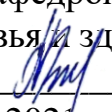


УТВЕРЖДАЮ  
Зав. кафедрой общественного  
здоровья и здравоохранения  
 В.Л.Аджиенко  
11.06.2021

**Методические указания для студентов**  
второго курса ЛЕЧЕБНОГО факультета

к проведению практического занятия  
по дисциплине «Медицинская информатика»

***Тема 10. Относительные и средние величины.  
Графическое представление данных***

Волгоград  
2021

## ВОПРОСЫ ДЛЯ ПОДГОТОВКИ К ВЫПОЛНЕНИЮ ЗАДАНИЙ

1. Для чего применяются относительные величины?
2. Назовите основные типы относительных величин и методику их вычисления?
3. Принцип вычисления интенсивных показателей?
4. Как вычисляются экстенсивные относительные величины?
5. Особенности вычисления показателей соотношения?
6. Дайте определение динамического ряда.
7. Назовите основные показатели динамического ряда.
8. Для чего применяют графические изображения данных?
9. Перечислите основные виды графических изображений.
10. Какой тип показателей применяется для построения внутрестолбиковой диаграммы?
11. Какие показатели принято изображать секторной диаграммой?
12. На основе каких показателей изготавливают столбиковые диаграммы?
13. Какие показатели изображают линейной диаграммой?
14. Дайте определение вариационного ряда
15. С какой целью в медицинских исследованиях используются средние величины?
16. Дайте определение средней величины.
17. Назовите основные элементы вариационного ряда.
18. Как вычисляется простая средняя арифметическая?
19. Как вычисляется взвешенная средняя арифметическая?
20. Назовите основные свойства средней арифметической величины.
21. Что такое среднее квадратическое отклонение и что показывает эта характеристика?
22. Как вычисляется среднее квадратическое отклонение?
23. В каких случаях возникает необходимость вычисления коэффициента вариации? Приведите пример.
24. Как вычисляется коэффициент вариации.
25. Понятие ошибки репрезентативности.
26. Как вычисляется ошибка репрезентативности?
27. Какова формула вычисления ошибки репрезентативности для относительных величин?
28. Как определяются доверительные границы средней арифметической в генеральной совокупности и с какой целью?

## ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

Для характеристики изучаемой статистической совокупности используются относительные величины, расчет которых проводится на третьем этапе статистического исследования.

Относительные величины (показатели, коэффициенты) весьма распространены и постоянно, наряду с абсолютными величинами, применяются в медицине и здравоохранении, прежде всего для сопоставления одной совокупности с другой. Вычисление относительных величин выполняется на основе абсолютных значений учетных признаков (абсолютных величин).

Абсолютные величины, полученные непосредственно при измерении учетных признаков каждой единицы наблюдения, сами по себе несут важную информацию о размере того или иного явления (количество родившихся, умерших; число коек в каждой больнице города, число дней болезни каждого больного и др.). Они служат основой для вычисления относительных величин, поскольку без преобразования имеют ограниченное познавательное значение.

Относительные величины вычисляются путем отношения (деления) одной абсолютной величины на другую с последующим умножением полученной дроби на требуемое основание (100, 1000, 10 000, 100 000 и т.д.). Смысл расчета относительных величин состоит в нахождении общей меры, приведении к общему знаменателю (основанию). При этом подбор того или иного множителя связан с тем, что относительные величины целесообразно представлять в целых числах, легко воспринимающихся при анализе. Соответственно, относительные показатели могут быть выражены в процентах (%), промилле (‰), продецимилле (‱), просантимилле (‱). Для их условного обозначения применяется знак “Р”.

Наиболее часто используются следующие ВИДЫ относительных показателей: интенсивные, экстенсивные, соотношения и наглядности.

ИНТЕНСИВНЫЕ ПОКАЗАТЕЛИ отображают распространенность (частоту, уровень) развития явления в своей среде, т.е. в среде, которая продуцирует это явление. Применяются они, чаще всего, в медицине и демографии, отвечая на вопрос: как часто явление встречается в известной среде? Интенсивные величины отражают ЧАСТОТУ (встречаемость) явления в СРЕДЕ, ПОРОДИВШЕЙ это явление и, как правило, вычисляются на основании 1000 (в промилле, ‰). Если полученное значение меньше единицы, целесообразно использование множителей 10 000 (‱), 100 000 (‱).

$$P_{\text{Интенс.}} = \frac{\text{Абсолютный размер явления} * 1000}{\text{Абсолютный размер среды}}.$$

ЭКСТЕНСИВНЫЕ ПОКАЗАТЕЛИ отражают структурный состав изучаемой совокупности. Они характеризуют отношение части статистической совокупности к совокупности в целом (долю, удельный вес, часть от целого), т.е. отношение отдельного элемента к итогу. Эти показатели нельзя использовать для определения уровня изучаемого явления, они демонстрируют только соотношение его частей и чаще выражаются в %.

$$P_{\text{Экт.}} = \frac{\text{Абсолютный размер части явления} * 100}{\text{Абсолютный размер явления в целом}}.$$

ПОКАЗАТЕЛИ СООТНОШЕНИЯ применяются, когда необходимо определить взаимоотношение не связанных между собой совокупностей. Например, обеспеченность населения больничными койками или врачами, соотношение средних медработников и врачей и др. Они вычисляются как отношение величины одного явления к другому явлению и выражаются, в основном, в продецимилле, реже - в процентах, промилле и др.

$$P_{\text{Соотн.}} = \frac{\text{Абсолютный размер одного явления} * 10000}{\text{Абсолютный размер второго явления}}.$$

Для анализа изменения изучаемого явления во времени вычисляются ДИНАМИЧЕСКИЕ ПОКАЗАТЕЛИ на основе динамического ряда.

Динамическим называется ряд чисел, состоящий из однородных сопоставимых величин, характеризующих изменения какого-либо явления за определенные отрезки времени.

Основными ВИДАМИ динамических рядов являются:

- простой (состоит из абсолютных величин);
- сложный (из относительных и средних);
- моментный (состоит из величин, характеризующих размеры явлений на определенные даты);
- интервальный (состоит из величин, характеризующих размеры явления за определенный интервал времени).

Числа динамического ряда называются **УРОВНЕМ**.

К ПОКАЗАТЕЛЯМ ДИНАМИЧЕСКОГО ряда относятся:

- АБСОЛЮТНЫЙ ПРИРОСТ - разность уровней изучаемого явления в данном и предыдущем периоде, может быть положительным, нулевым или отрицательным (убыль);
- ПОКАЗАТЕЛЬ НАГЛЯДНОСТИ - отношение каждого последующего уровня к исходному, принятому за 100%;
- ПОКАЗАТЕЛЬ РОСТА, или темп роста - отношение каждого последующего уровня к предыдущему уровню, принятому за 100%;
- ПОКАЗАТЕЛЬ ПРИРОСТА, или темп прироста - отношение абсолютного прироста к предыдущему уровню, принятому за 100%. Как и абсолютный прирост этот показатель может выражаться

положительным (прирост), нулевым (стабильность) или отрицательным (убыль) значением.

При оценке показателей абсолютного прироста и темпа прироста имеет значение на знак полученной величины. Если показатель положительный, это указывает на прирост изучаемого явления. При его отрицательном значении говорят об убыли. В случае нулевого значения фиксируется стабильное состояние изучаемого явления.

Для анализа полученных относительных и динамических показателей необходимо не только уметь их рассчитывать, но и правильно выбирать графическое изображение с целью наглядного представления результатов исследования. Применение графического метода в статистическом исследовании делает изучаемые величины более доступными для понимания. При применении графического метода важно знать, что содержание каждого показателя должно строго соответствовать виду графического изображения.

ИНТЕНСИВНЫЕ показатели, а также показатели СООТНОШЕНИЯ могут быть наглядно представлены в виде 4-х основных типов диаграмм: столбиковые, линейные, картограммы и картодиаграммы. Картограмма - это географическая карта с различной штриховкой. Картодиаграмма - это географическая карта с нанесенными на нее диаграммами.

ЭКСТЕНСИВНЫЕ показатели графически могут быть изображены секторной или внутрестолбиковой диаграммой.

Для графического изображения динамических показателей наиболее часто применяются линейная и столбиковая диаграммы. Радиальная (лепестковая) диаграмма является частным видом линейной диаграммы, построенной на полярных координатах. Она используется при необходимости изобразить графически динамику явления за замкнутый цикл времени.

Основными требованиями, которые должны быть соблюдены в случае представления рисунков в публикациях, являются:

- Условные обозначения должны соответствовать виду диаграммы, если на рисунке отображается один ряд, то легенда не требуется;
- Единицы измерения и значения шкалы обозначаются на каждой оси диаграммы;
- Количественные значения представленных явлений должны быть показаны на каждом элементе диаграммы, например в секторах круга, над столбиками и так далее;
- Каждый рисунок диаграммы сопровождается названием, помещаемым под рисунком;
- Если в публикации размещаются две и более диаграммы, каждый рисунок обозначается порядковым номером.

## Пример вычисления относительных величин, заполнения статистических таблиц и графического отображения данных

Условие задачи: получены данные (абсолютные величины) статистического наблюдения заболеваемости в районе А города N за 2017 год. Число детей до 18 лет, проживающих в районе А города N, составило 3 500 человек, в том числе мальчиков 1 700 человек, девочек 1 800 человек. Число болевших детей на протяжении года составило 2 900 человек. Из них мальчиков 1 400 человек, девочек 1 500 человек. Общее число зарегистрированных заболеваний у детей за 2017 год составило 3820 случаев. Из них у мальчиков 1 900 случаев, у девочек 1 920 случаев. В том числе число случаев заболеваний скарлатиной у детей за этот год 66 случаев, из них у мальчиков 35 случаев, у девочек 31 случай.

Задание: а) требуется вычислить показатели числа больных лиц, частоту случаев всех заболеваний и случаев скарлатины (интенсивные величины);  
б) вычислить показатели доли (удельного веса) мальчиков и девочек среди детей района и долю заболеваний скарлатиной во всех заболеваниях детей, а также среди мальчиков и девочек (экстенсивные);  
в) построить статистическую таблицу, содержащую интенсивные показатели;  
г) изобразить экстенсивные показатели секторной и внутристолбиковой диаграммами.

Решение: необходимо запустить программу Excel и создать новый лист. На этом листе ввести условия задачи и решение, как показано на рис. 1, сохранить файл.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Получены данные статистического наблюдения - абсолютные величины заболеваемости в р-не N за 2010 год.											
3	1. Число детей, проживающих в районе N в 2010 году:						3 500	человек.				
4	Из них:	мальчики	-	1 700	чел.;							
5		девочки	-	1 800	чел.							
6	2. Число болевших детей на протяжении 2010 года:						2 900	человек.				
7	Из них:	мальчиков	-	1 400	чел.;							
8		девочек	-	1 500	чел.							
9	3. Общее число зарегистрированных заболеваний у детей за 2010 год:							3 820	случаев.			
10	Из них:	мальчики	-	1 900	случаев;							
11		девочки	-	1 920	случаев.							
12	4. Число заболеваний скарлатиной у детей за 2010 год:						66	случаев.				
13	Из них:	мальчики	-	35	случаев;							
14		девочки	-	31	случай.							
15	<b>Задание:</b> а) вычислить показатели числа больных лиц, частоту случаев всех заболеваний и случаев скарлатины (интенсивные величины); б) вычислить показатели доли (удельного веса) мальчиков и девочек среди детей района и долю заболеваний скарлатиной во всех заболеваниях детей, а также среди мальчиков и девочек (экстенсивные); в) построить статистическую таблицу, содержащую интенсивные показатели; г) изобразить экстенсивные показатели секторной и внутристолбиковой диаграммами.											
16												
17	<b>Решение:</b>											
18	<b>а) вычисление показателей числа больных лиц, частоты случаев всех заболеваний и случаев скарлатины (интенсивные величины)</b>											
19	$P_{\text{бол.лиц}} \text{ всего} = (2900/3500)*1000=$				828,6	%	$P_{\text{случ.}} = (3820/3500)*1000=$				1091,4	%
20	$P_{\text{бол.мальчики}} = (1400/1700)*1000=$				823,5	%	$P_{\text{случ.мальчики}} = (1900/1700)*1000=$				1117,6	%
21	$P_{\text{бол.девочки}} = (1500/1800)*1000=$				833,3	%	$P_{\text{случ.девочки}} = (1920/1800)*1000=$				1066,7	%
22												
23	$P_{\text{случ.скарп.}} = (66/3500)*1000=$				18,9	%						
24	$P_{\text{случ.у мальчиков}} = (35/1700)*1000=$				20,6	%						
25	$P_{\text{случ.у девочек}} = (31/1800)*1000=$				17,2	%						

Рис. 1. Пример вычисления относительных величин, а также их наглядного изображения в программе Excel

Вычисления выполняются с помощью формул, в которых используются адреса ячеек, содержащие абсолютные величины.

а) вычисление интенсивных показателей числа больных лиц, случаев всех заболеваний и заболеваний скарлатиной. Они отражают ЧАСТОТУ (встречаемость) явления в среде, ПОРОДИВШЕЙ это явление. Числитель – явление, знаменатель – среда, основание 1000.

$$P_{\text{Интенс.}} = \frac{\text{Абсолютный размер явления} * 1000}{\text{Абсолютный размер среды}}.$$

В программе Excel вычисления производятся с помощью формул, которые могут включать абсолютные значения или адреса ячеек, в которые введены данные условия задачи. Формулы расчета интенсивных показателей:

$$P_{\text{бол.лиц всего}} = (2900/3500) * 1000 = 828,6\text{‰}$$

$$P_{\text{бол.мальчики}} = (1400/1700) * 1000 = 823,5\text{‰}$$

$$P_{\text{бол.девочки}} = (1500/1800) * 1000 = 833,3\text{‰}$$

$$P_{\text{случ.}} = (3820/3500) * 1000 = 1091,4\text{‰}$$

$$P_{\text{случ.мальчики}} = (1900/1700) * 1000 = 1117,6\text{‰}$$

$$P_{\text{случ.девочки}} = (1920/1800) * 1000 = 1066,7\text{‰}$$

$$P_{\text{случ.скарл.}} = (66/3500) * 1000 = 18,9\text{‰}$$

$$P_{\text{случ.у мальчиков}} = (35/1700) * 1000 = 20,6\text{‰}$$

$$P_{\text{случ.у девочек}} = (31/1800) * 1000 = 17,2\text{‰}$$

б) вычисление экстенсивных показателей отражающих долю (УДЕЛЬНЫЙ ВЕС, часть явления в этом же явлении) числа мальчиков и девочек, проживающих в изучаемом районе, долю заболеваний скарлатиной в общей заболеваемости детей, а также среди мальчиков и девочек соответственно. Числитель – часть явления, знаменатель – всё явление, основание 100 (выражается в %).

$$P_{\text{Экт.}} = \frac{\text{Абсолютный размер части явления} * 100}{\text{Абсолютный размер явления в целом}}.$$

Формулы расчета интенсивных показателей:

$$P_{\text{доля мальчиков}} = (1700/3500) * 100 = 48,6\%$$

$$P_{\text{уд.вес случаев скарлатины}} = (66/3820) * 100 = 1,7\%$$

$$P_{\text{доля девочек}} = (1800/3500) * 100 = 51,4\%$$

$$P_{\text{уд.вес скарлатины среди мальчиков}} = (35/1900) * 100 = 1,8\%$$

$$P_{\text{уд.вес скарлатины среди девочек}} = (31/1920) * 100 = 1,6\%$$

в) заполнение статистической таблицы интенсивных относительных величин (таблица 1).

Таблица 1. Распространенность заболеваний среди детей в 2017 году

Пол	Число болевших на 1000 детей (‰)	Число случаев заболеваний на 1000 детей (‰)	Число случаев скарлатины на 1000 детей (‰)
Мальчики	823,5	1117,6	17,2
Девочки	833,3	1066,7	20,6
Итого:	828,6	1091,4	18,9

г) создание секторной (рис. 2) и внутрискладчатой (рис. 3) диаграмм, наглядно демонстрирующих экстенсивные показатели. В программе Excel изготовление таких рисунков производится командой «Вставка» с предварительным выделением диапазона ячеек содержащих данные для изображения. Значение изображаемых величин могут быть взяты из ячеек, в которых вычислены доли мальчиков и девочек, проживающих в изучаемом районе.

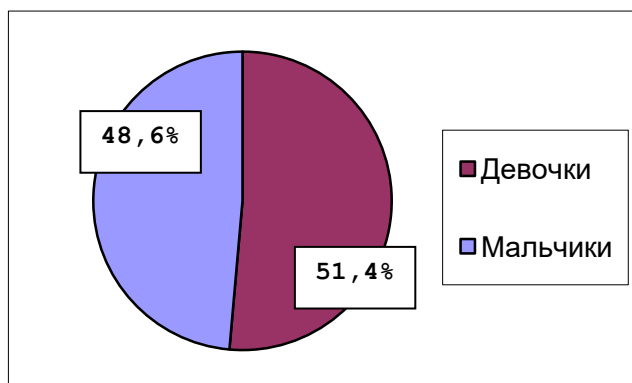


Рис. 2. Доля мальчиков и девочек, проживающих в районе А города N в 2017 году

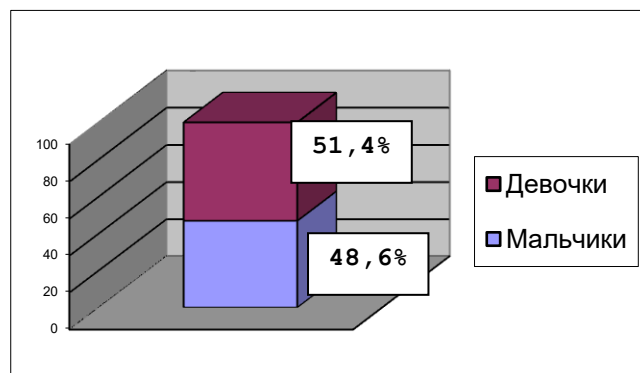


Рис. 3. Доля мальчиков и девочек, проживающих в районе А города N в 2017 году

### Пример расчета и анализа показателей первичной заболеваемости населения и структуры заболеваемости

Условие задачи: получены данные статистического наблюдения: среднегодовая численность жителей города N в 2016 году составляла 300 тысяч человек. За год зарегистрировано 216 850 случаев заболеваний, в том числе по группам: заболевания органов дыхания — 92 000 случаев, болезни органов кровообращения — 82 000 случаев, травмы — 16 500 случаев, новообразования — 16 000 случаев, инфекционные болезни - 9000 случаев, прочие заболевания — 1350.

Задание: составьте статистическую таблицу, вычислите интенсивные показатели частоты и экстенсивные показатели структуры заболеваемости жителей района N за 2016 год. Отобразите эти показатели соответствующими диаграммами. Сделайте вывод.



Решение: необходимо запустить программу Excel, ввести условие задачи, создать макет статистической таблицы, ввести данные и формулы вычислений, как показано на рис. 4, сохранить файл. С целью наглядного изображения и анализа относительных величин требуется создать рисунки диаграмм соответствующие типу показателей.

Для вычисления относительных величин необходимо занести абсолютные данные из условия задачи в отдельные ячейки таблицы, а затем, в соответствующих ячейках поместить расчетные формулы. Правила и результаты вычислений интенсивных и экстенсивных показателей заболеваемости представлен в таблице 2, а их графическое изображение на рис. 5 и рис. 6.

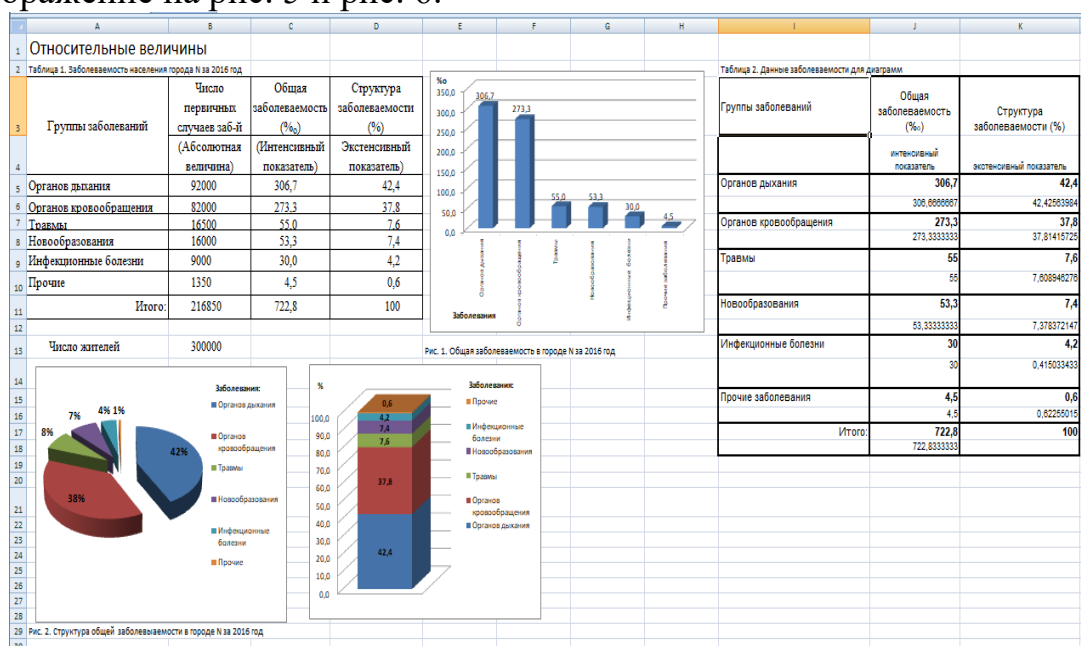


Рис. 4. Вычисление относительных величин в программе Excel

Таблица 2. Показатели заболеваемости населения города N за 2016 год

Группы заболеваний	Общая заболеваемость (%) интенсивный показатель	Структура заболеваемости (%) экстенсивный показатель
Органов дыхания	<b>306,7</b> =92000/300000*1000	<b>42,4</b> =92000/216850*100
Органов кровообращения	<b>273,3</b> =82000/300000*1000	<b>37,8</b> =82000/216850*100
Травмы	<b>55,0</b> =16500/300000*1000	<b>7,6</b> =16500/216850*100
Новообразования	<b>53,3</b> =16000/300000*1000	<b>7,4</b> =16000/216850*100
Инфекционные болезни	<b>30,0</b> =9000/300000*1000	<b>4,2</b> =9000/216850*100
Прочие	<b>4,5</b> =1350/300000*1000	<b>0,6</b> =1350/216850*100
Итого:	<b>722,8</b> =216850/300000*1000	<b>100</b>

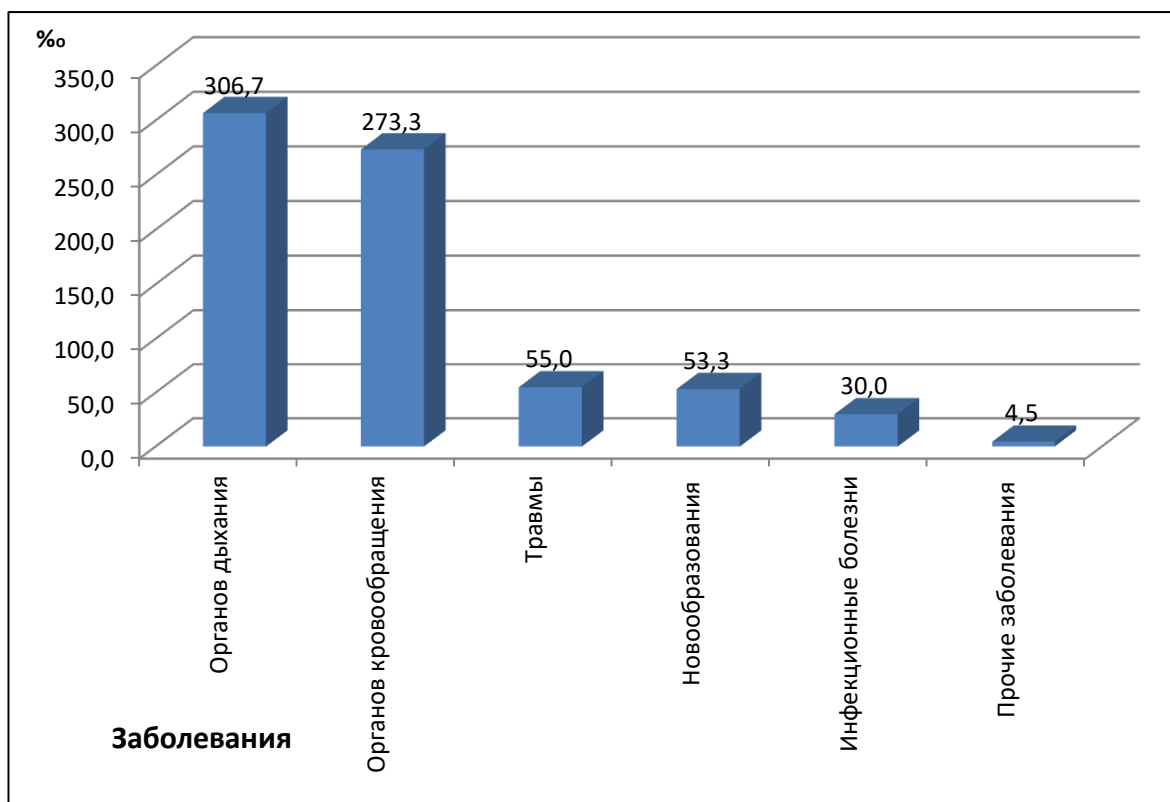


Рис. 5. Уровень заболеваемости по группам болезней жителей города N в 2016 году на 1000 жителей

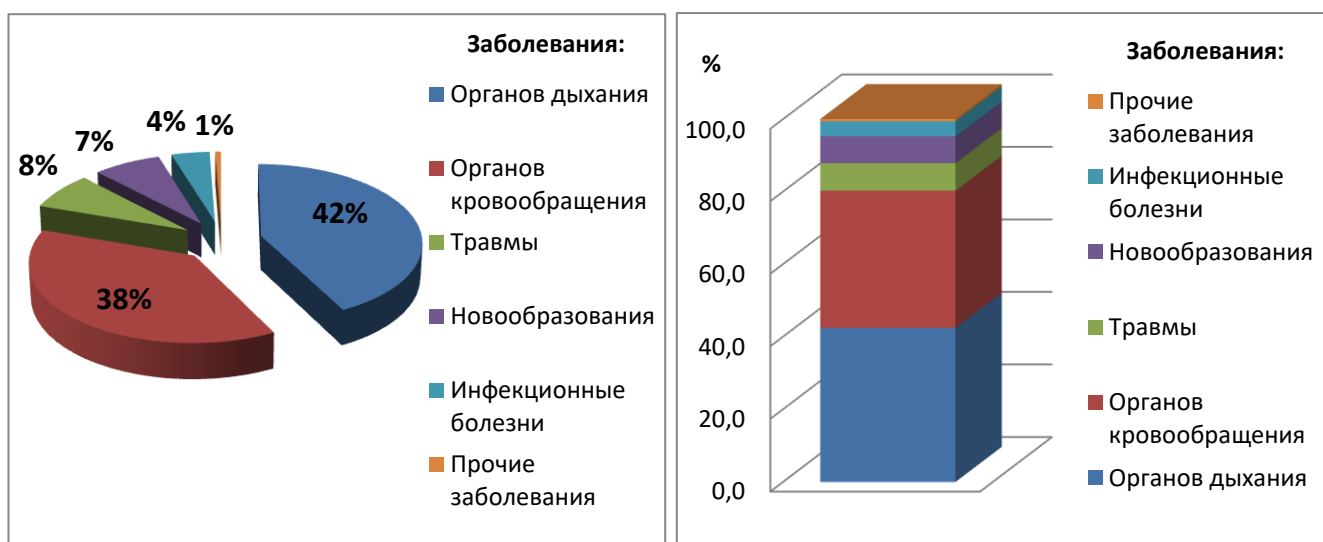


Рис. 6. Структура заболеваемости жителей города N в 2016 году (%)

**Вывод:** заболеваемость в районе N характеризуется повышенным уровнем болезней органов дыхания и сердечно-сосудистой системы по сравнению с другими классами заболеваний, причем их доли в структуре заболеваемости приблизительно равны.

## Пример расчета и анализа показателей динамики численности студентов в Волгоградской области

Условие задачи: получены данные о численности студентов и населения Волгоградской области с 2004 по 2009 годы. В 2004 году число студентов составило 25 000 человек, в 2005 – 26 300 человек, в 2006 – 27 000, в 2007 – 26 800 человек, в 2008 – 28 300 в 2009 – 26 100 человек.

Задание: вычислить показатели динамики числа студентов, представить их графическим изображением, проанализировать и сделать вывод.

Решение: необходимо запустить программу Excel и на отдельном листе введите данные и решение задачи, как показано ниже, сохранить файл.

В программе Excel для вычисления динамических величин необходимо занести абсолютные данные из условия задачи в отдельные ячейки таблицы, а затем изготовить макет статистической таблицы и в соответствующих ячейках этого макета ввести расчетные формулы. Принцип вычисления динамических показателей представлен в таблице 3, а графическое изображение динамики на основе показателя наглядности демонстрирует рисунок 7.

Таблица 3. Динамика количества студентов в Волгоградской области за 6 лет

Годы	Абсолютное значение числа студентов (чел.)	Динамика числа студентов в показателях:			
		абсолютного прироста или убыли (чел.)	наглядности (%)	роста/ убыли (%)	прироста/ убыли (%)
2004	25 000	-	100,0	-	-
2005	26 300	1300 =26300-25000	105,2 =26300/25000*100	105,2 =26300/25000*100	5,2 =1300/25000*100
2006	27 000	700 =27000-26300	108,0 =27000/25000*100	102,7 =27000/26300*100	2,7 =700/26300*100
2007	26 800	-200	107,2	99,3	-0,7
2008	28 300	1 500	113,2	105,6	5,6
2009	26 100	-2 200	104,4	92,2	-7,8

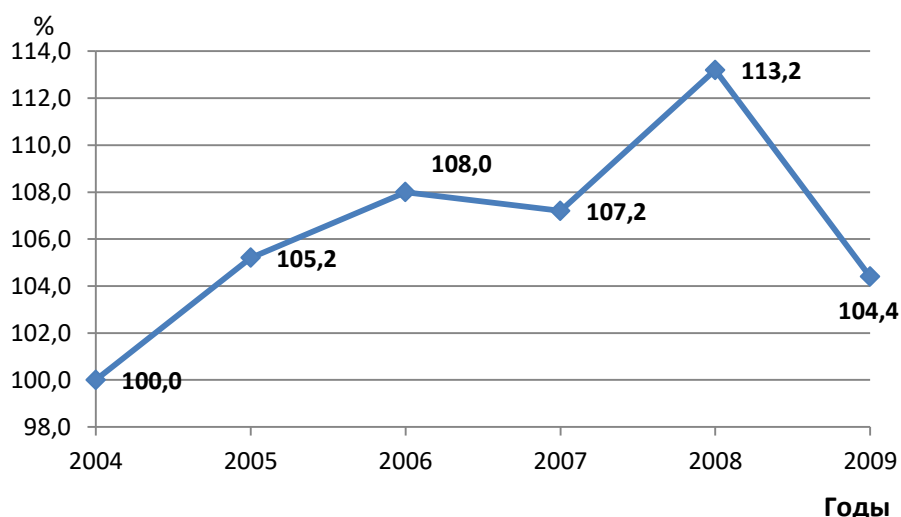


Рис. 7. Динамика числа студентов в Волгоградской области с 2004 по 2009 годы

**Вывод:** изменение относительного числа студентов в Волгоградской области за 6 лет (с 2004 по 2009 годы) характеризуется увеличением до 2008 года с последующей тенденцией к снижению до уровня 2005 года.

## ВАРИАЦИОННЫЕ РЯДЫ. ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Особое место в статистическом анализе принадлежит определению среднего уровня изучаемого признака или явления. Средний уровень признака измеряют средними величинами.

Средняя величина характеризует общий количественный уровень изучаемого признака и является групповым свойством статистической совокупности. Она нивелирует, ослабляет случайные отклонения индивидуальных наблюдений в ту или иную сторону и выдвигает на первый план основное, типичное свойство изучаемого признака.

Средние величины широко используются:

1. Для оценки состояния здоровья населения: характеристики физического развития (рост, вес, окружность грудной клетки и пр.), выявления распространенности и длительности различных заболеваний, анализа демографических показателей (естественного движения населения, средней продолжительности предстоящей жизни, воспроизводства населения, средней численности населения и др.).
2. Для изучения деятельности лечебно-профилактических учреждений, медицинских кадров и оценки качества их работы, планирования и определения потребности населения в различных видах медицинской помощи (среднее число обращений или посещений на одного жителя в год, средняя длительность пребывания больного в стационаре, средняя

продолжительность обследования больного, средняя обеспеченность врачами, койками и пр.).

3. Для характеристики санитарно-эпидемиологического состояния (средняя запыленность воздуха в цехе, средняя площадь на одного человека, средние нормы потребления белков, жиров и углеводов и т. д.).
4. Для определения медико-физиологических показателей в норме и патологии, при обработке лабораторных данных, для установления достоверности результатов выборочного исследования в социально-гигиенических, клинических, экспериментальных исследованиях.

Вычисление средних величин выполняется на основе вариационных рядов. ВАРИАЦИОННЫЙ РЯД – это однородная в качественном отношении статистическая совокупность, отдельные единицы которой характеризуют количественные различия изучаемого признака или явления.

Количественная вариация может быть двух типов: прерывная (дискретная) и непрерывная.

Прерывный (дискретный) признак выражается только целым числом и не может иметь никаких промежуточных значений (например, число посещений, численность населения участка, число детей в семье, степень тяжести болезни в баллах и др.).

Непрерывный признак может принимать любые значения в определенных пределах, в том числе и дробные, и выражается лишь приближенно (например, вес – для взрослых можно ограничиться килограммами, а для новорожденных – граммами; рост, артериальное давление, время, потраченное на прием больного, и т. д.).

Цифровое значение каждого отдельного признака или явления, входящего в вариационный ряд, называется вариантой и обозначается буквой *V*. В математической литературе встречаются и другие обозначения, например *x* или *y*.

Вариационный ряд, где каждая варианта указана один раз, называется ПРОСТЫМ. Такие ряды используются в большинстве статистических задач в случае компьютерной обработки данных.

При увеличении числа наблюдений, как правило, встречаются повторяющиеся значения вариантов. В этом случае создается СГРУППИРОВАННЫЙ ВАРИАЦИОННЫЙ РЯД, где указывается число повторений (частота, обозначается буквой «*p*»).

РАНЖИРОВАННЫЙ ВАРИАЦИОННЫЙ РЯД состоит из вариантов, расположенных в порядке возрастания или убывания. Как простой, так и сгруппированный ряды могут быть составлены с ранжированием.

ИНТЕРВАЛЬНЫЙ ВАРИАЦИОННЫЙ РЯД составляют с целью упрощения последующих вычислений, выполняемых без использования компьютера, при очень большом числе единиц наблюдения (более 1000).

НЕПРЕРЫВНЫЙ ВАРИАЦИОННЫЙ РЯД включает значения вариантов, которые могут выражаться любыми значениями.

Если в вариационном ряде значения признака (варианты) заданы в виде отдельных конкретных чисел, то такой ряд называют ДИСКРЕТНЫМ.

Общими характеристиками значений признака, отражаемого в вариационном ряде, являются средние величины. Среди них наиболее применяемые: средняя арифметическая величина  $M$ , мода  $Mo$  и медиана  $Me$ . Каждая из этих характеристик своеобразна. Они не могут подменить друг друга и лишь в совокупности достаточно полно и в сжатой форме представляют собой особенности вариационного ряда.

МОДОЙ ( $Mo$ ) называют значение наиболее часто встречающейся варианты.

МЕДИАНА ( $Me$ ) – это значение варианты, делящей ранжированный вариационный ряд пополам (с каждой стороны медианы находится половина вариантов). В редких случаях, когда имеется симметричный вариационный ряд, мода и медиана равны между собой и совпадают со значением средней арифметической.

Наиболее типичной характеристикой значений вариантов является средняя арифметическая величина ( $M$ ). В математической литературе она обозначается  $\bar{X}$ .

СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ ВЕЛИЧИНА ( $M, \bar{X}$ ) – это общая количественная характеристика определенного признака изучаемых явлений, составляющих качественно однородную статистическую совокупность. Различают среднюю арифметическую простую и взвешенную. Средняя арифметическая простая вычисляется для простого вариационного ряда путем суммирования всех вариантов и делением этой суммы на общее количество вариантов, входящих в данный вариационный ряд. Вычисления проводятся по формуле:

$$M = \frac{\sum V}{n},$$

где:  $M$  - средняя арифметическая простая;

$\sum V$  - сумма вариантов;

$n$  - число наблюдений.

В сгруппированном вариационном ряде определяют взвешенную среднюю арифметическую. Формула ее вычисления:

$$M = \frac{\sum Vp}{n},$$

где:  $M$  - средняя арифметическая взвешенная;

$\sum Vp$  - сумма произведений вариант на их частоты;

$n$  - число наблюдений.

При большом числе наблюдений в случае ручных вычислений может применяться способ моментов.

Средняя арифметическая имеет следующие свойства:

- сумма отклонений вариант от средней ( $\sum d$ ) равна нулю (см. табл. 15);
- при умножении (делении) всех вариант на один и тот же множитель (делитель) средняя арифметическая умножается (делится) на тот же множитель (делитель);
- если прибавить (вычесть) ко всем вариантам одно и то же число, средняя арифметическая увеличивается (уменьшается) на это же число.

Средние арифметические величины, взятые сами по себе, без учета вариабельности рядов, из которых они вычислены, могут не в полной мере отражать свойства вариационного ряда, в особенности когда необходимо сопоставление с другими средними. Близкие по значению средние могут быть получены из рядов с различной степенью рассеяния. Чем ближе друг к другу отдельные варианты по своей количественной характеристике, тем меньше РАССЕЯНИЕ (КОЛЕБЛЕМОСТЬ, ВАРИАБЕЛЬНОСТЬ) ряда, тем типичнее его средняя.

Основными параметрами, которые позволяют оценить вариабельность признака, являются:

- Размах;
- Амплитуда;
- Среднее квадратическое отклонение;
- Коэффициент вариации.

Приблизительно о колеблемости признака можно судить по размаху и амплитуде вариационного ряда. Размах указывает на максимальную ( $V_{\max}$ ) и минимальную ( $V_{\min}$ ) варианты в ряду. Амплитуда ( $A_m$ ) является разностью этих вариантов:  $A_m = V_{\max} - V_{\min}$ .

Основной, общепринятой мерой колеблемости вариационного ряда являются ДИСПЕРСИЯ (D). Но наиболее часто применяется более удобный параметр, вычисляемый на основе дисперсии - среднее квадратическое отклонение ( $\sigma$ ). Оно учитывает величину отклонения ( $d$ ) каждой варианты вариационного ряда от его средней арифметической ( $d = V - M$ ).

Поскольку отклонения вариант от средней могут быть положительными и отрицательными, то при суммировании они дают значение «0» ( $\sum d = 0$ ). Чтобы избежать этого, величины отклонения ( $d$ ) возводятся во вторую степень и усредняются. Таким образом, дисперсия вариационного ряда является средним квадратом отклонений вариант от средней арифметической и вычисляется по формуле:

$$D = \frac{\sum d^2}{n}.$$

Она является важнейшей характеристикой вариабельности и применяется для вычисления многих статистических критериев.

Поскольку дисперсия выражается квадратом отклонений, ее величина не может использоваться в сопоставлении со средней арифметической. Для этих целей применяется СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ, которое обозначается знаком «Сигма» ( $\sigma$ ). Оно характеризует среднее отклонение всех вариантов вариационного ряда от средней арифметической величины в тех же единицах, что и сама средняя величина, поэтому они могут использоваться совместно.

Среднее квадратическое отклонение определяют по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}}.$$

Указанная формула применяется при числе наблюдений ( $n$ ) больше 30. При меньшем числе  $n$  значение среднего квадратического отклонения будет иметь погрешность, связанную с математическим смещением ( $n - 1$ ). В связи с этим, более точный результат может быть получен с помощью учета такого смещения в формуле расчета стандартного отклонения:

$$s = \sqrt{\frac{\sum d^2}{n-1}}.$$

СТАНДАРТНОЕ ОТКЛОНЕНИЕ ( $s$ ) – это оценка среднеквадратического отклонения случайной величины  $X$  относительно её математического ожидания на основе несмещённой оценки её дисперсии.

При значениях  $n > 30$  среднее квадратическое отклонение ( $\sigma$ ) и стандартное отклонение ( $s$ ) будут одинаковыми ( $\sigma = s$ ). Поэтому в большинстве практических пособий эти критерии рассматриваются как равнозначные. В программе Excel вычисление стандартного отклонения может быть выполнено функцией =СТАНДОТКЛОН(диапазон). А с целью расчета среднего квадратического отклонения требуется создать соответствующую формулу.

Среднее квадратическое или стандартное отклонение позволяет определить, насколько значения признака могут отличаться от среднего значения. Предположим, существуют два города с одинаковой средней дневной температурой в летний период. Один из этих городов расположен на побережье, а другой на континенте. Известно, что в городах, расположенных на побережье, различия дневных температур меньше, чем у городов, расположенных внутри континента. Поэтому среднее квадратическое отклонение дневных температур у прибрежного города будет меньше, чем у второго города. На практике это означает, что средняя температура воздуха каждого конкретного дня в городе, расположенного



на континенте будет сильнее отличаться от среднего значения, чем в городе на побережье. Кроме того стандартное отклонение позволяет оценить возможные отклонения температуры от средней с требуемым уровнем вероятности.

Согласно теории вероятности, в явлениях, подчиняющихся нормальному закону распределения, между значениями средней арифметической, среднего квадратического отклонения и вариантами существует строгая зависимость (ПРАВИЛО ТРЕХ СИГМ). Например, 68,3% значений варьирующего признака находятся в пределах  $M \pm 1\sigma$ , 95,5% — в пределах  $M \pm 2\sigma$  и 99,7% — в пределах  $M \pm 3\sigma$ .

Величина среднего квадратического отклонения позволяет судить о характере однородности вариационного ряда и исследуемой группы. Если величина среднего квадратического отклонения небольшая, то это свидетельствует о достаточно высокой однородности изучаемого явления. Среднюю арифметическую в таком случае следует признать вполне характерной для данного вариационного ряда. Однако слишком малая величина сигмы заставляет думать об искусственном подборе наблюдений. При очень большой сигме средняя арифметическая в меньшей степени характеризует вариационный ряд, что говорит о значительной вариабельности изучаемого признака или явления или о неоднородности исследуемой группы. Однако сопоставление величины среднего квадратического отклонения возможно только для признаков одинаковой размерности. Действительно, если сравнивать разнообразие веса новорожденных детей и взрослых, мы всегда получим более высокие значения сигмы у взрослых.

Сравнение вариабельности признаков различной размерности может быть выполнено с помощью КОЭФФИЦИЕНТА ВАРИАЦИИ. Он выражает разнообразие в процентах от средней величины, что позволяет производить сравнение различных признаков. Коэффициент вариации в медицинской литературе обозначается знаком «С», а в математической «v» и вычисляется по формуле:

$$C = \frac{\sigma}{M} \times 100\% .$$

Значения коэффициента вариации менее 10% свидетельствует о малом рассеянии, от 10 до 20% – о среднем, более 20% – о сильном рассеянии вариант вокруг средней арифметической.

Средняя арифметическая величина, как правило, вычисляется на основе данных выборочной совокупности. При повторных исследованиях под влиянием случайных явлений средняя арифметическая может изменяться. Это обусловлено тем, что исследуется, как правило, только часть возможных единиц наблюдения, то есть выборочная совокупность.

Информация обо всех возможных единицах, представляющих изучаемое явление, может быть получена при изучении всей генеральной совокупности, что не всегда возможно. В то же время с целью обобщения данных эксперимента представляет интерес величина средней в генеральной совокупности. Поэтому для формулировки общего вывода об изучаемом явлении, результаты, полученные на основе выборочной совокупности, должны быть, перенесены на генеральную совокупность статистическими методами.

Чтобы определить степень совпадения выборочного исследования и генеральной совокупности, необходимо оценить величину ошибки, которая неизбежно возникает при выборочном наблюдении. Такая ошибка называется «ОШИБКОЙ РЕПРЕЗЕНТАТИВНОСТИ» или «Средней ошибкой средней арифметической». Она фактически является разностью между средними, полученными при выборочном статистическом наблюдении, и аналогичными величинами, которые были бы получены при сплошном исследовании того же объекта, т.е. при изучении генеральной совокупности. Поскольку выборочная средняя является случайной величиной, такой прогноз выполняется с приемлемым для исследователя уровнем вероятности. В медицинских исследованиях он составляет не менее 95%.

Ошибку репрезентативности нельзя смешивать с ошибками регистрации или ошибками внимания (описки, просчеты, опечатки и др.), которые должны быть сведены до минимума адекватной методикой и инструментами, применяемыми при проведении эксперимента.

Величина ошибки репрезентативности зависит как от объема выборки, так и от вариабельности признака. Чем больше число наблюдений, тем ближе выборка к генеральной совокупности и тем меньше ошибка. Чем более изменчив признак, тем больше величина статистической ошибки.

На практике для определения ошибки репрезентативности в вариационных рядах пользуются следующей формулой:

$$m = \frac{\sigma}{\sqrt{n-1}},$$

где:  $m$  – ошибка репрезентативности;

$\sigma$  – среднее квадратическое отклонение;

$n$  – число наблюдений в выборке.

Из формулы видно, что размер средней ошибки прямо пропорционален среднему квадратическому отклонению, т. е. вариабельности изучаемого признака, и обратно пропорционален корню квадратному из числа наблюдений.

При выполнении статистического анализа на основе вычисления относительных величин построение вариационного ряда не является обязательным. При этом определение средней ошибки для относительных показателей может выполняться по упрощенной формуле:

$$m = \sqrt{\frac{Pq}{n-1}},$$

где:  $P$  – величина относительного показателя, выраженного в процентах, промилле и т.д.;

$q$  – величина, обратная  $P$  и выраженная как  $(1-P)$ ,  $(100-P)$ ,  $(1000-P)$  и т. д., в зависимости от основания, на которое рассчитан показатель;

$n$  – число наблюдений в выборочной совокупности.

Однако, указанная формула вычисления ошибки репрезентативности для относительных величин может применяться только в том случае, когда значение показателя меньше его основания. В ряде случаев расчета интенсивных показателей такое условие не соблюдается, и показатель может выражаться числом более 100% или 1000‰. В такой ситуации выполняется построение вариационного ряда и вычисление ошибки репрезентативности по формуле для средних величин на основе среднего квадратического отклонения.

Прогнозирование величины средней арифметической в генеральной совокупности выполняется с указанием двух значений – минимального и максимального. Эти крайние значения возможных отклонений, в пределах которых может колебаться искомая средняя величина генеральной совокупности, называются «ДОВЕРИТЕЛЬНЫЕ ГРАНИЦЫ».

Постулатами теории вероятностей доказано, что при нормальном распределении признака с вероятностью 99,7%, крайние значения отклонений средней будут не больше величины утроенной ошибки репрезентативности ( $M \pm 3m$ ); в 95,5% – не больше величины удвоенной средней ошибки средней величины ( $M \pm 2m$ ); в 68,3% – не больше величины одной средней ошибки ( $M \pm 1m$ ) (рис. 7).

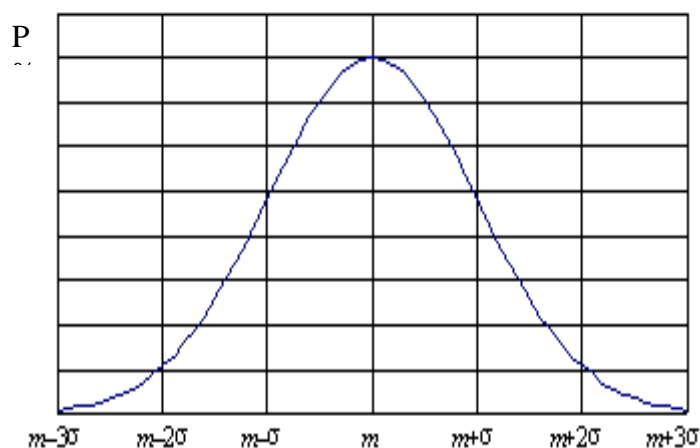


Рис. 7. Плотность вероятностей нормального распределения.

Отметим, что приведенное выше утверждение справедливо только для признака, который подчиняется нормальному закону распределения Гаусса.

Большинство экспериментальных исследований, в том числе и в области медицины, связано с измерениями, результаты которых могут принимать практически любые значения в заданном интервале, поэтому, как правило, описываются моделью непрерывных случайных величин. В связи с этим в большинстве статистических методов рассматриваются непрерывные распределения. Одним из таких распределений, имеющим основополагающую роль в математической статистике, является НОРМАЛЬНОЕ, ИЛИ ГАУССОВО, РАСПРЕДЕЛЕНИЕ.

Это объясняется целым рядом причин.

1. Прежде всего, многие экспериментальные наблюдения можно успешно описать с помощью нормального распределения. Следует сразу же отметить, что не существует распределений эмпирических данных, которые были бы в точности нормальными, поскольку нормально распределенная случайная величина находится в пределах от  $-\infty$  до  $+\infty$ , чего никогда не встречается на практике. Однако нормальное распределение очень часто хорошо подходит как приближение.

Проводятся ли измерения веса, роста и других физиологических параметров организма человека - везде на результаты оказывает влияние очень большое число случайных факторов (естественные причины и ошибки измерения). Причем, как правило, действие каждого из этих факторов незначительно. Опыт показывает, что результаты именно в таких случаях будут распределены приближенно нормально.

2. Многие распределения, связанные со случайной выборкой, при увеличении объема последней переходят в нормальное.

3. Нормальное распределение хорошо подходит в качестве приближенного описания других непрерывных распределений (например, асимметричных).

4. Нормальное распределение обладает рядом благоприятных математических свойств, во многом обеспечивших его широкое применение в статистике.

В то же время следует отметить, что в медицинских данных встречается много экспериментальных распределений, описание которых моделью нормального распределения невозможно. Для этого в статистике разработаны методы, которые принято называть «Непараметрическими».

Выбор статистического метода, который подходит для обработки данных конкретного эксперимента, должен производиться в зависимости от принадлежности полученных данных к нормальному закону

распределения. Проверка гипотезы на подчинение признака нормальному закону распределения выполняется с помощью гистограммы распределения частот (графика), а также ряда статистических критериев. Среди них:

- Критерий асимметрии ( $\beta$ );
- Критерий проверки на эксцесс ( $\gamma$ );
- Критерий Шапиро – Уилкса ( $W$ ) .

Анализ характера распределения данных (его еще называют проверкой на нормальность распределения) осуществляется по каждому параметру. Чтобы уверенно судить о соответствии распределения параметра нормальному закону, необходимо достаточно большое число единиц наблюдения (не менее 30 значений).

Для нормального распределения критерии асимметрии и эксцесса принимают значение 0. Если распределение смещено вправо  $\beta > 0$  (положительная асимметрия), при  $\beta < 0$  - график распределения смещен влево (отрицательная асимметрия). Критерий асимметрии проверяет форму кривой распределения. В случае нормального закона  $\gamma = 0$ . При  $\gamma > 0$  кривая распределения острее, если  $\gamma < 0$  пик более сглаженный, чем функция нормального распределения.

Для проверки на нормальность по критерию Шапиро – Уилкса требуется найти значение этого критерия по статистическим таблицам при необходимом уровне значимости и в зависимости от числа единиц наблюдения (степеней свободы). Приложение 1. Гипотеза о нормальности отвергается при малых значениях этого критерия, как правило, при  $W < 0,8$ .

### **Пример построения вариационных рядов, вычисления средних величин, создания графика распределения признака и проверки на нормальность распределения**

Задача: Для выявления общей характеристики частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделении №1 больницы выполнено 16 измерений пульса у детей: Иванов Василий – 120 ударов в минуту, Сидоров Костя – 130 уд/мин, Казаков Василий – 115 уд/мин, Быков Иван - 120 уд/мин, Ежков Дмитрий - 120 уд/мин, Попов Александр - 125 уд/мин, Комин Алексей - 110 уд/мин, Ким Эльдар - 125 уд/мин, Грибин Петр - 115 уд/мин, Котов Дмитрий - 120 уд/мин, Носов Семен - 125 уд/мин, Лобанов Семен - 135 уд/мин, Бычков Евгений - 115 уд/мин, Мокин Юрий - 130 уд/мин, Захаров Михаил - 125 уд/мин, Чернов Павел - 120 уд/мин.

**Задание:** а) составьте простой не ранжированный вариационный ряд, вычислите среднюю арифметическую величину и определите степень рассеяния вариантов в вариационном ряду;

б) выполните упорядочение (ранжирование) ряда по возрастанию и определите медиану;

в) составьте сгруппированный ранжированный вариационный ряд, постройте график распределения признака, определите моду и среднюю взвешенную величину;

г) определите вид распределения признака и статистические критерии нормальности распределения;

д) вычислите ошибку репрезентативности и доверительные границы колебания средней арифметической в генеральной совокупности.

**Решение:** необходимо запустить программу Excel, создать новый лист, и на этом листе ввести условие и выполнить вычисления, как показано на рис.8, сохранить файл.

	A	B	C	D	E	F	G	H
1	Расчет средних вариационного ряда.							
2								
3	Частота пульса детей, поступивших в отделение №1 больницы.							
4								
5		V	d	d2		Расчет мастером "Анализ данных -		
6	1 Иванов Е	120	-1,88	3,52	1	Описательная статистика		
7	2 Сидоров	130	8,13	66,02	2	Столбец1		
8	3	115	-6,88	47,27	3	Среднее	121,875	
9	4	120	-1,88	3,52	4	Стандартная ошибка	1,63777	
10	5	120	-1,88	3,52	5	Медиана	120	
11	6	125	3,13	9,77	6	Мода	120	
12	7	110	-11,88	141,02	7	Стандартное отклонение	6,551081	
13	8	125	3,13	9,77	8	Дисперсия выборки	42,91667	
14	9	115	-6,88	47,27	9	Эксцесс	-0,16979	
15	10	120	-1,88	3,52	10	Асимметричность	0,209598	
16	11	125	3,13	9,77	11	Интервал	25	
17	12	135	13,13	172,27	12	Минимум	110	
18	13	115	-6,88	47,27	13	Максимум	135	
19	14	130	8,13	66,02	14	Сумма	1950	
20	15	125	3,13	9,77	15	Счет	16	
21	16	120	-1,88	3,52	16	Наибольший(1)	135	
22	Сумма=	1950	0,00	643,75		Наименьший(1)	110	
23				643,75		Уровень надежности(95,0%); 3,490827		
24	M <sub>1</sub> =	121,8				$\sigma = \sqrt{\frac{\sum d^2}{n}}$		
25	m <sub>1</sub> =	1,64	$m = \frac{\sigma}{\sqrt{n-1}}$		Сигма:		6,343057	
26	C <sub>1</sub> =	5,2					$s = \sqrt{\frac{\sum d^2}{n-1}}$	
27					Стандартное отклонение		6,551081	
28	Уровень значимости		0,05		Довер-й интервал=		3,209971	3,209
29	Довер-е границы		от: 118,6		до: 125,2			
30	Асимметрия		0,209598					
31	Эксцесс=		-0,16979					
	Вывод:	Средняя частота пульса пациентов 2-го отделения с вероятностью 95,5% составляет от 118,6 до 125,2 ударов в минуту, вариабельность малая.						
32								
33								

Рис. 8. Пример построения простого вариационного ряда и вычисления средних величин в программе Excel.

а) Построение простого вариационного ряда частоты пульса детей, поступивших в отделение №1 больницы. Простой вариационный ряд представляет собой таблицу, в которой подлежащим является изучаемый признак, обозначаемый знаком  $V$  (варианта). Все полученные в эксперименте данные вносят в таблицу в порядке их записи в журнал регистрации (см. условие задачи).

**ВНИМАНИЕ!** Если значения вариант повторяются, то каждое из них заносится в отдельную ячейку таблицы. Затем вычисляют сумму вариант, среднюю арифметическую, отклонения ( $d$ ) каждой варианты от средней величины и квадрат отклонения ( $d^2$ ) соответствующими формулами (таблица 4).

Таблица 4. Простой не ранжированный вариационный ряд

	$V$	$d=V-M$	$d^2$
1 Иванов Василий	120	-1,88	3,52
2 Сидоров Костя	130	8,13	66,02
3 ...	115	-6,88	47,27
4 ...	120	-1,88	3,52
5 ...	120	-1,88	3,52
6 ...	125	3,13	9,77
7 ...	110	-11,88	141,02
8 ...	125	3,13	9,77
9 ...	115	-6,88	47,27
10 ...	120	-1,88	3,52
11 ...	125	3,13	9,77
12 ...	135	13,13	172,27
13 ...	115	-6,88	47,27
14 ...	130	8,13	66,02
15 ...	125	3,13	9,77
16 ...	120	-1,88	3,52
Сумма ( $\Sigma$ ) =	<b>1950</b>	0,00	<b>643,75</b>

$n=16$

Вычисление средней арифметической ( $M$ ) в простом вариационном ряду выполняется по формуле:

$$M = \frac{\Sigma V}{n} = 1950/16 = \underline{\underline{121,9}} \text{ уд/мин.}$$

В программе Excel для вычисления средней арифметической может применяться функция =СРЗНАЧ(Диапазон ячеек). Использование этой функции даст такой же результат.

Определение степени рассеяния вариант в вариационном ряду с помощью среднего квадратического отклонения ( $\sigma$  - Сигма) и коэффициента вариации (C). Среднее квадратическое отклонения вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum d^2}{n}} = \text{КОРЕНЬ}(643,75 / 16) = \underline{\underline{\pm 6,343}}.$$

Для упрощения расчета среднего квадратического отклонения при  $n > 30$  в программе Excel может использоваться функция =СТАНДОТКЛОН(Диапазон). Если выполнить вычисления этой функцией в нашем вариационном ряду, то будет получено значение стандартного отклонения = 6,55. Это значение отличается от величины среднего квадратического отклонения, что обусловлено малым числом наблюдений ( $n = 16$ ).

Коэффициент вариации (C) вычисляется формулой:

$$C = \frac{\sigma}{M} \times 100\% = 6,343/121,9 * 100 = \underline{\underline{5,2\%}}, \text{ что}$$

является малым рассеянием ( $<10\%$ ).

Вариабельность признака (рассеяние) оценивается как малая при  $C < 10\%$ , средняя при  $10\% < C < 20\%$ , высокая при  $C > 20\%$

**Вывод: средняя частота пульса пациентов изучаемой группы составляет 121,9 ударов в минуту, вариабельность пульса малая.**

б) Преобразование не ранжированного вариационного ряда в ранжированный (таблица 5) производится в имеющей таблице простого ряда с помощью команды сортировки по столбцу V. С этой целью необходимо предварительно выделить диапазон ячеек таблицы, включающий данные из сортируемого столбца, а затем на вкладке «Главная» нажать кнопку «Сортировка и фильтр». **ВНИМАНИЕ!** При выполнении сортировки требуется расширить диапазон сортируемых ячеек.



Таблица 5. **Ранжированный** вариационный ряд

	$V$	$d=V-M$	$d^2$
1	110	-11,88	141,02
2	115	-6,88	47,27
3	115	-6,88	47,27
4	115	-6,88	47,27
5 Иванов Василий	120	-1,88	3,52
6	120	-1,88	3,52
7	120	-1,88	3,52
8	<b>120</b>	-1,88	3,52
9	<b>120</b>	-1,88	3,52
10	125	3,13	9,77
11	125	3,13	9,77
12	125	3,13	9,77
13	125	3,13	9,77
14 Сидоров Костя	130	8,13	66,02
15	130	8,13	66,02
16	135	13,13	172,27
Сумма( $\Sigma$ ) = =СУММ(Диапазон ячеек)	1950	0	643,75
Средняя арифметическая = =СРЗНАЧ(Диапазон ячеек)	121,875		

Для определения медианы ( $Me$ ) в ранжированном вариационном ряду находят значение варианты, расположенной в центре ряда.

$Me$  = варианта, занимающая срединное положение (если ряд состоит из четного числа вариантов, медианой является полусумма двух центральных вариантов) = **120 уд/мин.**

В программе Excel для определения медианы применяется функция =МЕДИАНА(Диапазон ячеек).

в) Построение сгруппированного ранжированного ряда выполняется в отдельной таблице (таблица 6), которая должна включать столбец частоты повторений вариант ( $p$ ). Данные из сгруппированного вариационного ряда используются для построения точечной диаграммы распределения частот. Он демонстрирует частоту встречаемости вариант ( $p$ ), указываемую по вертикальной оси точечной диаграммы (рис. 9). Число повторений вариант может быть вычислено из диапазона ячеек простого вариационного ряда функцией =СЧЁТЕСЛИ(Диапазон; критерий).

Таблица 6. Сгруппированный  
ранжированный  
вариационный ряд

	<i>V</i>	<i>p</i>
1	110	1
2	115	3
3	<b>120</b>	<b>5</b>
4	125	4
5	130	2
6	135	1
Σ=	735	16

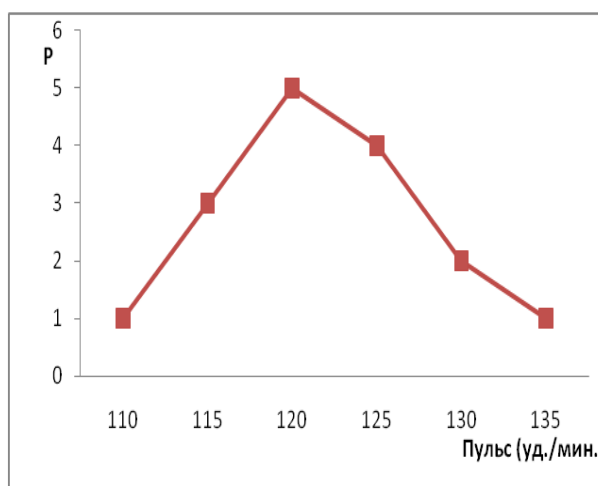


Рис. 9. График распределения признака.

Определение моды может выполняться в сгруппированном вариационном ряду выбором значения варианты, которая повторяется наибольшее число раз.

**Mo**= наиболее часто повторяющаяся варианта = **120 уд/мин**  
(встречается 5 раз).

В программе Excel для определения моды может применяться функция =МОДА(Диапазон данных).

**ВНИМАНИЕ!** Эта функция используется только в простом вариационном ряду.

В сгруппированном вариационном ряду средняя арифметическая вычисляется по модифицированной формуле и называется «Средняя взвешенная»:

$$M = \frac{\sum Vp}{n} = \frac{110*1 + 115*3 + 120*5 + 125*4 + 130*2 + 135*1}{1 + 3 + 5 + 4 + 2 + 1} = \frac{735}{16} = 121,9$$

Независимо от способа построения вариационного ряда и метода вычисления средней арифметической, для одинаковых данных значения их средней величины должны совпадать.

г) Оценка нормальности распределения выполняется с помощью графика распределения (рис. 2) по форме линии, которая должна приближаться к виду кривой Гаусса (функции нормального распределения). Статистические критерии нормальности распределения вычисляются с помощью функций или в модуле «Описательная статистика», который может быть вызван командой «Данные» - «Анализ данных» - «Описательная статистика».

Если указанная команда отсутствует разделе «Данные», необходимо выполнить установку надстройки программы Excel «Пакет анализа». Для этого требуется открыть меню настройки программы кнопкой «Office»,

выбрать «Параметры Excel» -> «Надстройки» -> «Перейти» -> «Пакет анализа» и «Пакет анализа - VBA» -> «ОК». Модуль анализа данных в программе Excel, как правило, не подключается при типовой установке пакета программ Microsoft Office. Поэтому для применения методов статистической обработки данных необходимо однократно выполнить указанную настройку программы.

**ВНИМАНИЕ!** Модуль «Описательная статистика» применяется **только в простых вариационных рядах**. В связи с этим необходимо использовать диапазон ячеек таблицы, в котором составлен ПРОСТОЙ вариационный ряд, созданный в пункте б) настоящей задачи. Результат вычислений, выполненный модулем «Описательная статистика», показан в таблице 7.

Таблица 7. Простой не ранжированный вариационный ряд и результат работы модуля «Описательная статистика»

	V	d=V-M	d <sup>2</sup>
1 Иванов Василий	120	-1,88	3,52
2 Сидоров Костя	130	8,13	66,02
3 ...	115	-6,88	47,27
4 ...	120	-1,88	3,52
5 ...	120	-1,88	3,52
6 ...	125	3,13	9,77
7 ...	110	-	141,02
8 ...	125	3,13	9,77
9 ...	115	-6,88	47,27
10 ...	120	-1,88	3,52
11 ...	125	3,13	9,77
12 ...	135	13,13	172,27
13 ...	115	-6,88	47,27
14 ...	130	8,13	66,02
15 ...	125	3,13	9,77
16 ...	120	-1,88	3,52
$\Sigma=$	1950	0,00	643,75

n=16

Результат выполнения команды «Анализ данных» -> «Описательная статистика».	
Столбец1	
Среднее	121,875
<b>Стандартная ошибка</b>	<b>1,63777</b>
Медиана	120
Мода	120
<b>Стандартное отклонение</b>	<b>6,551081</b>
	42,91667
Дисперсия выборки	7
	-
<b>Экссесс</b>	<b>0,169790,20959</b>
<b>Асимметричность</b>	<b>8</b>
Интервал	25
Минимум	110
Максимум	135
Сумма	1950
Счет	16
Наибольший(1)	135
Наименьший(1)	110
Уровень надежности(95,0%)	3,490827

Экссесс = -0,17. Для нормального распределения эксцесс = 0.

**Распределение близко к нормальному.**

Асимметричность = 0,2. Для нормального распределения асимметричность = 0. **Распределение близко к симметричному.**

Применяя стандартные функции =ЭКССЕСС(Диапазон) и =СКОС(Диапазон) мы получим эти же значения.

д) Определение ошибки средней арифметической ( $m$ , ошибка репрезентативности или стандартная ошибка) выполняется по формуле:

$$m = \frac{\sigma}{\sqrt{n-1}} = 6,343 / \text{КОРЕНЬ}(16-1) = \underline{1,64}$$

Программа Excel позволяет вычислять этот критерий с помощью модуля «Описательная статистика». При этом ошибка средней арифметической ( $m$ ) называется «Стандартная ошибка», так как в качестве среднего квадратического отклонения используется «Стандартное отклонение» (таблица 7). Поэтому в вариационных рядах с числом единиц наблюдения менее 30 возможно несущественное отличие значений полученных расчетной формулой и модулем описательной статистики.

Таким образом, с вероятностью  $P > 95\%$  доверительные границы колебания средней арифметической в генеральной совокупности  $M \pm 2m$   
 $= 121,9 \pm 2 \cdot 1,64 =$

**121,9 ± 3,28 уд/мин.**

Для вычисления доверительных границ в программе Excel может применяться статистическая функция =ДОВЕРИТ(0,05; Значение стандартного отклонения; Число единиц наблюдения). При ее использовании с вычисленными ранее значениями задачи получим:  
=ДОВЕРИТ(0,05; 6,55; 16) = 3,21.

**Вывод: Средняя частота пульса пациентов 1-го отделения с вероятностью 95,5% составляет от 118,6 до 125,2 ударов в минуту.**

### **Пример сравнения рассеяния вариационных рядов**

Условие задачи: Для выявления общей характеристики частоты сердечных сокращений (ЧСС) детей 1-го года жизни в отделении №2 больницы выполнено 17 измерений пульса у детей: 1. Казаков Саша – 130 уд.в мин.; 2. Литвинов Сережа – 135 уд.в мин.; 3...– 125; 4... – 115; 5 ...– 125; 6 ... – 125; 7 ... – 120; 8 ... – 125; 9 ... – 130; 10 ... – 120; 11 ... – 140; 12 ... – 145; 13 ... – 115; 14 ...– 130; 15 ... – 125; 16 ... – 120; 17 ... – 125.

- Задание:
- создать простой, ранжированный вариационный ряды, определить средние величины и доверительные границы средней арифметической;
  - составить сгруппированный, ранжированный вариационный ряд и построить график распределения признака, выполнить проверку его на нормальность;
  - определить параметры вариабельности признака: амплитуду, размах, среднее квадратическое отклонение, коэффициент вариации и ошибку репрезентативности;
  - сравнить характеристики рассеяния вариационных рядов, используемых в примерах данного раздела.

Решение: необходимо запустить программу Excel, создать новый лист, ввести условие задачи, выполнить вычисления с помощью формул и модуля «Описательная статистика», как показано ниже, сохранить файл.

а) Построение простого ранжированного вариационного ряда (таблица 7) и вычисление их основных характеристик модулем «Описательная статистика».

Таблица 7. Простой ранжированный вариационный ряд и его характеристики

	V	d=V-M	d <sup>2</sup>
1	115	-11,18	124,91
2	115	-11,18	124,91
3	115	-11,18	124,91
4	120	-6,18	38,15
5	120	-6,18	38,15
6	125	-1,18	1,38
7	125	-1,18	1,38
8	125	-1,18	1,38
9	125	-1,18	1,38
10	125	-1,18	1,38
11	125	-1,18	1,38
12 Казаков	130	3,82	14,62
13	130	3,82	14,62
14	130	3,82	14,62
15 Литвинов	135	8,82	77,85
16	140	13,82	191,09
17	145	18,82	354,33
Σ=	2145	0,00	1126,47

$$n=17$$

$$M_2 = \underline{126,2}$$

$$m_2 = \underline{2,04}$$

$$\text{Сигма } (\sigma) = 8,390734$$

Результат выполнения команды «Анализ данных» -> «Описательная статистика».	
<i>Столбец1</i>	
	<b>126,176470</b>
<b>Среднее</b>	<b>6</b>
	<b>2,03505183</b>
<b>Стандартная ошибка</b>	<b>9</b>
<b>Медиана</b>	<b>125</b>
<b>Мода</b>	<b>125</b>
<b>Стандартное отклонение</b>	<b>8,39073368</b>
	<b>5</b>
	70,4044117
Дисперсия выборки	6
Эксцесс	0,3573165
	0,65906347
Асимметричность	6
Интервал	30
Минимум	115
Максимум	145
Сумма	2145
Счет	17
Наибольший(1)	145
Наименьший(1)	115
Уровень	4,31411620
надежности(95,0%)	3

Вычисление коэффициента вариации:  $C=8,39/126,2*100=6,6\%$

**Вывод: Средняя частота пульса пациентов 2-го отделения с вероятностью 95,5% составляет  $126,2 \pm 2,04$  ударов в минуту, вариабельность малая.**

б) Проверка на нормальность распределения признака производится на основе сгруппированного вариационного ряда (таблица 8), который создается в новой таблице, как продемонстрировано в предыдущем примере. В этой таблице указывают величину вариант и число их повторений (частоту,  $p$ ), сортируют по возрастанию (ранжируют), а затем определяют моду и медиану. Затем изготавливается точечная диаграмма распределения признака (рис. 10).

Таблица 8. Сгруппированный, ранжированный вариационный ряд

	$V$	$p$
1	115	3
2	120	2
3	<b>125</b>	<b>6</b>
4	130	3
5	135	1
6	140	1
7	145	1
	$\Sigma p=$	17

$n=17$

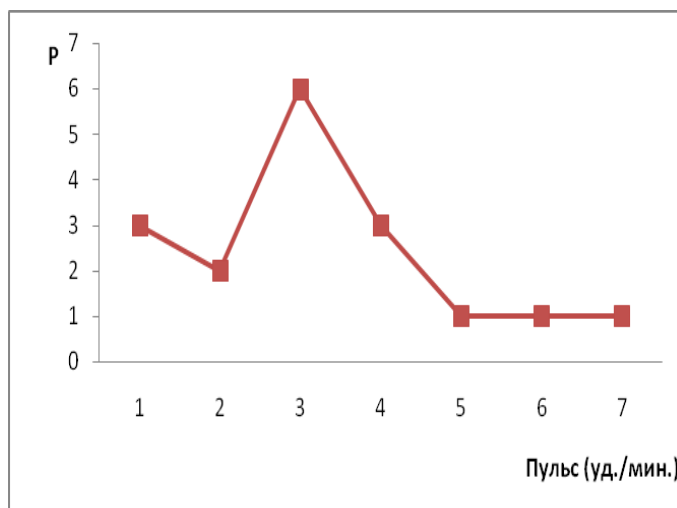


Рис. 10. График распределения признака.

$Me$  = варианта, занимающая срединное положение  
=МЕДИАНА(Диапазон) = 125 уд/мин.

$Mo$  = наиболее часто повторяющаяся варианта  
=МОДА(Диапазон) = 125 уд/мин.

в) Параметры вариабельности определяются по результатам обработки данных модулем «Описательная статистика» (таблица 7). Амплитуда (интервал) = 30 уд/мин, размах от 115 до 145 уд/мин, среднее квадратическое отклонение = 8,4 уд/мин, коэффициент вариации = **6,6%**, ошибка репрезентативности =  $\pm 2,04$  уд/мин.

г) Сравнение рассеяния вариационных рядов, используемых в примерах данного раздела выполняется на основе коэффициента вариации первого ( $C_1=5,4\%$ ) и второго вариационных рядов ( $C_2=6,6\%$ ).

**Вывод: вариабельность пульса пациентов 1-го отделения ниже, чем пациентов 2-го отделения больницы и в обоих случаях малая ( $<10\%$ ).**