нарича **отрицателно изтеглено или лява асиметрия**. Такъв вид има разпределението на резултатите от теста при 50 студента (табл. 5.4).

Обратно, когато повечето са ниски и неголям брой стойности са високи, опашката е изтеглена към високите стойности и такова разпределение се нарича **положително изтеглено или дясна асиметрия**.

Кумулативният полигон представлява линия, отразяваща кумулативно честотно разпределение. Може да се използва за оценка на честотата на поява на стойност, която е по-малка или равна на конкретна стойност на променливата (например, процент на мъжете с диастолно налягане по-малко или равно на 90 mmHg).

5.5. Заключение

Табличното представяне очертава много по-ясно характеристиките на изучаваните променливи, отколкото необработените данни. Графичното

изобразяване, от своя страна, предоставя по-добро визуално възприемане на характеристиките на променливите величини, отколкото табличното представяне. Основният недостатък е в това, че в таблиците и графиките се представят обобщени данни и се губи индивидуалността на информацията.

При **подбора на подходящи методи** за представяне на данните трябва да се отчита **конкретната ситуация и характера на данните**. В табл. 5.7 са обобщени някои основни изисквания в тази насока.

Табл. 5.7. Подходящо използване на таблично и графично представяне на данни

Таблични методи	
Характер на данните	Табличен метод
Малък масив от данни (под 20 на брой)	Ординално подреждане
Индивидуални наблюдения, много на брой, включващо само една променлива величина	Честотна таблица
Индивидуални наблюдения, включващи две или повече променливи	Крос-табулация (комбинирана таблица)
Графични методи, подходящи за таблични данни	
Таблични данни	Графичен метод
Честотна таблица, количествена променлива, един масив от данни	Хистограма или честотен полигон
Честотна таблица, количествена променлива, два масива от данни	Честотен полигон
Честотна таблица, категорийни данни	Сълбова или кръговосекторна диаграма

5.6. Въпроси за самоподготовка

- Номинални и ординални данни се представят най-добре като хистограми.
А. вярно Б. невярно
- Описателните статистически методи:
А. се опират на принципите на вероятност
Б. представляват подходи за планиране на изследванията
В. представляват подходи за обобщаване и организация на данните

3. Кои от следните данни за страните от Балканския регион могат да се изобразят като кръгово-секторна диаграма, използвайки отделен сектор за всяка страна?

- А. населението на всяка страна от Балканския регион
- Б. брутният вътрешен продукт във всяка страна от Балканския регион
- В. и двата вида посочени в А и Б данни

4. Голям масив от данни може да бъде обобщен образно чрез:

- А. размах на вариационния ред
- Б. хистограма
- В. честотна таблица

5. Ако за посочения масив от данни

x	f
15	3
17	5
19	4
23	3
29	1

се построи диаграма на абсолютните честоти и диаграма на относителните честоти, то:

- А. двете диаграми ще имат едни и същи стойности на хоризонталната ос
- Б. диаграмата на абсолютните честоти е по-неточна от тази на относителните честоти
- В. двете ще изобразяват непрекъснати данни

6. Ако за поредица от числени стойности на количествена променлива се построи диаграма на абсолютните честоти и диаграма на относителните честоти, то:

- А. диаграмата на абсолютните честоти е по-неточна от тази на относителните честоти
- Б. двете диаграми ще имат различни стойности на вертикалната ос
- В. двете ще изобразяват непрекъснати данни

7. При нормално (камбановидно) разпределение по-голямата част от резултатите са разположени в двата края на разпределението.

- А. вярно
- Б. невярно

8. Процентите (относителните дялове) на всяка стойност спрямо общия брой наблюдавани случаи представляват:

- А. хистограма
- Б. относително честотно разпределение
- В. кумулативен честотен полигон
- Г. нормално разпределение

9. Ако най-ниският резултат в едно честотно разпределение е 40, а най-високият е 189 и ние групираме резултатите в 15 интервала, то ширината на всеки интервал ще бъде:

- А. 5 Б. 10 В. 12 Г. 15

10. Ако в една хистограма на честотно разпределение ширината на интервалите се намали:

- А. броят на интервалите и броят на случаите в един интервал ще намалее
- Б. броят на интервалите ще намалее, а броят на случаите в един интервал ще се увеличи
- В. броят на интервалите ще нарасне, а броят на случаите в един интервал ще намалее
- Г. броят на интервалите и броят на случаите в един интервал ще се увеличи

11. При голяма извадка третокласници е проведен тест за проверка на способностите за четене с възможен най-висок резултат 50 точки. Данните сочат симетрично разпределение от 5 до 48 точки. Ако същият тест бъде приложен при петокласници, каква ще бъде очакваната форма на разпределение?

- А. отрицателно изтеглено (лява асиметрия)
- Б. симетрично и камбановидно
- В. симетрично, но не камбановидно
- Г. положително изтеглено (дясна асиметрия)

12. Дневната посещаемост на студентите през семестъра показва, че повечето посещават редовно занятията, с изключение на петъчните дни. При графично представяне (хоризонталната ос - брой посещения и на вертикалната ос – дните от седмицата) разпределението ще бъде:

- А. ляво изтеглено
- Б. симетрично
- В. дясно изтеглено

13. Абсолютното честотно разпределение предоставя информация за:
- А. стойността на измерването и броя на индивидите с тази стойност
 - Б. стойността на измерването и процентът на индивидите с тази стойност
 - В. стойността на измерването и процентът индивиди с тази стойност или с по-малка от нея
14. Относителното честотно разпределение предоставя информация за:
- А. стойността на измерването и броя на индивидите с тази стойност
 - Б. стойността на измерването и процентът на индивидите с тази стойност
 - В. стойността на измерването и процентът индивиди с тази стойност или с по-малка от нея
15. Кумулативното честотно разпределение предоставя информация за:
- А. стойността на измерването и броя на индивидите с тази стойност
 - Б. стойността на измерването и процентът на индивидите с тази стойност
 - В. стойността на измерването и процент индивиди с тази стойност или с по-малка от нея
16. Относителното честотно разпределение позволява сравнение между различни по размери масиви от данни.
- А. вярно Б. невярно
17. Всичко посочено по-долу за честотните разпределения е вярно, **с изключение на:**
- А. използват се за изследване на характеристиките на всеки наблюдаван случай
 - Б. полезни са за обобщаване на данните
 - В. полезни са за представяне на големи масиви от данни в по-сбита форма
 - Г. илюстрират вариабилността на данните
18. Всичко посочено за графичните изображения на данните е вярно, **с изключение на:**
- А. илюстрират варирането на данните
 - Б. илюстрират къде приблизително се разполага средната аритметична
 - В. позволяват сравняване на сходни данни
 - Г. имат абсолютно еднаква форма независимо от използваните единици върху осите

19. Честотен полигон е подходящ за графично представяне на непрекъснати променливи.
А. вярно Б. невярно
20. При отрицателно (ляво изтеглено) разпределение болшинството стойности, разпределени по оста x , са ниски и относително малко резултати имат високи стойности.
А. вярно Б. невярно
21. При нормално разпределение повечето стойности са разположени в двата екстремни края на разпределението.
А. вярно Б. невярно
22. Ако една крива е отрицателно изтеглена (лява асиметрия), разпределението на резултатите има “опашка” в посока на ниските стойности на променливата.
А. вярно Б. невярно
23. Кръгово-секторната диаграма е подходяща за графично изобразяване на данни, представени в номинална скала.
А. вярно Б. невярно
24. Пациенти изразяват удовлетворението си от проведеното лечение като отговарят на въпрос с четири варианта на отговор: 1. силно неудовлетворен; 2. неудовлетворен; 3. удовлетворен, 4. силно неудовлетворен. На каква скала са представени отговорите на пациентите и как би трябвало да бъде изобразено честотното разпределение на данните?
А. номинална скала и стълбова диаграма
Б. ординална скала и стълбова диаграма
В. интервална скала и хистограма
25. Данни, представени на интервална или пропорционална скала, се изобразяват като:
А. хистограма
Б. стълбова диаграма
В. честотен полигон
Г. верни са А и В

26. Група студенти се явяват на изпит по вътрешни болести. При условие, че само малък брой студенти са добре подготвени, разпределението на резултатите ще бъде:

- А. симетрично
- Б. отрицателно изтеглено (лява асиметрия)
- В. положително изтеглено (дясна асиметрия)

27. Кръгово-секторна диаграма е неподходяща за номинални данни.

- А. вярно
- Б. невярно

28. Група студенти се явяват на изпит по статистика. При условие, че болшинството студенти са добре подготвени, разпределението на резултатите ще бъде:

- А. симетрично
- Б. отрицателно изтеглено (лява асиметрия)
- В. положително изтеглено (дясна асиметрия)

29. Честотните разпределения на непрекъснатите данни могат да бъдат представени графично като хистограми или честотни полигоми.

- А. вярно
- Б. невярно

30. Височината на стълба (колоната) представя честотата, а не стойността на променливата. А. вярно

- Б. невярно

Отговори въпросите от глава 5:

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