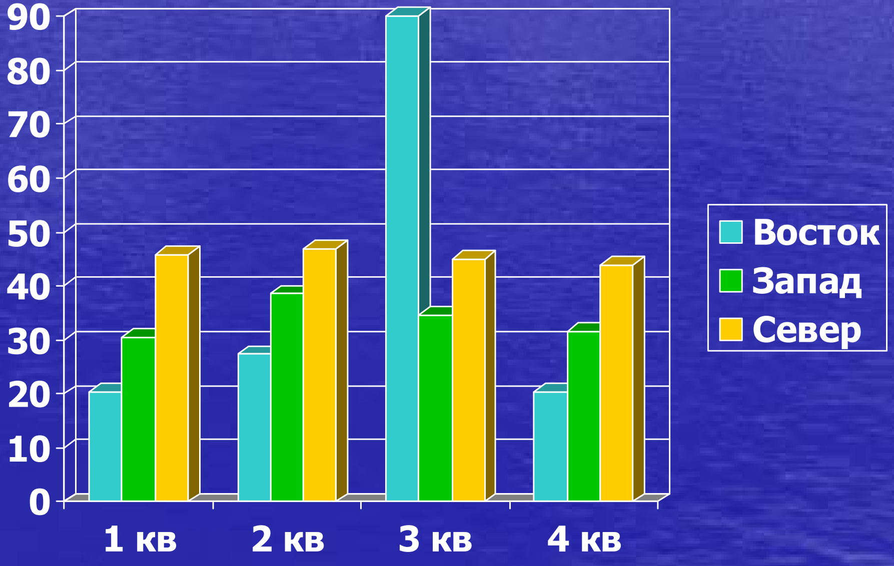
Министерство образования Республики Беларусь

УО «Витебский государственный технологический университет»

6. Элементы математической статистики.



Кафедра теоретической и прикладной математики.

разработана доц.Дуниной Е.Б.

6.1 Выборочный метод. Основные понятия.

Статистической совокупностью называется множество однородных объектов, объединенных по некоторому общему отличительному признаку.

Примеры статистических совокупностей: множество населения данной страны при исследовании ее трудовых ресурсов, множество рабочих цеха при изучении вопроса о количестве выпускаемой продукции.

Для изучения интересующего признака применяется выборочный метод.

Суть метода заключается в том, что обследованию подвергаются не все объекты совокупности, а только некоторая их часть, случайно выбранная из данной совокупности.

Выводы, полученные при изучении этой части, распространяются на всю совокупность объектов.

Генеральной совокупностью называется совокупность всех однородных объектов, подлежащих изучению.

Выборочной совокупностью, или выборкой, называется совокупность объектов, случайно отобранных из генеральной совокупности.

Объемом совокупности (генеральной или выборочной) называется число ее объектов.

Например, если из *10 000* изготовленных изделий для обследования отобрано 100, то объем генеральной совокупности *N=10 000*, объем выборки *n=100*

Выборка бывает повторной (с возвращением исследуемого объекта в генеральную совокупность) и бесповторной (без указанного возвращения).

На практике чаще используется бесповторная выборка.

Выборка должна быть *репрезентативной (представительной),* т.е. такой по которой можно уверенно судить об интересующем признаке всей генеральной совокупности.

Выборка будет репрезентативной, если ее осуществить случайным образом.

6.2 Вариационный ряд. Эмпирическая функция распределения.

При изучении некоторого признака выборочной совокупности производятся испытания.

Пусть посредством независимых испытаний, проведенных в одинаковых условиях, получены следующие числовые значения:

$$x_1, x_2, \ldots, x_n,$$

где *n*-объем выборки .

Наблюдаемые значения x_i **называют вариантами,**

а последовательность вариант, записанных в возрастающем порядке – вариационном рядом.

Пусть из генеральной совокупности извлечена выборка, причем

x_1 наблюдалось n_1 раз,

x_2 - n_2 раз,

x_k - n_k раз,

$\sum n_i = n$ - объем выборки.

Числа наблюдений называют **частотами,**

а их отношение к объему выборки

$$\frac{n_i}{n} = w_i$$

- относительными частотами.

Причем

$$\sum_{i=1}^{k} w_i = 1,$$

т.е. сумма относительных частот всех вариант равна единице.

Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот.

Пример.

Задано распределение частот выборки объема *n=20.*

x_i	2	6	12
n_i	3	10	7

Написать распределение относительных частот.

Решение.

$$w_i = \frac{n_i}{n},$$

$$w_1 = \frac{3}{20} = 0{,}15,$$

$$w_2 = \frac{10}{20} = 0{,}5,$$

$$w_3 = \frac{7}{20} = 0{,}35.$$

Напишем распределение относительных частот

x_i	2	6	12
w_i	0,15	0,5	0,35

Контроль:
0,15+0,5+0,35=1.

Эмпирической функцией распределения, или функцией распределения выборки, называется функция, определяющая для каждого значения *x* относительную частоту события *X<x*.

Эмпирическую функцию распределения обозначают $F^*(x).$

Если n_x **- число вариант, меньших *x*, *n* – объем выборки, то по определению**

$$F^*(x) = \frac{n_x}{n}. \quad (6.1)$$

В отличие от эмпирической функции распределения выборки функцию распределения *F(x)* генеральной совокупности называют теоретической функцией распределения.

Отличие состоит в том, что теоретическая функция *F(x)* определяет вероятность события *X<x,* а эмпирическая функция $F^*(x)$

определяет относительную частоту этого события.

$F^*(x)$ **обладает следующими свойствами:**

1. Значения $F^*(x)$ **принадлежат отрезку [0,1].**

2. $F^*(x)$ **- неубывающая функция.**

3. Если x_1 **- наименьшая варианта, то**

$$F^*(x) = 0 \quad \text{при} \quad x \le x_1,$$

если x_k **- наибольшая варианта, то**

$$F^*(x) = 1 \quad \text{при} \quad x > x_k.$$

Пример.

Построить эмпирическую функцию по данному распределению выборки

x_i	2	6	10
n_i	12	18	30

Решение.

Найдем объем выборки: *12+18+30=60.*

Наименьшая варианта равна 2.

Следовательно, $F^*(x)=0 \quad \text{при} \quad x\le 2.$

Значение *X<6*, а именно $x_1=2,$

наблюдалось 12 раз, следовательно

$$F^*(x)=\frac{n_x}{n},$$

$$F^*(x)=\frac{12}{60}=0{,}2 \quad \text{при} \quad 2<x\le 6.$$

Значение *X<10*, а именно $x_1=2, x_2=6$

наблюдалось *12+18=30* раз, следовательно

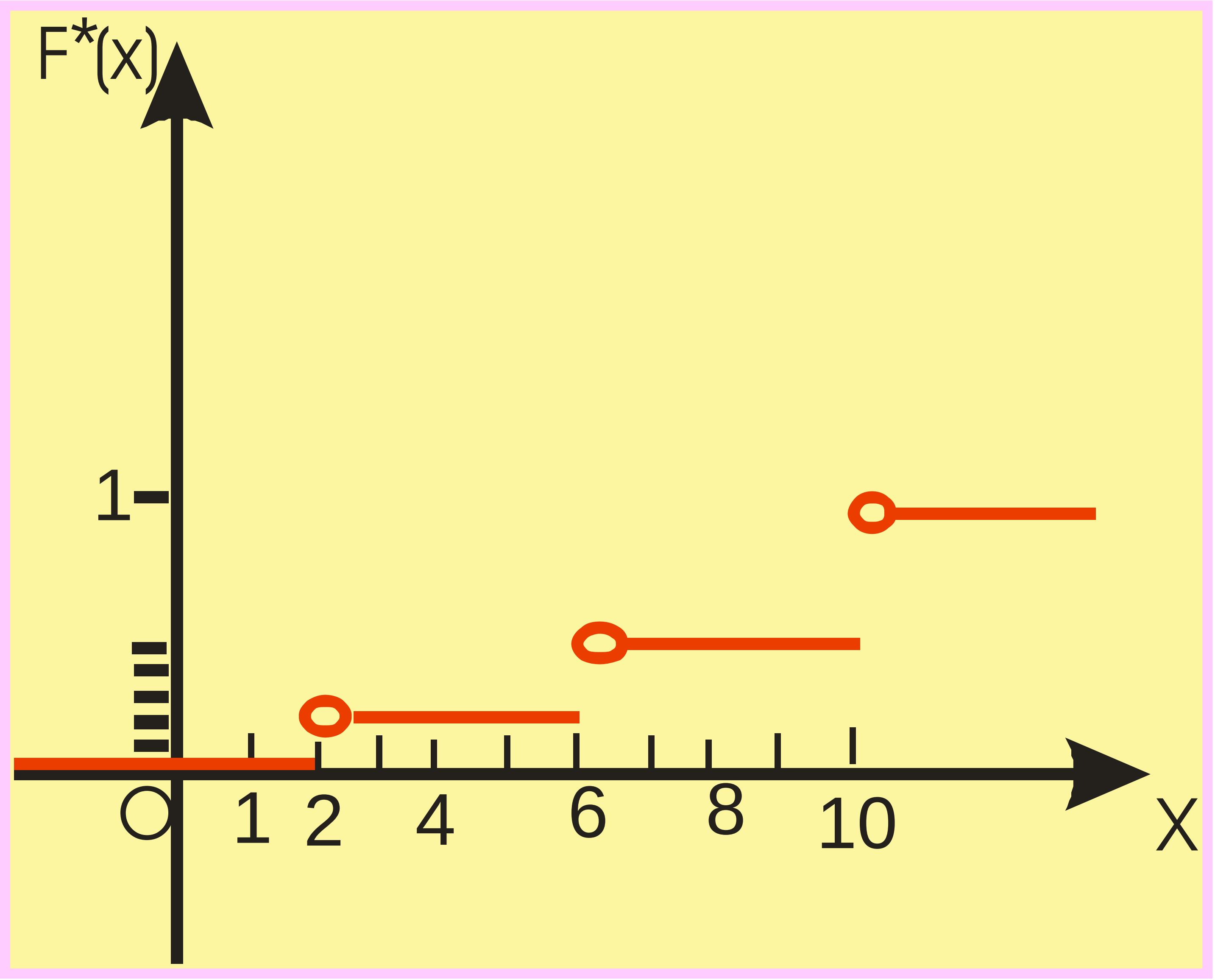
$$F^*(x) = \frac{30}{60} = 0{,}5 \quad \text{при} \quad 6 < x \le 10.$$

Т.к. *X=10* наибольшая варианта, то

$$F^*(x) = 1 \quad \text{при} \quad x > 10.$$

Таким образом,

$$F^*(x) = \begin{cases} 0, & x \le 2 \\ 0{,}2, & 2 < x \le 6 \\ 0{,}5, & 6 < x \le 10 \\ 1, & x > 10. \end{cases}$$


F*(x)
1
O
1
2
4
6
8
10
X

6.3 Полигон и гистограмма.

В целях наглядности статистическое распределение изображается графически.

Полигоном частот называют ломаную, отрезки которой соединяют точки

$$(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k).$$

Для построения полигона частот на оси абсцисс откладывают варианты x_i,

а на оси ординат – соответствующие им частоты n_i

Точки (x_i, n_i) соединяют отрезками прямых и получают полигон частот.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки

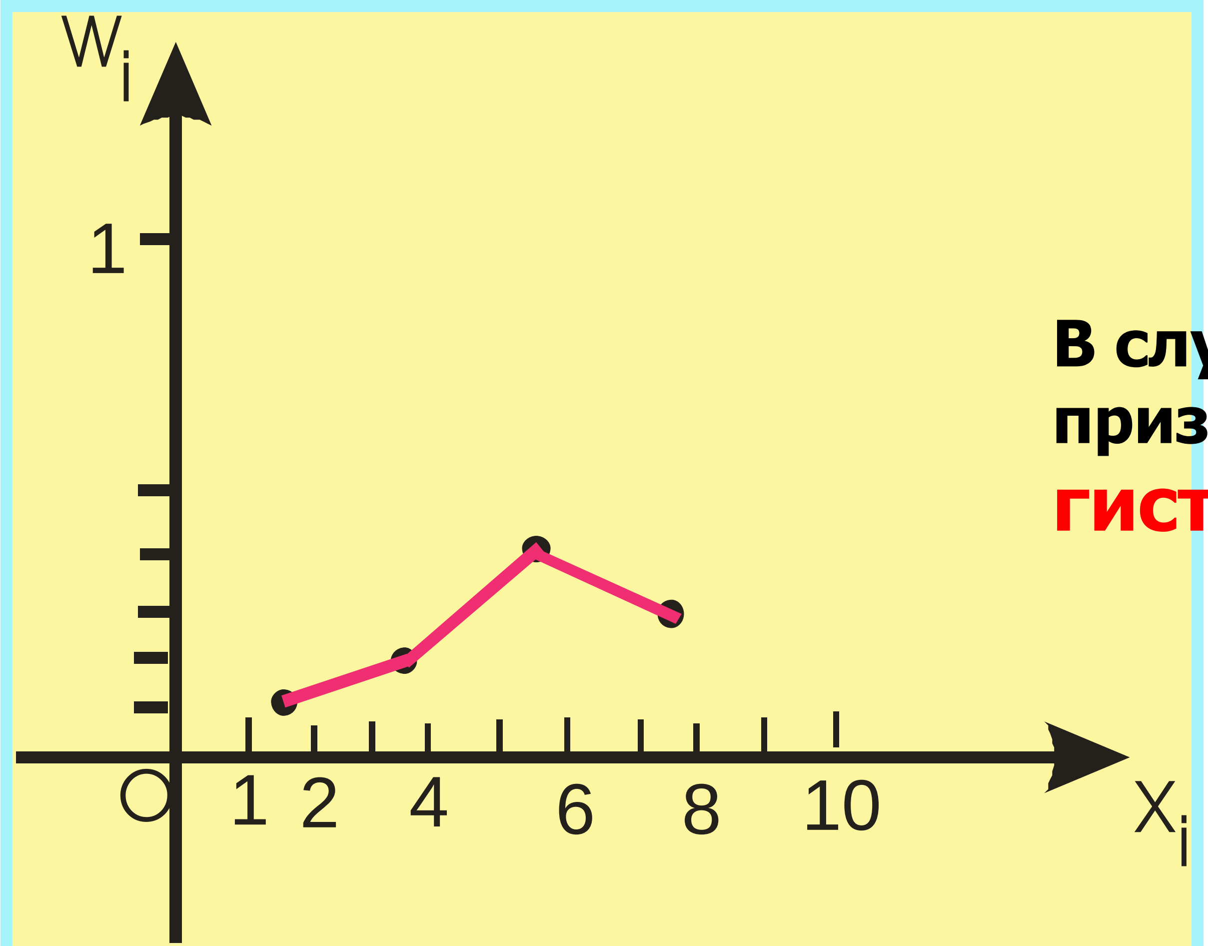
$$(x_1, w_1), (x_2, w_2), \ldots, (x_k, w_k).$$

Пример.

Построить полигон относительных частот следующего распределения

x_i	1,5	3,5	5,5	7,5
w_i	0,1	0,2	0,4	0,3

Решение.



В случае непрерывного признака строят гистограмму.

Для этого интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько частичных интервалов длиною *h* и находят для каждого частичного интервала n_i **- сумму частот вариант, попавших в *i* интервал.**

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению

$$\frac{n_i}{h}.$$

Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии

$$\frac{n_i}{h}.$$

Площадь i-прямоугольника равна

$$\frac{n_i}{h} \cdot h = n_i$$

- сумме частот вариант i-го интервала, следовательно, площадь гистограммы частот равна сумме всех частот, т.е. объему выборки.

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною *h*, а высоты равны отношению

$$\frac{w_i}{h}.$$

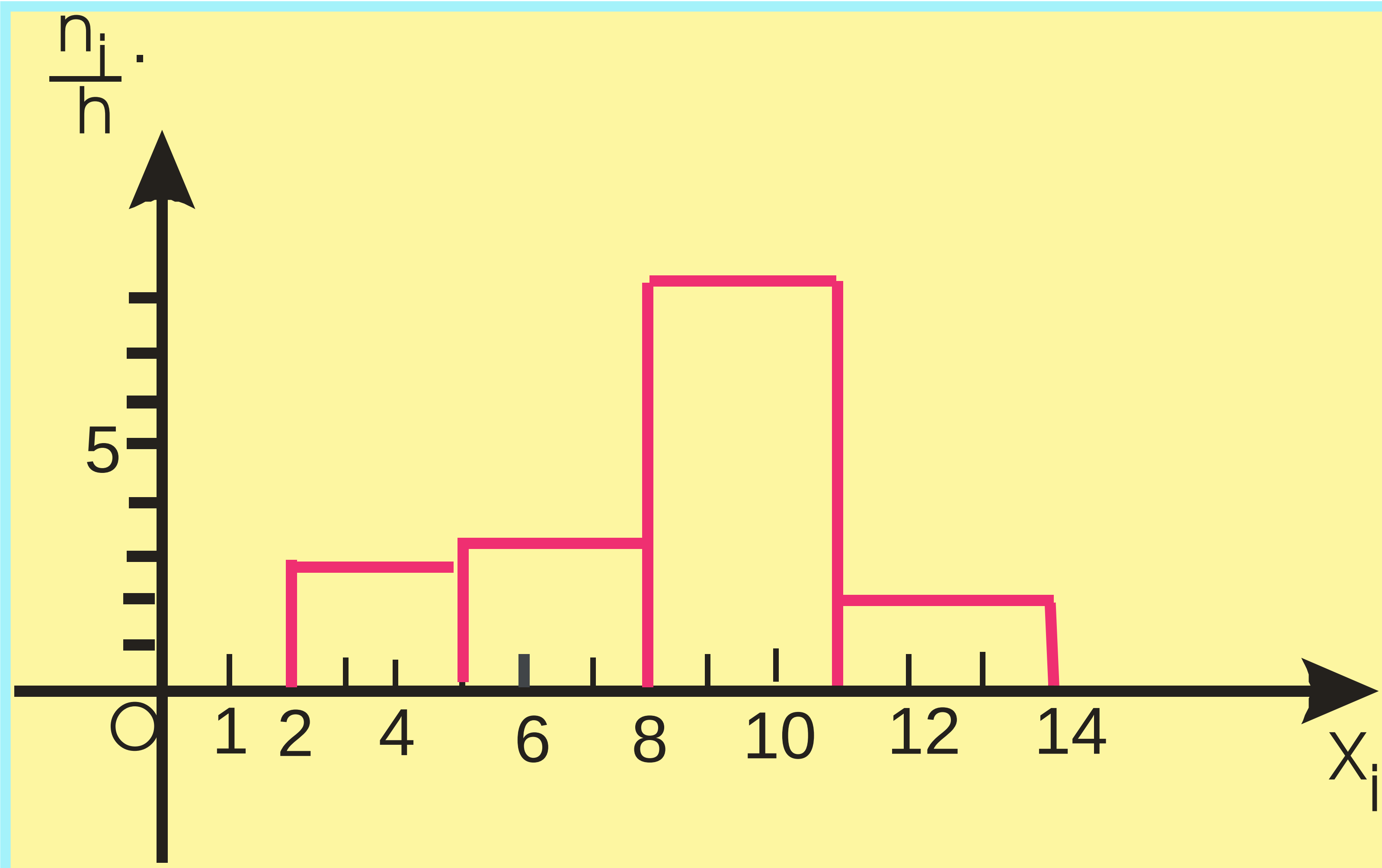
Площадь гистограммы относительных частот равна сумме всех относительных частот, т.е. единице.

Пример.

Построить гистограмму частот

частичный интервал $h=3$	сумма частот частичного интервала n_i	$\frac{n_i}{h}$
2-5	9	3
5-8	10	3,3
8-11	25	8,3
11-14	6	2

Решение.



$\frac{n_i}{h}$
5
O
1
2
4
6
8
10
12
14
x_i

6.4 Несмещенные, эффективные и состоятельные оценки.

Обычно в распоряжении исследователя имеются лишь данные выборки, например значения количественного признака

$$x_1, x_2, \ldots, x_n$$

полученные в результате *n* наблюдений.

Наблюдения предполагаются независимыми.

Найти статистическую оценку неизвестного параметра – значит найти функцию от наблюдаемых случайных величин, которая и дает приближенное значение оцениваемого параметра.

Пусть θ^* **статистическая оценка неизвестного параметра** θ **теоретического распределения.**

Допустим, что по выборке объема *n* найдена оценка

$$\theta_1^*.$$

Повторим опыт, т.е. извлечем из генеральной совокупности другую выборку того же объема и по ее данным найдем оценку θ_2^*.

Повторяя опыт многократно, получим числа $\theta_1^*, \theta_2^*, \theta_3^*, \ldots, \theta_k^*$ **которые вообще говоря различны между собой.**

Тогда оценку θ^* можно рассматривать как случайную величину, а числа $\theta_1^*, \theta_2^*, \theta_3^*, \ldots, \theta_k^*$ как ее возможные значения.

Пусть оценка θ^* дает приближенное значение θ с избытком; тогда каждое найденное по данным выборок число $\theta_i^* \, (i = 1, \ldots, k)$ больше истинного значения $\theta.$

Ясно, что в этом случае и математическое ожидание (среднее значение) случайной величины θ^* больше чем θ

$$\boxed{M(\theta^*) > \theta}$$

Если θ^* дает оценку с недостатком, то

$$\boxed{M(\theta^*) < \theta.}$$

Таким образом, использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, привело бы к ***систематическим ошибкам.***

Несмещенной называют статистичекую оценку θ^*, математичекое ожидание которой равно оцениваемому параметру θ при любом объеме выборки, т.е.

$$\boxed{M(\theta^*) = \theta.}$$

Смещенной называют оценку математическое ожидание которой не равно оцениваемому параметру.

Кроме этого возможные значения θ^*

могут быть сильно рассеяны вокруг среднего значения θ^*, т.е. дисперсия $D(\theta^*)$

может быть значительной.

Эффективной называют статистическую оценку, которая имеет наименьшую возможную дисперсию

$$\boxed{D(\theta^*) = D_{\min}}$$

Состоятельной называют статистическую оценку, которая при $n \to \infty$

стремится по вероятности к оцениваемому параметру.

6.5 Генеральная и выборочная средняя.

Генеральной средней $\bar{x}_{\text{г}}$

называют среднее арифметическое значений признака генеральной совокупности.

Если все значения $x_1, x_2, \ldots, x_N$ **признака**

генеральной совокупности объема *N* различны, то

$$\bar{x}_{\Gamma} = \frac{x_1 + x_2 + \ldots + x_N}{N}.$$

Если же значения признака $x_1, x_2, \ldots, x_k$

имеют соответственно частоты $N_1, N_2, \ldots, N_k$,

причем $N_1 + N_2 + \ldots + N_k = N$, **то**

$$\bar{x}_{\text{г}} = \frac{x_1 N_1 + x_2 N_2 + \ldots + x_k N_k}{N}.$$

Пусть генеральная совокупность объема *N* содержит объекты с различными значениями признака *X*, равными

$$x_1, x_2, \ldots, x_N.$$

Вероятность того, что будет извлечен объект со значением признака x_1 **равна** $\frac{1}{N}$.

С этой же вероятностью будет извлечен и любой другой объект.

Найдем математическое ожидание

$$M(X) = x_1 \frac{1}{N} + x_2 \frac{1}{N} + \ldots + x_k \frac{1}{N} = \bar{x}_Г \qquad (6.2)$$

Выборочной средней $\bar{x}_{\text{В}}$ **называют среднее арифметическое значение признака выборочной совокупности.**

Если все значения $x_1, x_2, \ldots, x_n$ признака выборки объема *n* различны, то

$$\bar{x}_{\text{В}} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Если же значения признака $x_1, x_2, \ldots, x_k$ имеют соответственно частоты $n_1, n_2, \ldots, n_k,$

причем $n_1 + n_2 + \ldots + n_k = n,$ то

$$\bar{x}_{\text{В}} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n} = \frac{\left(\sum_{i=1}^{k} n_i x_i\right)}{n}.$$

Выборочную среднюю принимают в качестве оценки генеральной средней.

Т.е. несмещенной оценкой генеральной средней является выборочная средняя

$$M(\bar{x}_{\text{В}}) = \bar{x}_{Г}$$

Пример.

Из генеральной совокупности извлечена выборка объема *n=50*

x_i	2	5	7	10
n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

Решение.

Несмещенной оценкой генеральной средней является выборочная средняя

$$\overline{x}_{\text{В}} = \frac{\left(\sum_{i=1}^{k} n_i x_i\right)}{n} = .\frac{16\cdot 2 + 12\cdot 5 + 8\cdot 7 + 14\cdot 10}{50} = 5{,}76$$

6.6 Генеральная и выборочная дисперсия

Для характеристики рассеяния количественного признака *X* генеральной совокупности вокруг своего среднего значения служат понятия генеральной дисперсии и генерального среднего квадратического отклонения.

Генеральной дисперсией $D_Г$

называют среднее квадратическое квадратов отклонений значений признака генеральной совокупности от их среднего значения $\overline{x}_г$.

Если все значения $x_1, x_2, \ldots, x_N$ **признака генеральной совокупности объема *N* различны, то**

$$\boxed{D_Г = \frac{\sum\limits_{i=1}^{N}(x_i - \overline{x}_Г)^2}{N}. \qquad (6.3)}$$

Если же значения признака $x_1, x_2, \ldots, x_k$

имеют соответственно частоты $N_1, N_2, \ldots, N_k$, **то**

$$D_Г = \frac{\sum_{i=1}^{k} N_i (x_i - \bar{x}_Г)^2}{N}. \qquad (6.4)$$

Генеральным средним квадратическим отклонением называется корень квадратный из генеральной дисперсии

$$\sigma_Г = \sqrt{D_Г}. \qquad (6.5)$$

Выборочной дисперсией D_B

называют среднее квадратическое квадратов отклонения наблюдаемых значений выборки от их среднего значения $\bar{x}_B$.

Если все значения $x_1, x_2, \ldots, x_n$

признака выборки объема *n* различны, то

$$D_B = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_B)^2}{n}. \quad (6.6)$$

Если же значения признака $x_1, x_2, \ldots, x_k$

имеют соответственно частоты $n_1, n_2, \ldots, n_k,$

то

$$D_B = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n}. \quad (6.7)$$

Выборочным средним квадратическим отклонением называется корень квадратный из выборочной дисперсии

$$\sigma_B = \sqrt{D_B}. \qquad (6.8)$$

Для вычисления выборочной дисперсии можно пользоваться формулой

$$D_B = \overline{x_B^{\,2}} - \left[\overline{x}_B\right]^2, \quad (6.9)$$

где

$$\overline{x}_{\text{В}} = \frac{\sum_{i=1}^{k} n_i x_i}{n},$$

$$\overline{x_B^{\,2}} = \frac{\sum_{i=1}^{k} n_i x_i^2}{n}$$

Выборочная дисперсия D_B

является смещенной оценкой генеральной дисперсии $D_Г,$ **т.к.**

$$M(D_B) = \frac{n-1}{n} D_Г,$$ **т.е.** $$M(D_B) \neq D_Г.$$

Чтобы получить несмещенную оценку генеральной дисперсии, вводят понятие эмпирической (или исправленной) дисперсии s^2.

Эмпирическая, или исправленная, дисперсия определяется формулой

$$s^2 = \frac{n}{n-1} D_B = \frac{n}{n-1} \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n}, \quad (6.10)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2. \quad (6.11)$$

В случае, когда все значения $x_1, x_2, \ldots, x_n$

различны, т.е. все $n_i = 1$ и $k = n$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_B)^2. \quad (6.12)$$

Для оценки среднего квадратического отклонения генеральной совокупности служит «исправленное» среднее квадратическое отклонение

$$s=\sqrt{s^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{k}n_i(x_i-\bar{x}_B)^2}. \qquad (6.13)$$

В случае, когда значения $x_1, x_2, \ldots, x_n$ **различны**

$$s=\sqrt{s^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_B)^2}. \qquad (6.14)$$

Пример.

В итоге пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): 92; 94; 103; 105; 106.

Найти: а) выборочную среднюю длину стержня; б) выборочную и исправленную дисперсии ошибок прибора.

Решение.

а) Найдем выборочную среднюю

$$\overline{x}_{\text{В}} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{92+94+103+105+106}{5} = 100.$$

б)Найдем выборочную дисперсию, для этого воспользуемся формулой (6.6)

$$D_B = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_B)^2}{n} =$$

$$= \frac{(92-100)^2 + (94-100)^2 + (103-100)^2 + (105-100)^2 + (106-100)^2}{5} = 34.$$

Выборочную дисперсию можно также найти по формуле (6.9)

$$D_B = \overline{x_B^{\,2}} - \left[\overline{x}_B\right]^2$$

$$\overline{x}_{\text{В}} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = 100,$$

$$\overline{x_B^2} = \frac{\sum_{i=1}^{k} n_i x_i^2}{n} = \frac{92^2 + 94^2 + 103^2 + 105^2 + 106^2}{5} = 10034$$

$$D_B = \overline{x_B^2} - \left[\overline{x}_B\right]^2 = 10034 - 10000 = 34.$$

Найдем исправленную дисперсию

$$s^2 = \frac{n}{n-1} D_B = \frac{5}{4} \cdot 34 = 42{,}5.$$

6.7 Интервальные оценки. Доверительная вероятность (надежность). Доверительный интервал.

Точечной называют оценку, которая определяется одним числом.

Интервальной называют оценку, которая определяется двумя числами – концами интервала.

Пусть найденная по данным выборки статистическая характеристика θ^*

служит оценкой неизвестного параметра θ.

Ясно, что θ^* тем точнее определяет параметр θ,

чем меньше абсолютная величина разности $\left|\theta - \theta^*\right|$.

Другими словами, если $\delta > 0$ и $\left|\theta - \theta^*\right| < \delta$,

то чем меньше δ, тем точнее оценка.

Положительное число δ характеризует точность оценки.

Надёжностью (доверительной вероятностью) оценки θ^* параметра θ называют вероятность γ,

с которой осуществляется неравенство $\left|\theta - \theta^*\right| < \delta$

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

Обычно надежность γ задается заранее, в качестве γ берут число, близкое к единице.

Заменим неравенство $\left|\theta - \theta^*\right| < \delta$

равносильным ему двойным неравенством

$$-\delta < \theta - \theta^* < \delta \text{ или } \theta^* - \delta < \theta < \theta^* + \delta.$$

Тогда

$$\boxed{P[\theta^* - \delta < \theta < \theta^* + \delta] = \gamma,}$$

т.е. вероятность того, что интервал $(\theta^* - \delta, \theta^* + \delta)$ **заключает в себе неизвестный параметр** θ **равна** $\gamma.$

Доверительным называют интервал $(\theta^* - \delta, \theta^* + \delta),$

который покрывает неизвестный параметр с заданной надежностью.

6.8 Доверительные интервалы для оценки математического ожидания нормального распределения.

Пусть количественный признак *X* генеральной совокупности распределен нормально, причем среднее квадратическое отклонение σ **известно.**

Требуется оценить математическое ожидание a **по выборочной средней.**

Выборочная средняя $\overline{X}_B$ **изменяется от выборки к выборке,**

ее можно рассматривать как случайную величину $\overline{X}$

Будем полагать, если случайная величина $\overline{X}$

распределена нормально, то выборочная средняя также распределена нормально.

Мы получали, что математическое ожидание среднего арифметического одинаково распределенных взаимно независимых случайных величин равно математическому ожиданию a

каждой из величин

$$M(\overline{X}) = a, \quad \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

Потребуем, чтобы выполнялось соотношение

$$P\left(\left|\overline{X}-a\right|<\delta\right)=\gamma$$

γ **– заданная надежность.**

Пользуясь формулой $P(|X-a|<\delta)=2\Phi\left(\dfrac{\delta}{\sigma}\right),$

заменив ***X*** **на** $\overline{X}$ **и** σ **на** $\sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$

получим

$$P(|\overline{X}-a|<\delta)=2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)=2\Phi(\mathrm{t}),$$

где $\mathrm{t}=\dfrac{\delta\sqrt{n}}{\sigma}$ или $\delta=\dfrac{\sigma t}{\sqrt{n}}.$

Формулу можно записать в виде

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

Учитывая, что

$$\left|\bar{X}-a\right|<\frac{\sigma t}{\sqrt{n}},$$

$$-\frac{\sigma t}{\sqrt{n}}<\bar{X}-a<\frac{\sigma t}{\sqrt{n}},$$

$$-\bar{X}-\frac{\sigma t}{\sqrt{n}}<-a<-\bar{X}+\frac{\sigma t}{\sqrt{n}},$$

$$\bar{X}-\frac{\sigma t}{\sqrt{n}}<a<\bar{X}+\frac{\sigma t}{\sqrt{n}},$$

$$P\left(\overline{X}_B - \frac{\sigma t}{\sqrt{n}} < a < \overline{X}_B + \frac{\sigma t}{\sqrt{n}}\right) = 2\Phi(\mathrm{t}) = \gamma \qquad (6.15)$$

Смысл полученного соотношения таков:

с надежностью γ **можно утверждать, что**

доверительный интервал $\left(\overline{X}_B - \frac{\sigma t}{\sqrt{n}}, \overline{X}_B + \frac{\sigma t}{\sqrt{n}}\right)$

покрывает неизвестный параметр a,

точность оценки $\delta = \frac{t\sigma}{\sqrt{n}}.$

Число *t* определяется из равенства

$$2\Phi(t) = \gamma \quad \text{или} \quad \Phi(t) = \gamma / 2.$$

По таблице функции Лапласа находят аргумент *t*, которому соответствует значение функции Лапласа равное

$$\gamma / 2.$$

Пример.

Случайная величина *X* имеет нормальное распределение со средним квадратическим отклонением $\sigma = 3.$

Найти доверительный интервал для оценки неизвестного математического ожидания a **по выборочному среднему** $\overline{X}_B = 4{,}1,$ **если объем выборки *n=36* и задана надежность оценки**

$$\gamma = 0{,}95.$$

Решение.

Найдем *t* из соотношения

$$2\Phi(\mathrm{t}) = \gamma \quad \text{или} \quad \Phi(\mathrm{t}) = \gamma / 2.$$

$$\Phi(\mathrm{t}) = 0{,}95 / 2 = 0{,}475.$$

По таблице находим *t=1,96* .

Найдем

$$\delta = \frac{\sigma t}{\sqrt{n}} = \frac{1{,}96 \cdot 3}{\sqrt{36}} = 0{,}98.$$

Доверительный интервал

$$\left(\overline{X}_B - \frac{\sigma t}{\sqrt{n}}, \overline{X}_B + \frac{\sigma t}{\sqrt{n}} \right)$$

Вычислим $$\overline{X}_B - \frac{\sigma t}{\sqrt{n}} = 4{,}1 - 0{,}98 = 3{,}12$$

$$\overline{X}_B + \frac{\sigma t}{\sqrt{n}} = 4{,}1 + 0{,}98 = 5{,}08$$

Доверительный интервал ***3,12<*** a ***<5,08.***

Если требуется оценить неизвестное математическое ожидание a **, при неизвестном среднем квадратическом отклонении** σ,

то доверительный интервал

$$\left(\overline{X}_B - \frac{t_\gamma S}{\sqrt{n}}, \overline{X}_B + \frac{t_\gamma S}{\sqrt{n}} \right)$$

покрывает неизвестный параметр a

с надежностью γ.

Здесь ***S*** **– «исправленное» среднее квадратическое отклонение**

$$s^2 = \frac{n}{n-1} D_B,$$

значение t_γ **находят по таблице по заданным** ***n*** ***и*** γ.

Пример.

Из генеральной совокупности извлечена выборка объема ***n=10***

x_i	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

Оценить с надежностью *0,95* математическое ожидание a нормально распределенного признака генеральной совокупности по выборочной средней при помощи доверительного интервала.

Решение.

Определим выборочную среднюю по формуле

$$\overline{x}_{\text{В}} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n} = \left(\sum_{i=1}^{k} n_i x_i \right) / n.$$

$$\bar{x}_{\text{В}} = \frac{-2\cdot 2 + 1\cdot 1 + 2\cdot 2 + 3\cdot 2 + 4\cdot 2 + 5\cdot 1}{10} =$$

$$= \frac{-4 + 1 + 4 + 6 + 8 + 5}{10} = 2.$$

Определим «исправленное» среднее квадратичное отклонение по формуле (6.13)

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2},$$

$$s = \sqrt{\frac{1}{10-1}\left[2(-2-2)^2 + 1(1-2)^2 + 2(2-2)^2 + 2(3-2)^2 + 2(4-2)^2 + 1(5-2)^2\right]} =$$

$$= \sqrt{\frac{1}{9}\left[32+1+0+2+8+9\right]} = \sqrt{\frac{52}{9}} \approx 2{,}4$$

Найдем t_γ.

Пользуясь таблицей по

$$\gamma = 0{,}95 \quad \text{и} \quad n = 10$$

находим

$$t_\gamma = 2{,}26$$

Найдем искомый доверительный интервал

$$\overline{X}_B - \frac{t_\gamma S}{\sqrt{n}} < a < \overline{X}_B + \frac{t_\gamma S}{\sqrt{n}},$$

$n \backslash \gamma$	0,95	0,99	0,999
5	2,78	4,60	8,61
6	2,57	4,03	6,86
7	2,45	3,71	5,96
8	2,37	3,50	5,41
9	2,31	3,36	5,04
10	2,26	3,25	4,78
11	2,23	3,17	4,59

$$2-\frac{2{,}26\cdot 2{,}4}{\sqrt{10}}<a<2+\frac{2{,}26\cdot 2{,}4}{\sqrt{10}},$$

$$0{,}3<a<3{,}7.$$

Пример.

Пусть произведено девять независимых измерений физической величины

1.	42,1
2.	43,2
3.	42,9
4.	42,7
5.	42,8
6.	42,6
7.	43,1
8.	43,3
9.	42,5

Требуется оценить истинное значение измеряемой величины с надежностью $\gamma=0{,}95.$

Решение.

Истинное значение измеряемой величины равно ее математическому ожиданию.

Поэтому задача сводится к оценке математического ожидания (при неизвестном σ) при помощи доверительного интервала

$$\left(\overline{X}_B - \frac{t_\gamma S}{\sqrt{n}}, \overline{X}_B + \frac{t_\gamma S}{\sqrt{n}}\right),$$

покрывающего a с заданной надежностью $\gamma = 0{,}95.$

Найдем t_γ. **Пользуясь таблицей по**

$\gamma = 0{,}95$ и $n = 9$ **находим** $t_\gamma = 2{,}26$

Определим выборочную среднюю по формуле

$$\overline{X}_B = \frac{\sum_{i=1}^{k} x_i}{\text{n}}.$$

$$\overline{X}_B = \frac{42{,}1 + 43{,}2 + 42{,}9 + 42{,}7 + 42{,}8 + 42{,}6 + 43{,}1 + 43{,}3 + 42{,}5}{9} =$$

$$= \frac{385{,}2}{9} = 42{,}8.$$

Выборочную дисперсию определим по формуле

$$D_B = \frac{\sum_{i=1}^{n} (x_i - \overline{x}_B)^2}{n}.$$

Для удобства расчетов составим таблицу

№	x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
1.	42,1	0,7	0,49
2.	43,2	-0.4	0,16
3.	42,9	-0,1	0,01
4.	42,7	0,1	0,01
5.	42,8	0	0
6.	42,6	0,2	0,04
7.	43,1	-0,3	0,09
8.	43,3	-0,5	0,25
9.	42,5	0,3	0,09
Σ	385,2	0	1,14

Выборочная дисперсия

$$D_B = \frac{\sum_{i=1}^{n} (x_i - \overline{x}_B)^2}{n} =$$

$$= \frac{1{,}14}{9} = 0{,}127.$$

Исправленная дисперсия

$$s^2 = \frac{n}{n-1} D_B = \frac{9}{9-1} \cdot 0{,}127 = 0{,}143.$$

Тогда

$$s = \sqrt{0{,}143} = 0{,}378$$

Найдем искомый доверительный интервал

$$\overline{X}_B - \frac{t_\gamma S}{\sqrt{n}} < a < \overline{X}_B + \frac{t_\gamma S}{\sqrt{n}},$$

$$42{,}8 - \frac{2{,}35 \cdot 0{,}378}{\sqrt{9}} < a < 42{,}8 + \frac{2{,}35 \cdot 0{,}378}{\sqrt{9}},$$

Итак с надежностью 0,95 истинное значение измеряемой величины заключено в интервале

$$42{,}5 < a < 43{,}1$$

6.9 Другие характеристики вариационного ряда.

Модой M_0 называют варианту, которая имеет наибольшую частоту.

Например, для ряда

x_i	1	4	7	9
n_i	5	1	20	6

мода равна 7

$$M_0 = 7.$$

Медианой m_e **называют варианту, которая делит вариационный ряд на две части, равные по числу вариант.**

Например, для ряда $2 \quad 3 \quad 5 \quad 6 \quad 7$

с нечетным числом вариант медиана равна *5.*

При четном числе вариант *n=2k* медиана

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

Например, для ряда $2 \quad 3 \quad 5 \quad 6 \quad 7 \quad 9$

медиана равна

$$\frac{5+6}{2} = 5{,}5.$$

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

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

Например, для ряда $1 \quad 3 \quad 4 \quad 5 \quad 6 \quad 10$

размах равен *10-1=9.*

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 безразмерная величина.

Коэффициент вариации служит для сравнения величин рассеяния по отношению к выборочной средней двух вариационных рядов.

Ряд у которого коэффициент вариации больше, имеет большее рассеяние по отношению к выборочной средней.