# Раздел. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

## Тема 2. Интервальные оценки параметров распределения

# План лекции

**&1. Интервальные статистические распределения выборки**

**&2. Доверительный интервал**

**&3. Другие характеристики вариационного ряда**

# &1. Интервальные статистические распределения выборки

**В случае непрерывных случайных величин рассматривают *интервальные* статистические распределения выборки, которые оформляются в виде следующей таблицы:**

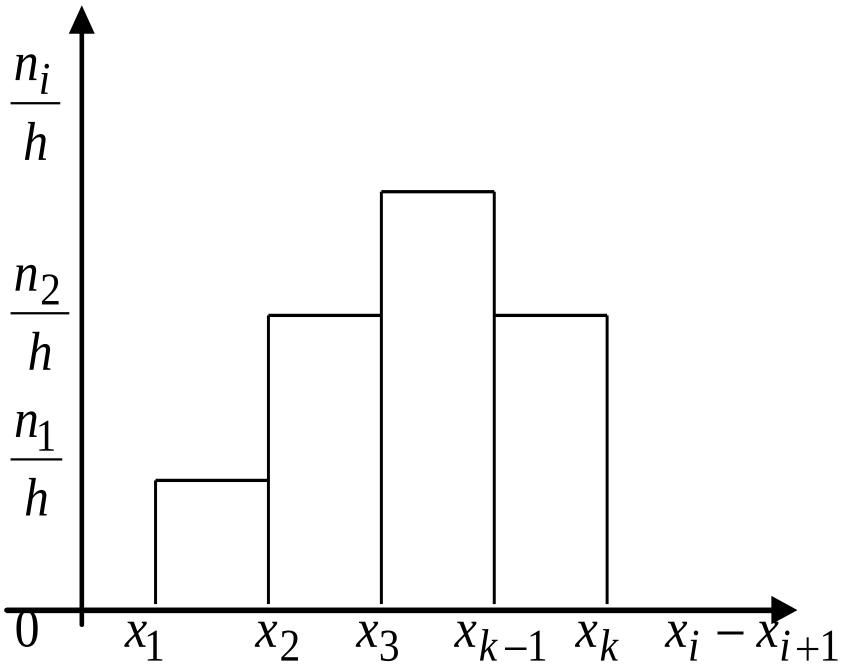
| $(x_i; x_{i+1})$ | $(x_1; x_2)$ | $(x_2; x_3)$ | $\ldots$ | $(x_{k-1}; x_k)$ |
|---|---|---|---|---|
| $n_i$ | $n_1$ | $n_2$ | $\ldots$ | $n_{k-1}$ |
| $W_i$ | $W_1$ | $W_2$ | $\ldots$ | $W_{k-1}$ |

$$\sum_{i=1}^{k} n_i = n; \quad \sum_{i=1}^{k} W_i = 1.$$

**Разница между двумя соседними вариантами называется *шагом* интервала**

$$h = x_i - x_{i+1}$$

**Графической характеристикой интервальных распределений является гистограмма частот (гистограмма относительных частот).**



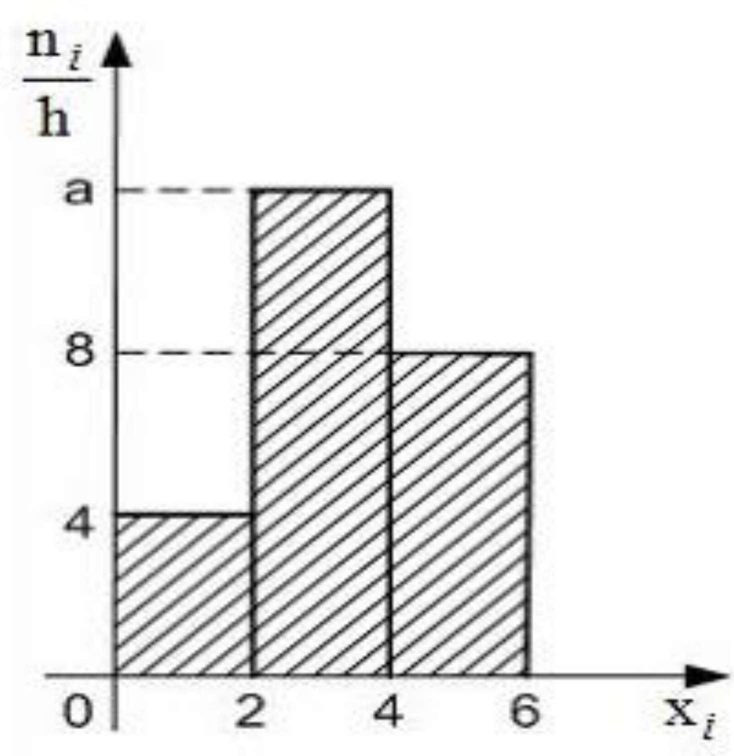
# ЗАМЕЧАНИЕ

**Для построения эмпирической функции распределения и нахождения точечных оценок ряда необходимо преобразовать его к дискретному виду по формуле**

$$\boxed{x_i^* = \frac{x_i + x_{i+1}}{2}} \qquad (3)$$

## Задача №1

**Из генеральной совокупности извлечена выборка объема *n*=50, гистограмма частот которой имеет вид:**




**Найти значение а.**

# РЕШЕНИЕ

$$n = (a + 8 + 4) \cdot h$$, то

$$a = \frac{50}{2} - 8 - 4 = 13$$.

**Задача 2**

Дана интервальная оценка (8,45; 9,15) математического ожидания нормально распределенного количественного признака. Найти точечную оценку математического ожидания.

# РЕШЕНИЕ

**Интервальная оценка математического ожидания нормального распределенного количественного признака представляет собой интервал, симметричный относительно точечной оценки. Тогда точечная оценка будет равна**

$$\frac{8{,}45+9{,}15}{2}=8{,}8$$

## &2. <u>Доверительный интервал</u>

***Интервальной*** называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Пусть найденная по данным выборки статистическая характеристика $\theta^*$ служит оценкой неизвестного параметра $\theta$. Ясно, что $\theta^*$ тем точнее определяет параметр $\theta$, чем больше абсолютная величина разности $\left|\theta-\theta^*\right|$. Другими словами, если $\delta>0$ и $\left|\theta-\theta^*\right|<\delta$, то чем меньше $\delta$, тем оценка точнее. Таким образом, положительное число $\delta$ характеризует ***точность оценки***.

Однако статистические методы не позволяют категорически утверждать, что оценка $\theta^*$ удовлетворяет неравенству $\left|\theta-\theta^*\right|<\delta$; можно лишь говорить о вероятности $\gamma$, с которой это неравенство осуществляется.

***Надежностью ( доверительной вероятностью)*** оценки $\theta$ по $\theta^*$ называют вероятность $\gamma$, с которой осуществляется неравенство $\left|\theta-\theta^*\right|<\delta$.

Обычно надежность оценки задается наперед, причем в качестве $\gamma$ берут число, близкое к единице. Наиболее часто задают надежность, равную 0,95; 0,99 и 0,999.

Пусть вероятность того, что $\left|\theta-\theta^*\right|<\delta$, равна $\gamma$

$$P\left(\left|\theta-\theta^*\right|<\delta\right)=\gamma.$$

Заменив неравенство $\left|\theta-\theta^*\right|<\delta$ равносильным ему двойным неравенством $-\delta<\theta-\theta^*<\delta$ или $\theta^*-\delta<\theta<\theta^*+\delta$, имеем

$$P\left(\theta^*-\delta<\theta<\theta^*+\delta\right)=\gamma.$$

Это соотношение следует понимать так: вероятность того, что интервал $\left(\theta^*-\delta;\theta^*+\delta\right)$ заключает в себе (покрывает) неизвестный параметр $\theta$, равно $\gamma$.

***Доверительным*** называют интервал $\left(\theta^*-\delta;\theta^*+\delta\right)$, который покрывает неизвестный параметр с заданной надежностью $\gamma$.

## Задача 3

**Дана интервальная оценка (10,45; 11,55) математического ожидания нормально распределенного количественного признака.**

**Найти точность этой оценки.**

## РЕШЕНИЕ

**Точность интервальной оценки (*a*, *b*) определяется как**

$$\boldsymbol{\delta = \frac{b-a}{2} = \frac{11{,}55-10{,}45}{2} = 0{,}55}$$

## Доверительный интервал для оценки математического ожидания нормального распределения при известном $\sigma$

Приняв во внимание, что вероятность $P$ задана и равна $\gamma$, окончательно имеем

$$P\left(\bar{x}-\frac{t\sigma}{\sqrt{n}}<a<\bar{x}+\frac{t\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma.$$

Таким образом, с надежностью $\gamma$ можно утверждать, что доверительный интервал $\left(\bar{x}-\frac{t\sigma}{\sqrt{n}};\bar{x}+\frac{t\sigma}{\sqrt{n}}\right)$ покрывает неизвестный параметр $a$; точность оценки $t=\frac{\delta\sqrt{n}}{\sigma}$.

**Замечание.** Число $t$ определяется из равенства $2\Phi(t)=\gamma$ или $\Phi(t)=\frac{\gamma}{2}$; по таблице функций Лапласа (см. приложение 3) находят аргумент $t$, которому соответствует значение функции Лапласа, равное $\frac{\gamma}{2}$.

## Доверительный интервал для оценки математического ожидания нормального распределения при неизвестном $\sigma$

доверительный интервал $\left(\bar{x} - t_\gamma \frac{s}{\sqrt{n}}; \bar{x} + t_\gamma \frac{s}{\sqrt{n}}\right)$, покрывающий неизвестный параметр $a$ с надежностью $\gamma$.

## Доверительные интервалы для оценки среднего квадратического отклонения $\sigma$ нормального распределения

$$s(1-q) < \sigma < s(1+q).$$

Параметр $q$ определяется соответствующими $n$ и $\gamma$ по приложению 5.

**Пример.** 4 Количественный признак $X$ генеральной совокупности распределен нормально. По выборке объема $n$=25 найдено «исправленное» среднее квадратическое отклонение $s$=0,8. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение $\sigma$ с надежностью 0,95.

Решение. По таблице приложения 5 по данным $\gamma = 0{,}95$ и $n = 25$ найдем $q = 0{,}32$.

Искомый доверительный интервал таков:

$$0{,}8(1-0{,}32)<\sigma<0{,}8(1+0{,}32) \text{ или } 0{,}544<\sigma<1{,}056.$$

**Замечание.** Выше предполагалось, что $q$<1. Если $q$>1, то неравенство примет вид (учитывая, что $\sigma$>0)

$$0 < \sigma < s(1+q).$$

# <u>Задача 5</u>

Выборочно обследование 30 предприятий машиностроительной промышленности по валовой продукции и получены следующие данные, в млн. руб.:

18,0; 12,0; 11,9; 1,9; 5,5; 14,6; 4,8; 5,6; 4,8; 10,9; 9,7; 7,2; 12,4; 7,6; 9,7; 11,2; 4,2; 4,9; 9,6; 3,2; 8,6; 4,6; 6,7; 8,4; 6,8; 6,9; 17,9; 9,6; 14,8; 15,8.

Составить интервальное распределение выборки с началом $x_0 = 1$ и длиной частичного интервала $h = 3$. Построить гистограмму частот.

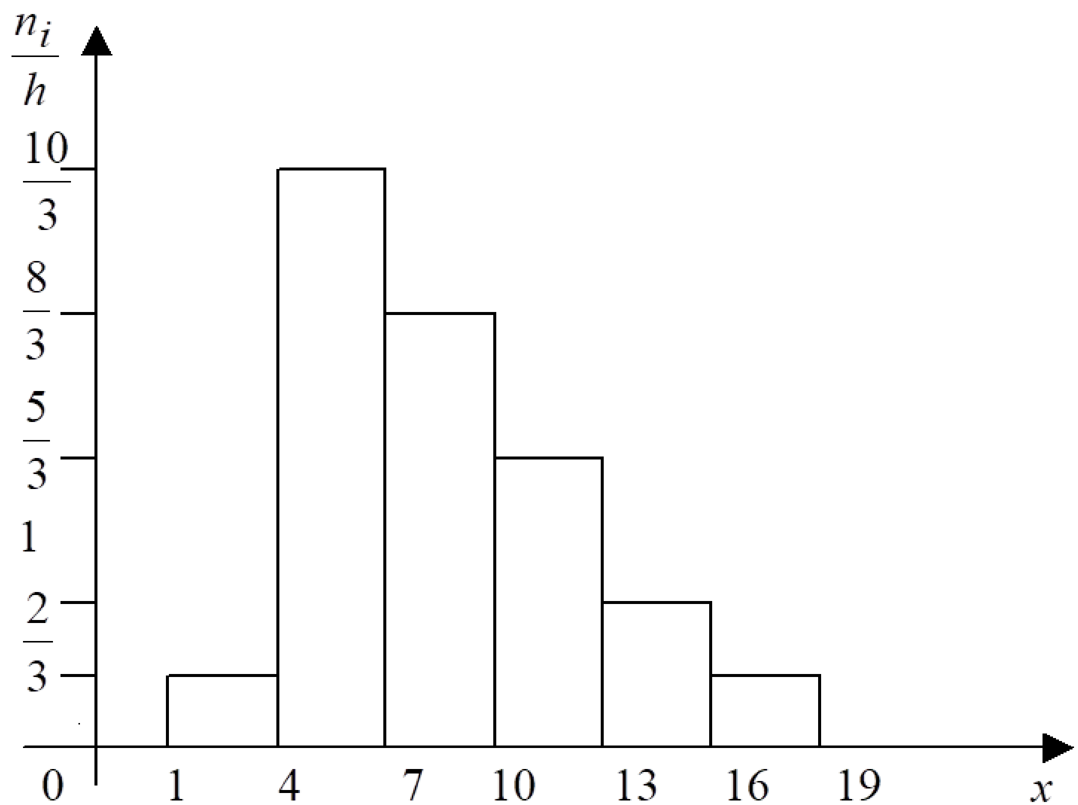
# РЕШЕНИЕ

Для составления интервального распределения составим таблицу, в первой строке которой расположим в порядке возрастания интервалы, длина каждого из которых $h = 3$. Во второй строке запишем количество значений признака в выборке, попавших в этот интервал (т.е. сумму частот вариант, попавших в соответствующий интервал):

| $(x_i; x_{i+1})$ | 1-4 | 4-7 | 7-10 | 10-13 | 13-16 | 16-19 |
|---|---|---|---|---|---|---|
| $n_i$ | 2 | 10 | 8 | 5 | 3 | 2 |

Объем выборки $n$ = 2 + 10 + 8 + 5 + 3 + 2 = 30.

Для построения гистограммы частот на оси абсцисс откладываем частичные интервалы, на каждом из них строим прямоугольники высотой $\frac{n_i}{h}$, где $n_i$ – частота $i$-го частичного интервала, $h$ – шаг (длина интервала), таким образом, гистограмма примет вид:



**Указание.** Для построения эмпирической функции распределения и нахождения точечных оценок ряда необходимо преобразовать его к дискретному виду по формуле

$$x_i^* = \frac{x_i + x_{i+1}}{2}.$$

Получим

| $x_i^*$ | 2,5 | 5,5 | 8,5 | 11,5 | 14,5 | 17,5 |
|---|---|---|---|---|---|---|
| $n_i$ | 2 | 10 | 8 | 5 | 3 | 2 |

# Задача 6

Из большой партии электроламп случайным образом отобрано 100. Средняя продолжительность горения ламп в выборке оказалась равной 1000 ч. Найти с надежностью $\gamma = 0{,}95$ доверительный интервал для средней продолжительности $a$ горения ламп во всей партии, если известно, что среднее квадратическое отклонение продолжительности горения лампы $\sigma = 40$ ч и продолжительность горения ламп распределена по нормальному закону.

Решение. По условию $\overline{x}_B = 1000,\ \gamma = 0{,}95,\ \sigma = 40.$ Для решения воспользуемся формулой

$$\overline{x}_B - \frac{\sigma t}{\sqrt{n}} < a < \overline{x}_B + \frac{\sigma t}{\sqrt{n}}$$

По приложению 3 находим $t$ из условия:

$$\Phi(t) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475; \Rightarrow t = 1{,}96.$$

Тогда доверительный интервал:

$$1000 - \frac{1{,}96 \cdot 40}{\sqrt{100}} < a < 1000 + \frac{1{,}96 \cdot 40}{\sqrt{100}}$$

$$992{,}16 < a < 1007{,}84.$$

# &3. Другие характеристики вариационного ряда

**Определение_1**

***Модой*** $M_0$ называют варианту,

**которая имеет наибольшую частоту**

## Задача 7

# Найти моду вариационного ряда

а) 1, 2, 2, 2, 3, 4, 5, 6.

б) 4, 4, 5, 6, 6 ,7 , 9

# РЕШЕНИЕ

**а) Такой вариантой является варианта 2, частота которой равна трем.**

**б) Ответ: 4 и 6**

# Определение_2

***Медианой*** $m_e$ называют варианту, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно, т.е. $n = 2k + 1$, то $m_e = x_{k+1}$; при четном $n = 2k$ медиана

$$m_e = \frac{(x_k + x_{k+1})}{2}.$$

Например, для ряда 2 3 5 6 7 медиана равна 5; для ряда 2 3 5 6 7 9 медиана равна $\frac{(5+6)}{2} = 5{,}5$.

## Задача 8

**Найти медианы вариационного ряда:**

**а) 1, 2, 2, 2, 4, 4, 5, 6;**

**б) 2, 3, 3, 4, 5, 6, 8.**

# РЕШЕНИЕ

**Медианой вариационного ряда называется варианта, расположенная в середине вариационного ряда.**

**а) Число вариант вариационного ряда четно, то $\frac{2+4}{2} = 3$**

**б) Число вариант вариационного ряда нечетно, то ответ 4.**

# Определение_3

***Размахом варьирования*** $R$ называют

разность между наибольшей и наименьшей вариантами:

$$R = x_{\max} - x_{\min}\,.$$

## <u>Задача 9</u>

**Найти размах варьирования вариационного ряда**
**1, 2, 4, 7, 10.**

# РЕШЕНИЕ

***Размахом варьирования*** $R$ называют разность между наибольшей и наименьшей вариантами:

$$\boxed{R = x_{\max} - x_{\min}\,.} \quad \textbf{(8)}$$

$$R = 10\text{-}1 = 9.$$

Размах является простейшей характеристикой рассеяния вариационного ряда.

# Определение_4

***Коэффициентом вариации*** *V* называют выраженное в процентах отношение выборочного среднего квадратического отклонения к выборочной средней:

$$V = \frac{\sigma_B}{\overline{x}_B} \cdot 100\%.$$

Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов: тот из рядов имеет большее рассеяние по отношению к выборочной средней, у которого коэффициент вариации больше. Коэффициент вариации – безразмерная величина, поэтому он пригоден для сравнения рассеяний вариационных рядов, варианты которых имеют различную размерность.

$$\boxed{\mathrm{V} = \frac{\sigma_{\text{в}}}{\bar{x}_{\text{в}}} \cdot 100\%} \quad (9)$$

# ОСНОВНЫЕ ФОРМУЛЫ

**Середины интервалов**

$$x_i^* = \frac{x_i + x_{i+1}}{2}$$

**Размах варьирования**

$$R = x_{\max} - x_{\min}.$$

**Коэффициенты вариации, асимметрии и эксцесса**

$$\mathrm{V} = \frac{\sigma_{\text{в}}}{\bar{x}_{\text{в}}} \cdot 100\%$$

**Интервальные оценки параметров распределения:**

$$\overline{X}_{в} - t\frac{\sigma}{\sqrt{n}} < a < \overline{X}_{в} + t\frac{\sigma}{\sqrt{n}};$$

$$\overline{X}_{в} - t_{\gamma}\frac{s}{\sqrt{n}} < a < \overline{X}_{в} + t_{\gamma}\frac{s}{\sqrt{n}};$$

$$s(1-q) < \sigma < s(1+q);$$

$$0 < \sigma < s(1+q);$$

# ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Какие оценки являются интервальными? В каких случаях следует пользоваться интервальной оценкой?
2. Что называется гистограммой?
3. Каким способом можно перейти от интервального статистического распределения к точечному?
4. Как найти доверительные интервалы для оценки математического ожидания нормального распределения?
5. Что называется модой?
6. Что называется медианой вариационного ряда?
7. Что называется размах?
8. Что называется коэффициентом вариации?

# ЗАДАЧИ ДЛЯ САМОКОНТРОЛЯ

**№1** Произведено выборочное обследование 25 школ. Получены следующие результаты: 42,5 60,0 63,5 70,5 82,0 83,5 92,0 95,5 100,0 101,0 105,0 108,5 110,0 115,5 120,0 120,5 122,0 130,0 138,5 140,0 142,0 150,5 160,0 162,1 180,5

$$\gamma = 0{,}96;\ \sigma_{\Gamma} = 31;\ h = 20;\ x_0 = 42{,}5.$$

Требуется:

1. Составить интервальное распределения выборки с шагом $h$, взяв за начало первого интервала $x_0$.
2. Построить гистограмму частот.
3. Найти $\overline{x}_B; D_B; \sigma_B; S$.
4. Найти с надежностью $\gamma$ доверительный интервал для оценки неизвестного математического ожидания признака $X$ генеральной совокупности, если признак $X$ распределен по нормальному закону и его среднее квадратическое отклонение равно $\sigma_{\Gamma}$.