

Министерство образования, науки и молодежной политики Краснодарского края
Государственное бюджетное профессиональное образовательное учреждение
Краснодарского края
«АРМАВИРСКИЙ МАШИНОСТРОИТЕЛЬНЫЙ ТЕХНИКУМ»

Учебное пособие

Элементы

математической статистики

для студентов специальности 09.02.07
«Информационные системы и программирование»

Содержание

1.	Предмет математической статистики.....	4
2.	Генеральная совокупность и выборки. Дискретные и интервальные вариационные ряды.....	4
2.1.	Генеральная и выборочная совокупность	4
2.2.	Вариационные ряды	5
3.	Графическое изображение вариационных рядов	7
3.1.	Функция распределения выборки и ее свойства	7
3.2.	Полигон и гистограмма.....	9
4.	Числовые характеристики выборки.....	12
4.1.	Среднее арифметическое и его свойства	12
4.2.	Выборочная дисперсия и ее свойства.....	14
5.	Методы расчета сводных характеристик выборки	16
5.1.	Метод произведения.....	16
5.2.	Метод сумм	18
6.	Понятие о точечной и интервальной оценке параметров распределения	19
6.1.	Точечные оценки генеральных числовых характеристик	19
6.2.	Интервальная оценка параметров распределения	21
7.	Элементы теории корреляции. Линейная корреляция	23
7.1.	Совместное распределение случайных величин	23
7.2.	Понятия коэффициента корреляции и функции регрессии	23
7.3.	Линейная корреляция	24
	Контрольные вопросы.....	28
	Задания для самостоятельного решения	29
	Тест для самоконтроля	31

1. Предмет математической статистики

Математические законы теории вероятностей имеют определенное физическое содержание. Они выражают закономерности массовых случайных явлений природы. Математическая обработка результатов наблюдений, эксперимента, массовых случайных явлений составляет предмет специальной науки - математической статистики.

Математическая статистика и теория вероятностей – две неразрывно связанные науки. Но если теория вероятностей формально-логически изучает закономерности случайных явлений, то математическая статистика имеет дело с результатами наблюдений случайных явлений, которые можно рассматривать как значения некоторой случайной величины.

Опр. *Математическая статистика* – это раздел математики, в котором изучаются методы обработки результатов наблюдений массовых случайных явлений с целью выявления их закономерностей.

Основными задачами математической статистики являются:

- 1) Описание закона распределения СВ
- 2) Отыскание неизвестных параметров распределения
- 3) Проверка правдоподобия гипотез

2. Генеральная совокупность и выборки.

Дискретные и интервальные вариационные ряды

2.1. Генеральная и выборочная совокупность

Опр. Совокупность всех мысленно-возможных объектов данного вида, над которыми производится наблюдение с целью получения конкретных значений определенной случайной величины, называется *генеральной совокупностью*.

В зависимости от того, конечна или бесконечна совокупность составляющих ее элементов, генеральная совокупность бывает *конечной* или *бесконечной*.

Например: По статистическим данным оценивается доля мальчиков среди всех родившихся за год детей.

Генеральная совокупность – это все родившиеся за год дети. Она конечна. Если же рассматривать до бесконечности непрерывное воспроизводство населения, то такая совокупность будет бесконечной.

Опр. Часть отобранных объектов из генеральной совокупности называется *выборочной совокупностью (выборкой)*, а число объектов выборки - ее *объемом*.

Если обозначить объем генеральной совокупности N , а объем выборочной совокупности n , то предполагается, что $N \gg n$ (много больше).

Возможны два различных варианта выборки:

- 1) Случайная выборка с возвратом (повторная) – это когда объект перед отбором следующего объекта возвращается в генеральную совокупность.
- 2) Случайная выборка без возврата (бесповторная).

Т.к. о свойствах генеральной совокупности приходится судить по выборке, то для точного суждения о случайной величине выборка должна быть *представительной (репрезентативной)*. Это означает, что объекты выборки должны достаточно правильно представлять генеральную совокупность, т.е. несмотря на случайный характер отбора объектов, они должны иметь одинаковую вероятность попасть в выборку.

Например, в цехе работают квалифицированные и начинающие токари. На контроль взята партия втулок. Если втулки изготовлены квалифицированным токарем, то представление о качестве будет завышенным, а если начинающим, то заниженным. Такого не должно быть.

2.2. Вариационные ряды

Т.к. основная задача математической статистики состоит в получении обоснованных выводов о генеральной совокупности по известным свойствам извлеченной из нее выборки, то сама выборка должна быть представлена в удобном для дальнейшего использования виде.

Обычно полученные наблюдаемые данные представляют собой множество расположенных в беспорядке чисел, просматривая которые бывает трудно выявить какую-либо закономерность их варьирования (изменения). Для изучения закономерностей варьирования значений случайной величины опытные данные подвергают обработке. Сначала их ранжируют.

Опр. *Ранжированием* опытных данных называется операция, заключающаяся в том, что результаты наблюдений над случайной величиной располагают в порядке не убывания.

После этой операции опытные данные объединяют в группы так, чтобы в каждой группе значения случайной величины были одинаковыми.

Опр. Значение случайной величины, соответствующее определенной группе сгруппированного ряда наблюдаемых данных, называется *вариантой* и обозначается x_i , где i – индекс варианты.

Опр. Численность отдельной группы сгруппированного ряда наблюдаемых данных называется *частотой* варианты (ее *весом*) и обозначается m_i .

Опр. Отношение частоты данной варианты к общей сумме частот всех вариантов называется *относительной частотой* (*частостью*) варианты и обозначается p_i^* .

Т.о.:
$$p_i^* = \frac{m_i}{\sum_{i=1}^k m_i}, \text{ где } k - \text{число всех вариантов}$$

Т.к. $n = \sum_{i=1}^k m_i$, то $p_i^* = \frac{m_i}{n}$. Очевидно, что $\sum_{i=1}^k p_i^* = 1$.

(!!) *Частость является статистической вероятностью появления варианты x_i .*

Опр. Ранжированная совокупность вариантов x_i с соответствующими им частотами m_i (или частостями p_i^*) называется *дискретным вариационным рядом распределения*.

Удобнее всего такой ряд представлять в виде таблицы.

ПР.: На телефонной станции проводились наблюдения над числом неправильных соединений в минуту (X). Получились следующие результаты: 0,

7, 4, 2, 1, 3, 7, 5, 6, 7, 7, 0, 0, 1, 2, 2, 4, 4, 6, 6, 0, 1, 1, 3, 4, 4, 5, 7, 1, 7. Составить ряд распределения случайной величины X .

x_i	0	1	2	3	4	5	6	7
m_i	4	5	3	2	5	2	3	6
p_i^*	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{5}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{6}{30}$

Если изучаемая случайная величина является непрерывной, то ранжирование и группировка не позволяют выделить характерные черты варьирования ее значений. Это объясняется тем, что значения случайной величины могут мало отличаться друг от друга и поэтому в совокупности наблюдаемых данных одинаковые значения величины могут встречаться редко, а частоты вариантов мало отличаются друг от друга. Если число возможных значений дискретной случайной величины велико, то построение дискретного ряда также нецелесообразно. В таких случаях строят интервальный ряд распределения.

Опр. *Интервальным вариационным рядом распределения* называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами (частостями) попадания значений величины в каждый из них.

Иногда интервальный ряд для простоты исследования условно заменяют дискретным. В этом случае срединное значение i -го интервала принимают за варианту x_i , а соответствующую интервальную частоту m_i – за частоту этой варианты.

3. Графическое изображение вариационных рядов

3.1. Функция распределения выборки и ее свойства

Пусть имеется выборочная совокупность значений некоторой случайной величины X объема n и каждой variante из этой совокупности поставлена в соответствие ее частота.

Опр. *Выборочной (эмпирической) функцией распределения* называется функция $F^*(x)$, задающая для каждого значения x относительную частоту события $X < x$.

Если обозначить через m_x число выборочных значений величины X , меньших некоторого действительного числа x , то $F^*(x) = \frac{m_x}{n}$.

(!!) Если $F(x)$ определяет вероятность события $X < x$, то $F^*(x)$ – относительную частоту этого события.

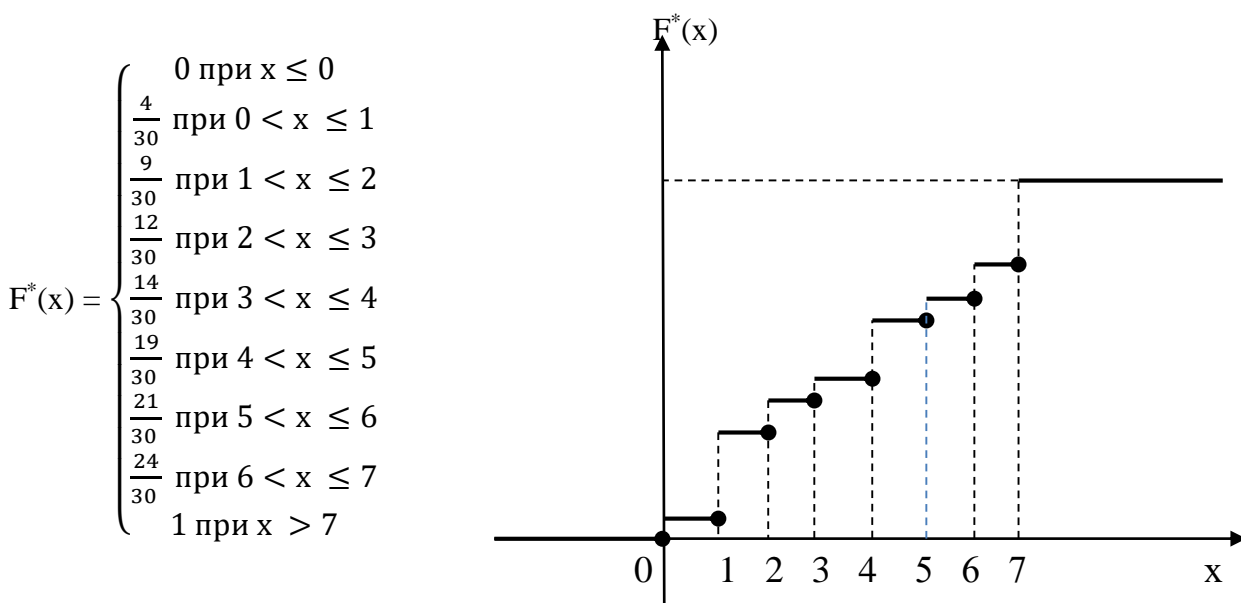
Функция $F^*(x)$ обладает такими же свойствами, что и $F(x)$, т.е.:

- 1) $0 \leq F^*(x) \leq 1$
- 2) Неубывающая
- 3) Если x_1 – наименьшая варианта, а x_k – наибольшая, то при $x \leq x_1$ $F^*(x) = 0$, а при $x > x_k$ $F^*(x) = 1$.

Опр. График эмпирической функции распределения называется *кумулятой*.

Пр. Построим эмпирическую функцию распределения и ее график по данным предыдущей таблицы:

x_i	0	1	2	3	4	5	6	7
P_i^*	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{3}{30}$	$\frac{2}{30}$	$\frac{5}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{6}{30}$



В нашем примере $F^*(x)$ построена по дискретному вариационному ряду. Если результаты наблюдений представлены в виде интервального вариационного ряда, то эмпирическую функцию распределения в таком виде построить нельзя. Для нее график будет представлять собой непрерывную линию.

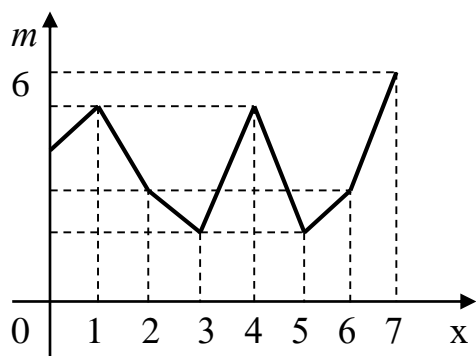
3.2. Полигон и гистограмма

Наблюдаемые данные, представленные в виде вариационного ряда, можно изобразить графически, используя не только функцию $F^*(x)$. К наиболее распространенным видам графического изображения вариационных рядов относятся *полигон* и *гистограмма*.

Полигон используется, в основном, для изображения дискретного вариационного ряда. При этом различают полигон частот и полигон относительных частот.

Опр. *Полигоном частот (относительных частот)* называется ломаная, отрезки которой соединяют точки $(x_1; t_1)$, $(x_2; t_2)$, ..., $(x_1; p^*_1)$, $(x_2; p^*_2)$, ...).

Построим полигон частот для предыдущей задачи.



(!!) Полигон является выборочным аналогом многоугольника распределения вероятностей.

Гистограмма строится только для интервальных вариационных рядов.

Опр. *Гистограммой частот* называется ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы

длины h , а высоты равны отношению $\frac{m_i}{h}$ (плотность частоты), где m_i – сумма частот вариант, попавших в i -ый интервал.

(!!) Очевидно, что площадь частичного i -ого прямоугольника равна $h \cdot \frac{m_i}{h} = m_i$, т.е. сумме частот вариант, попавших в i -ый интервал. Площадь гистограммы частот равна сумме всех частот (объему выборки n).

Аналогично гистограммы частот определяется гистограмма относительных частот. При этом: $\frac{p_i^*}{h}$ – плотность относительной частоты.

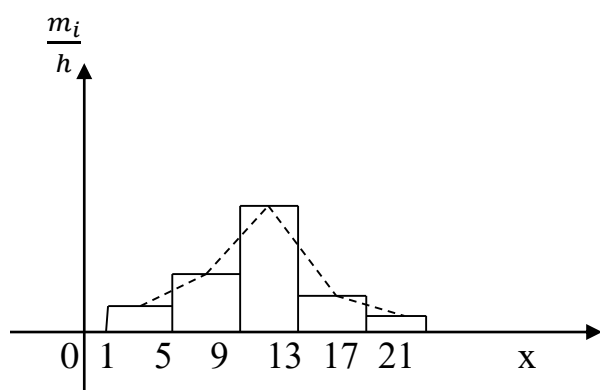
(!!) Площадь частичного i -ого прямоугольника равна $h \cdot \frac{p_i^*}{h} = p_i^*$, а площадь гистограммы относительных частот равна 1.

Если длины интервалов разные, то при построении гистограммы это надо учитывать. Например, если интервалы имеют длину 10, кроме крайнего, который имеет длину 50, то все попавшие в него данные можно мысленно разбить на 5 одинаковых частей, каждая из которых попала бы в свой интервал длины 10. Следовательно, высота прямоугольника над этим интервалом длины 50 должна быть в 5 раз меньше.

ПР. Построить гистограмму частот по данному распределению выборки.

Интервал $X_i - X_{i+1}$	Сумма частот вариант Интервала m_i	Плотность частоты $\frac{m_i}{h}$
1 - 5	10	2,5
5 - 9	20	5
9 - 13	50	12,5
13 - 17	12	3
17 - 21	8	2

$n = 100, h = 4$



Для графического изображения интервального вариационного ряда можно использовать полигон, если этот ряд преобразовать в дискретный. Для этого каждый интервал заменяют его срединным значением и ставят ему в соответствие интервальную частоту (частость). Для полученного дискретного ряда строят полигон (см. пунктирную линию).

При построении интервального распределения существуют правила выбора числа интервалов и величины каждого интервала. Критерием здесь служит оптимальное соотношение, т.к. при увеличении числа интервалов улучшается репрезентативность, но увеличивается объем данных и время на их обработку.

Для подсчета числа интервалов k обычно применяют эмпирическую формулу Стерджесса (подразумевается округление до ближайшего целого):

$$k = 1 + 3,322 \cdot \lg n$$

Тогда величину каждого интервала h можно вычислить по формуле:

$$h = \frac{x_{\max} - x_{\min}}{k}$$

Разность между наибольшим и наименьшим значениями вариант называют **размахом выборки**.

(!!) Желательно, чтобы каждый интервал содержал не менее пяти вариант.

ПР. Даны результаты изменения напряжения в электросети (в вольтах). Составить вариационный ряд и построить гистограмму относительных частот, если значения напряжения следующие: 227, 215, 230, 232, 223, 220, 228, 222, 221, 226, 226, 215, 218, 220, 216, 220, 225, 212, 217, 220.

По условию: $n = 20$, $x_{\max} = 232$, $x_{\min} = 212$

$$k = 1 + 3,322 \cdot \lg 20 \approx 5 \Rightarrow h = \frac{232 - 212}{5} = 4$$

Составим интервальный вариационный ряд относительных частот:

x_i	212-217	217-222	222-227	227-232
P_i^*	0,25	0,35	0,25	0,15
$\frac{p_i}{h}$	0,05	0,07	0,05	0,03

4. Числовые характеристики выборки

Построив вариационный ряд и изобразив его графически, можно получить первоначальное представление о закономерностях, имеющих место в ряду наблюдений. Однако на практике этого недостаточно, особенно если имеется необходимость сравнивать ряды. Поэтому для дальнейшего изучения изменений значений СВ используют числовые характеристики вариационных рядов. Поскольку они вычисляются по статистическим данным, их называют *статистическими характеристиками* или *оценками*.

Пусть статистический материал представлен в виде вариационного ряда. Для того чтобы проанализировать ряд, удобно выделить некоторые постоянные, которые отражали бы присущие изучаемой совокупности закономерности.

Некоторые из этих постоянных отличаются тем, что вокруг них концентрируются остальные результаты наблюдений. Такие величины называются *средними величинами*. К ним относятся *среднее арифметическое*, *среднее геометрическое*, *среднее гармоническое* и т.п. Другие - отражают «величину изменчивости» наблюдаемых данных, например, величину разброса значений СВ вокруг среднего арифметического. Их называют *показателями вариации*. К ним относятся: *размах варьирования*, *среднеквадратическое отклонение*, *дисперсия* и т.д.

4.1. Среднее арифметическое и его свойства

Пусть x_1, x_2, \dots, x_n – данные наблюдений над случайной величиной X .

Опр. *Средним арифметическим* \bar{X} наблюдаемых значений случайной величины X называется частное от деления суммы всех этих значений на их число, т.е.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

(!!) Очевидно, что среднее арифметическое – величина той же размерности, что и значения СВ.

Если данные наблюдений представлены в виде дискретного вариационного ряда, где x_1, x_2, \dots, x_k – наблюдаемые варианты, а m_1, m_2, \dots, m_k – соответствующие им частоты ($\sum_{i=1}^k m_i = n$), то по определению,

$$\bar{X} = \frac{\sum_{i=1}^k x_i m_i}{n} \quad (2)$$

Вычисленное по этой формуле среднее арифметическое называется **взвешенным**, т.к. частоты m_i называются **веса́ми**, а операция умножения x_i на m_i – **взвешиванием**.

Преобразуем последнюю формулу: $\bar{X} = \frac{\sum_{i=1}^k x_i m_i}{n} = \sum_{i=1}^k x_i \cdot \frac{m_i}{n} = \sum_{i=1}^k x_i \cdot p_i^*$

$$\bar{X} = \sum_{i=1}^k x_i \cdot p_i^*$$

Для интервального вариационного ряда за x_i принимают середину i -го интервала, а за m_i – соответствующую интервальную частоту. В этом случае значения средних арифметических, вычисленных по формулам (1) и (2) могут не совпадать, т.к. в формуле (2) значения СВ внутри каждого интервала принимаются равными середине интервала, в то время, как они могут быть произвольно расположены в интервале.

Основные свойства среднего арифметического

1) Среднее арифметическое алгебраической суммы соответствующих друг другу значений, принадлежащих двум группам наблюдений, равна алгебраической сумме средних арифметических этих групп, т.е.

$$\overline{X \pm Y} = \bar{X} \pm \bar{Y}$$

2) Если ряд наблюдений состоит из 2-х непересекающихся групп наблюдений, то среднее арифметическое \bar{Z} всего ряда наблюдений равно взвешенному среднему арифметическому групповых средних \bar{X} и \bar{Y} , причем весами являются объемы групп n_1 и n_2 соответственно, т.е.

$$\bar{Z} = \frac{\bar{X}n_1 + \bar{Y}n_2}{n_1 + n_2}$$

3) Среднее арифметическое постоянной равно самой постоянной, т.е.

$$\bar{C} = C$$

4) Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство: $\overline{CX} = C \overline{X}$, т.е. постоянную можно выносить за знак среднего арифметического.

5) Если все результаты наблюдений увеличить (уменьшить) на одно и то же число, то среднее арифметическое увеличится (уменьшится) на то же число, т.е. $\overline{X \pm C} = \overline{X} \pm C$

6) Если все частоты вариант умножить на одно и то же число, то среднее арифметическое не изменится.

7) Сумма отклонений $x_i - \overline{X}$ результатов наблюдений от среднего арифметического равна нулю.

(!!) Между математическим ожиданием и средним арифметическим случайной величины видна большая схожесть. Среднее арифметическое – выборочный аналог математического ожидания. Однако, если математическое ожидание является постоянной величиной (что следует из определения), то среднее арифметическое не обладает этим свойством.

4.2. Выборочная дисперсия и ее свойства

На величину рассеивания наблюдаемых значений случайной величины вокруг ее среднего арифметического \overline{X} влияет каждое отклонение $x_i - \overline{X}$, однако сумма всех этих отклонений не может быть мерой рассеивания, т.к. она равна нулю. Чтобы на оценку меры рассеивания влияли все указанные отклонения, и их знаки при этом не играли роли, целесообразно использовать не сами отклонения, а их квадраты. Лучше всего для этой цели взять среднее арифметическое величин $(x_i - \overline{X})^2$.

Опр. *Выборочной дисперсией* $D^*(X)$ значений случайной величины X является среднее арифметическое квадратов отклонений наблюдаемых значений этой величины от их среднего арифметического.

Обозначение:

Т.о.:
$$D^*(X) = \frac{\sum_{i=1}^k (x_i - \overline{X})^2 m_i}{n}$$
 или
$$D^*(X) = \sum_{i=1}^k (x_i - \overline{X})^2 p_i^*$$

(!!) Если данные представлены в виде интервального ряда, то его необходимо заменить дискретным и только после этого применять данные формулы.

Выборочная дисперсия выражается в квадратных единицах. Поэтому вводят понятие **выборочного среднеквадратического отклонения**:

$$\sigma^*(X) = \sqrt{D^*(X)}$$

Основные свойства выборочной дисперсии

- 1) Дисперсия постоянной равна нулю, т.е. $D^*(C) = 0$
- 2) Если все результаты наблюдений увеличить (уменьшить) на одно и то же число, то дисперсия не изменится, т.е. $D^*(X \pm C) = D^*(X)$
- 3) Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство: $D^*(CX) = C^2 D^*(X)$.
- 4) Если все частоты вариант умножить на одно и то же число, то выборочная дисперсия не изменится.
- 5) Выборочная дисперсия равна разности между средним арифметическим квадратов наблюдаемых значений СВ и квадратом их среднего арифметического, т.е. $D^*(X) = \overline{X^2} - (\bar{X})^2$

(!!) Между $D(X)$ и $D^*(X)$ много общего: обе являются мерой рассеивания, формулы для их вычисления внешне похожи и они обладают схожими свойствами. Поэтому $D^*(X)$ можно брать в качестве приближенных значений $D(X)$.

Наряду с выборочной дисперсией используют еще одну величину

$$S^2 = \frac{\sum_{i=1}^k (x_i - \bar{X})^2 m_i}{n-1}, \text{ которая связана с } D^*(X) \text{ соотношением: } S^2 = \frac{n D^*(X)}{n-1}.$$

Ее называют **исправленной выборочной дисперсией**. Тогда очевидно, что S – исправленное среднеквадратичное отклонение.

(!!) Если первоначальные варианты x_i – большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число C , т.е. перейти к условным вариантам $u_i = x_i - C$, тогда:

- 1) $\bar{X} = C + \bar{U}$;
- 2) $D^*(X)$ не изменится

ПР. Найдите выборочную среднюю, выборочную и исправленную выборочную дисперсии по данному распределению выборки:

x_i	1250	1270	1280
m_i	2	5	3

/ Перейдем к условным вариантам $u_i = x_i - 1270$, получим

u_i	-20	0	10
m_i	2	5	3

$$\bar{X} = 1270 + \frac{-20 \cdot 2 + 0 \cdot 5 + 10 \cdot 3}{10} = 1270 - 1 = 1269$$

$$D^*(X) = D^*(U) = \frac{400 \cdot 2 + 0 \cdot 5 + 100 \cdot 3}{10} - 1 = 110 - 1 = 109$$

$$S^2 = \frac{1090}{9} \approx 121 \quad /$$

(!!) Если первоначальные варианты являются десятичными дробями с к десятичными знаками после запятой, то чтобы избежать действий с дробями, умножают первоначальные варианты на постоянное число $C = 10^k$, переходя к условным вариантам. При этом выборочная средняя увеличится в C раз, а выборочная дисперсия - в C^2 раз. Поэтому, найдя среднее арифметическое и дисперсию условных вариантов, надо разделить ее на C и на C^2 соответственно.

ПР. Найдите исправленную выборочную дисперсию по данному распределению выборки

x_i	0,01	0,05	0,09
m_i	2	3	5

$$/n=10 \quad u_i = 100x_i$$

u_i	1	5	9
m_i	2	3	5

5. Методы расчета сводных характеристик выборки

5.1. Метод произведения

Пусть выборка задана в виде распределения равноотстоящих вариантов и соответствующих им частот.

В этом случае удобно находить выборочные среднюю и дисперсию методом произведений по формулам:

$$\bar{X} = M_1^* \cdot h + C \quad \text{и} \quad D^*(X) = (M_2^* - (M_1^*)^2) \cdot h^2, \quad \text{где}$$

h – шаг (разность между двумя соседними вариантами)

C – ложный нуль (варианта, которая расположена примерно в середине вариационного ряда)

$u_i = \frac{x_i - C}{h}$ – условная варианта

$M_1^* = \frac{\sum m_i u_i}{n}$ – условный момент 1-го порядка

$M_2^* = \frac{\sum m_i u_i^2}{n}$ – условный момент 2-го порядка

ПР. Найти методом произведений выборочную среднюю и выборочную дисперсии по заданному распределению выборки объёма $n = 100$.

x_i	12	14	16	18	20	22
m_i	5	15	50	16	10	4

/ 1) Найдём шаг $h = 14 - 12 = 2$

2) В качестве ложного нуля возьмем варианту, имеющую наибольшую частоту, т.е. $C = 16$.

3) Составим расчетную таблицу

x_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$
12	5	-2	-10	20
14	15	-1	-15	15
16	50	0	0	0
18	16	1	16	16
20	10	2	20	40
22	4	3	12	36
Итого	100		23	127

4) Вычислим условные моменты:

$$M_1^* = \frac{\sum m_i u_i}{n} = \frac{23}{100} = 0,23$$

$$M_2^* = \frac{\sum m_i u_i^2}{n} = \frac{127}{100} = 1,27$$

5) Найдём искомые выборочные среднюю и дисперсию:

$$\bar{X} = M_1^* \cdot h + C = 0,23 \cdot 2 + 16 = 16,46$$

$$D^*(X) = (M_2^* - (M_1^*)^2) \cdot h^2 = (1,27 - 0,23^2) \cdot 4 = 4,87 /$$

Если первоначальные варианты не являются равноотстоящими, то интервал, в котором заключены все варианты выборки делят на несколько равных, длины h , частичных интервалов (каждый интервал должен содержать не

менее 8 – 10 вариант). Затем находят середины частичных интервалов, которые и образуют последовательность равноотстоящих вариантов. В качестве частоты каждой середины интервала принимают сумму частот вариантов, которые попали в соответствующий частичный интервал.

При вычислении выборочной дисперсии для уменьшения ошибки, вызванной группировкой, делают поправку Шеппарда, а именно: вычитают из найденной дисперсии $\frac{1}{12}$ квадрата длины частичного интервала.

Т.о., с учетом поправки:

$$D^{**}(X) = D^*(X) - \frac{h^2}{12}$$

ПР. Найти методом произведений выборочную среднюю и выборочную дисперсии по заданному распределению выборки объема $n = 100$.

x_i	2	3	7	9	11	12,5	16	18	23	25	26
m_i	3	5	10	6	10	4	12	13	8	20	9

/1) Разобьем интервал 2 - 26 на 4 частичных интервала длины $h = 6$: 2 - 8; 8 - 14; 14 - 20; 20 - 24. Приняв середины этих интервалов в качестве новых вариантов y_i , получим равноотстоящие варианты: 5; 11; 17 и 23.

2) Составим вариационный ряд:

y_i	5	11	17	23
m_i	18	20	25	37

3) Пользуясь методом произведений найдём \bar{X} (15,86) и D^* (45,14).

С учетом поправки Шеппарда $D^{**} = 45,14 - \frac{6^2}{12} = 42,14$ /

5.2. Метод сумм

Известно, что выборочные среднюю и дисперсию можно найти по формулам:

$$\bar{X} = M_1^* \cdot h + C$$

и

$$D^*(X) = (M_2^* - (M_1^*)^2) \cdot h^2, \text{ где}$$

M_1^* и M_2^* – условные моменты 1-го и 2-го порядка соответственно

При использовании метода эти моменты находятся по формулам:

$$M_1^* = \frac{d_1}{n}$$

и

$$M_2^* = \frac{s_1 + 2s_2}{n}, \text{ где}$$

$$d_1 = a_1 - b_1, \quad s_1 = a_1 + b_1, \quad s_2 = a_2 + b_2$$

Т.о., при расчете сводных характеристик выборки этим методом, в конечном счете, надо найти четыре числа: a_1 , a_2 , b_1 , b_2 .

ПР. Найти методом сумм выборочную среднюю и выборочную дисперсии по заданному распределению выборки объема $n = 100$.

x_i	48	52	56	60	64	68	72	76	80	84
m_i	2	4	6	8	12	30	18	8	7	5

/ 1) Составим расчетную таблицу:

x_i	m_i	$b_1 = 72$	$b_2 = 70$
48	2	2	2
52	4	6	8
56	6	12	20
60	8	20	40
64	12	32	0
68	30	0	0
72	18	38	0
76	8	20	37
80	7	12	17
84	5	5	5
	$n = 100$	$a_1 = 75$	$a_2 = 59$

В качестве ложного нуля C выберем варианту 68, которая имеет наибольшую частоту и находится примерно в середине таблицы. В клетках строки, содержащей ложный нуль, запишем нули. В 4-ом столбце над и под уже помещенным нулем запишем еще по одному нулю. В оставшихся незаполненными над нулем клетках 3-го столбца (исключая самую верхнюю) запишем последовательно накопленные частоты: 2, $2 + 4 = 6$, $6 + 6 = 12$, ... Сложив все накопленные частоты, получим число $b_1 = 72$. Аналогично поступим с клетками под нулем. При этом получим число a_1 .

Заполняя 4-ый столбец, суммируем частоты 3-го столбца. Т.о. находим b_2 и a_2 .

2) Найдем $d_1 = a_1 - b_1 = 3$

$$s_1 = a_1 + b_1 = 147$$

$$s_2 = a_2 + b_2 = 129$$

3) Вычислим условные моменты:

$$M_1^* = 0,03, \quad M_2^* = \frac{147 + 2 \cdot 129}{100} = 4,05$$

4) Учитывая, что $h = 4$ и ложный нуль $C = 68$, находим:

$$\bar{X} = M_1^* \cdot h + C = 0,03 \cdot 4 + 68 = 68,12$$

$$D^*(X) = (M_2^* - (M_1^*)^2) \cdot h^2 = (4,05 - 0,03^2) \cdot 14 = 64,87 /$$

6. Понятие о точечной и интервальной оценке параметров распределения

6.1. Точечные оценки генеральных числовых характеристик

Все выборочные числовые характеристики $(\bar{X}, D^*(X), \dots)$ используются в качестве приближенных значений неизвестных числовых характеристик случайной величины X (неизвестных генеральных характеристик $M(X)$ и $D(X)$).

Опр. Выборочная характеристика, используемая в качестве приближенного значения неизвестной генеральной характеристики, называется ее *точечной статистической оценкой*.

«Точечная» означает, что оценка представляет собой число или точку на числовой оси. «Статистическая» означает, что оценка рассчитывается по результатам наблюдений (т.е. по собранной исследователем статистике).

Обозначим через θ (тетта) некоторую генеральную характеристику. Ее числовое значение неизвестно, однако предложен некоторый алгоритм или формула вычисления точечной оценки θ^* этой характеристики по результатам x_1, x_2, \dots, x_n наблюдений СВ X .

Для того чтобы точечная оценка была хорошим приближением к неизвестной генеральной характеристике, необходимо, чтобы она обладала следующими свойствами:

- 1) была состоятельной (чем больше объем, тем ближе к истине);
- 2) была несмещенной (математическое ожидание точечной оценки равно оцениваемому параметру);
- 3) была эффективной (среди всех прочих оценок той же самой характеристики обладала бы наименьшей дисперсией).

Обозначим через p неизвестную вероятность появления случайного события A в единичном испытании. Найдём приближенное значение p^* вероятности p . Проведем n независимых испытаний, в каждом из которых событие A может произойти с вероятностью p и не произойти с вероятностью $q = 1 - p$. Пусть m – количество испытаний, в которых произошло событие A , тогда $p^* = \frac{m}{n}$ – это опытная вероятность или относительная частота появления события A . Можно показать, что p^* обладает свойствами состоятельности, не-

смещенности и эффективности, а значит, может служить точечной оценкой вероятности.

Можно доказать, что выборочное среднее удовлетворяет всем 3-м вышеперечисленным свойствам, а потому является точечной оценкой математического ожидания.

Наряду с выборочной дисперсией $D^*(X)$ в качестве приближенного значения генеральной дисперсии чаще используют исправленную выборочную дисперсию, т.к. она удовлетворяет свойству несмещенности.

Методы получения точечных оценок

- 1) Метод моментов
- 2) Метод наибольшего правдоподобия
- 3) Метод наименьших квадратов (частный случай 2-го метода)

6.2. Интервальная оценка параметров распределения

Вычисляя на основании результатов наблюдений точечную оценку θ^* неизвестной числовой характеристики θ , получают лишь приближенное значение θ . Если для большого числа наблюдений точность приближения бывает достаточной для практических выводов, то для выборок небольшого объема вопрос о точности оценок очень важен. В МС он решается следующим образом: по данной выборке находится точечная оценка θ^* , затем задают вероятность γ и по определенным правилам находят такое число $\varepsilon > 0$, чтобы выполнялось соотношение:

$$P(|\theta^* - \theta| < \varepsilon) = \gamma$$

При этом число ε называется **точностью оценки** θ (чем меньше ε , тем выше точность), числа $\theta_1 = \theta^* - \varepsilon$ и $\theta_2 = \theta^* + \varepsilon$ называются **доверительными границами**, интервал $(\theta_1; \theta_2)$ – **доверительным интервалом** (интервальной оценкой характеристики θ).

Вероятность γ называется **доверительной вероятностью** или **надежностью интервальной оценки**. Надежность принято выбирать, равной 0,95;

0,99; 0,999. Тогда событие, состоящее в том, что интервал $(\theta_1; \theta_2)$ накроет характеристику θ , будет практически достоверным.

Поскольку довольно часто встречаются нормально распределенные случайные величины, то рассмотрим интервальные оценки для параметров нормального распределения.

Интервальной оценкой (с надежностью γ) математического ожидания a нормально распределенного количественного признака X по выборочной средней \bar{X} при известном среднеквадратичном отклонении σ генеральной совокупности служит доверительный интервал:

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

где n – объем выборки, t – значение аргумента функции Лапласа $\Phi(t)$, при котором $\Phi(t) = \frac{\gamma}{2}$.

При неизвестном σ :

$$\bar{X} - t_\gamma \cdot \frac{S}{\sqrt{n}} < a < \bar{X} + t_\gamma \cdot \frac{S}{\sqrt{n}}$$

где S – исправленное выборочное среднеквадратическое отклонение, а t_γ находятся по таблице приложения «Таблица значений $t_\gamma = t(\gamma, n)$ ».

ПР. Из генеральной совокупности извлечена выборка объема $n = 10$.

x_i	-2	1	2	3	4	5
m_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.

/ 1) Найдем выборочную среднюю $\bar{X} = \frac{-2 \cdot 2 + 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 2 + 5 \cdot 1}{10} = 2$

2) Найдем исправленное среднеквадратичное отклонение

$$S = \sqrt{\frac{(-2-2)^2 \cdot 2 + (1-2)^2 \cdot 1 + (2-2)^2 \cdot 2 + (3-2)^2 \cdot 2 + (4-2)^2 \cdot 2 + (5-2)^2 \cdot 1}{10-1}} \approx 2,4$$

3) Найдем t_γ , пользуясь таблицей приложений при $\gamma = 0,95$ и $n = 10$

$$t_\gamma = 2,26$$

4) Найдем искомый доверительный интервал по формуле $\bar{X} - t_\gamma \cdot \frac{S}{\sqrt{n}} < a < \bar{X} + t_\gamma \cdot \frac{S}{\sqrt{n}}$

Подстановка дает: $0,3 < a < 3,7$ /

Интервальной оценкой (с надежностью γ) среднеквадратичного отклонения σ нормально распределенного количественного признака X по исправленному выборочному среднеквадратичному отклонению S служит доверительный интервал:

- $S(1 - q) < \sigma < S(1 + q)$ при $q < 1$,
- $0 < \sigma < S(1 + q)$ при $q > 1$,

где q находят по заданным n и γ в таблице приложений «Таблица значений $q = q(\gamma, n)$ ».

7. Элементы теории корреляции. Линейная корреляция

7.1. Совместное распределение случайных величин

В математическом анализе имеют место с функциональными зависимостями между 2-мя переменными, при которой каждому значению одной из них соответствует единственное значение другой.

Однако в жизни чаще приходится иметь дело с более сложной зависимостью, чем функциональная. Такая зависимость возникает тогда, когда одна их величин зависит не только от другой, но и от ряда меняющихся факторов. При этом каждому фиксированному значению одной переменной соответствует не одно, а множество значений другой, причем нельзя сказать заранее, какое именно значение она примет.

Такую зависимость 2-х переменных называют **стохастической** (вероятностной).

7.2. Понятия коэффициента корреляции и функции регрессии

Пусть X и Y – две случайные величины. Если при изменении значений одной из них изменяются выборочные средние значения другой, то говорят, что имеет место **корреляционная зависимость**.

Например: 1) зависимость массы от роста; 2) зависимость заболеваемости от воздействия внешних факторов (запыленности, уровня радиа-

ции, солнечной активности и т.д.); 3) зависимость между количеством пропущенных студентами лекций и оценкой на экзамене.

Для характеристики корреляционной зависимости между случайными величинами вводится понятие **коэффициента корреляции**. Коэффициент корреляции принимает значения от -1 до +1. Чем выше значение модуля коэффициента корреляции, тем больше зависимость между величинами. Если повышение уровня одной переменной сопровождается повышением уровня другой, то речь идет о положительной корреляции. Например, чем выше личностная тревожность, тем больше риск заболеть язвой желудка. Если рост уровня одной переменной сопровождается снижением уровня другой, то мы имеем дело с отрицательной корреляцией. Например, чем боязливей особь, тем меньше у нее шансов занять доминирующее положение в группе.

Если переменные являются независимы, то коэффициент корреляции равен нулю, при этом корреляцию называют нулевой.

Корреляция бывает линейной, криволинейной (параболической, гиперболической, показательной и т.п.) и ранговой.

Опр. Функция, задающая зависимость одной случайной величины от значений, которые принимает другая случайная величина, называется **функцией регрессии**, а аналитическая форма представления этой функции - **уравнением регрессии**.

7.3. Линейная корреляция

Если модуль коэффициента корреляции равен 1, то между случайными величинами X и Y имеет место **линейная корреляция**. Ее можно описать линейным уравнением.

Выборочное уравнение прямой линии регрессии имеет вид:

$$y - \bar{Y} = R \frac{\sigma(Y)}{\sigma(X)} (x - \bar{X}),$$

где R – выборочный коэффициент корреляции, причем

$$R = \frac{\sum xym_{xy} - n\bar{X}\bar{Y}}{n\sigma(X)\sigma(Y)}$$

Если данные наблюдений над признаками X и Y заданы в виде корреляционной таблицы с равноотстоящими вариантами, то целесообразно перейти к условным вариантам: $u_i = \frac{x_i - C_1}{h_1}$ и $v_i = \frac{y_i - C_2}{h_2}$, где C_1 - ложный нуль, h_1 - шаг вариант X, а C_2 - ложный нуль, h_2 - шаг вариант Y.

В этом случае выборочный коэффициент корреляции

$$R = \frac{\sum uv m_{uv} - n\bar{U}\bar{V}}{n\sigma(U)\sigma(V)}$$

Слагаемое $\sum m_{uv} uv$ удобно вычислять, используя специальную расчетную таблицу.

Величины $\bar{U}, \bar{V}, \sigma(U), \sigma(V)$ могут быть найдены либо методом произведения, либо непосредственно по формулам:

$$\bar{U} = \frac{\sum u_i m_i}{n}, \quad \bar{V} = \frac{\sum v_i m_i}{n}, \quad \sigma(U) = \sqrt{\overline{U^2} - \bar{U}^2}, \quad \sigma(V) = \sqrt{\overline{V^2} - \bar{V}^2}$$

Зная эти величины, можно определить входящие в уравнение регрессии величины по формулам:

$$\bar{X} = \bar{U}h_1 + C_1, \quad \bar{Y} = \bar{V}h_2 + C_2, \quad \sigma(X) = \sigma(U)h_1, \quad \sigma(Y) = \sigma(V)h_2$$

ПР. Найти выборочное уравнение прямой линии регрессии Y на X по данным, приведенным в корреляционной таблице:

Y	X					m _y
	20	25	30	35	40	
16	4	6	-	-	-	10
26	-	8	10	-	-	18
36	-	-	32	3	9	44
46	-	-	4	12	6	22
56	-	-	-	1	5	6
m _x	4	14	46	16	20	n = 100

/ По условию: $h_1 = 5, h_2 = 10$.

Составим корреляционную таблицу в условных вариантах, выбрав в качестве ложных нулей $C_1 = 30$ и $C_2 = 36$.

V	U					m _v
	-2	-1	0	1	2	
-2	4	6	-	-	-	10
-1	-	8	10	-	-	18
0	-	-	32	3	9	44
1	-	-	4	12	6	22
2	-	-	-	1	5	6
m _u	4	14	46	16	20	n = 100

Найдем:

$$\bar{U} = \frac{\sum u_i m_i}{n} = \dots = 0,34$$

$$\bar{V} = \frac{\sum v_i m_i}{n} = \dots = -0,04$$

Найдем вспомогательные величины:

$$\overline{U^2} = \frac{\sum u_i^2 m_i}{n} = \dots = 1,26$$

$$\overline{V^2} = \frac{\sum v_i^2 m_i}{n} = \dots = 1,04$$

Найдем:

$$\sigma(U) = \sqrt{\overline{U^2} - \bar{U}^2} = \dots = 1,07$$

$$\sigma(V) = \sqrt{\overline{V^2} - \bar{V}^2} = \dots = 1,02$$

Найдем $\sum uv m_{uv}$, для чего составим расчетную таблицу:

V	U					$\sum u m_{uv}$	$v \cdot \sum u m_{uv}$
	-2	-1	0	1	2		
-2	-8 4 -8	-6 6 -12	-	-	-	-14	28
-1	-	-8 8 -8	0 10 -10	-	-	-8	8
0	-	-	0 32 0	3 3 0	18 9 0	21	0
1	-	-	0 4 4	12 12 12	12 6 6	24	24
2	-	-	-	1 1 2	10 5 10	11	22
$\sum v m_{uv}$	-8	-20	-6	14	16		82
$u \cdot \sum v m_{uv}$	16	20	0	14	32	82	

1) Произведение $u m_{uv}$ варианты u на частоту m_{uv} записывают в правом верхнем углу клетки, содержащей значение частоты

2) Складывают все числа, помещенные в правых верхних углах клеток одной строки.

3) Наконец, умножают варианту v на найденную сумму и складывают все числа последнего столбца, получая тем самым искомое число $\sum uv m_{uv}$.

Т.о.: $\sum uv m_{uv} = 82$

Для контроля аналогичные вычисления производят по столбцам: произведения vt_{uv} записывают в левый нижний угол соответствующей клетки; складывают все числа, помещенные в левых нижних углах клеток одного столбца; затем умножают полученные суммы на варианты u и складывают все числа последней строки. Полученное число также равно искомой сумме $\sum uv t_{uv}$.

Совпадение сумм свидетельствует о правильности вычислений.

Найдем выборочный коэффициент корреляции:

$$R = \frac{\sum uv t_{uv} - n\bar{U}\bar{V}}{n\sigma(U)\sigma(V)} = \frac{82 - 100 \cdot 0,34 \cdot (-0,04)}{100 \cdot 1,07 \cdot 1,02} = 0,76$$

$$\bar{X} = \bar{U}h_1 + C_1 = 0,34 \cdot 5 + 36 = 31,70$$

$$\bar{Y} = \bar{V}h_2 + C_2 = -0,04 \cdot 10 + 30 = 35,60$$

$$\sigma(X) = \sigma(U)h_1 = 1,07 \cdot 5 = 5,35$$

$$\sigma(Y) = \sigma(V)h_2 = 1,02 \cdot 10 = 10,2$$

$$\text{Подставим найденные величины в уравнение: } y - \bar{Y} = R \frac{\sigma(Y)}{\sigma(X)} (x - \bar{X})$$

$$\text{Получим: } y - 35,60 = 0,76 \cdot \frac{10,2}{5,35} (x - 31,70)$$

$$y = 1,45x - 10,36/$$

Контрольные вопросы

- 1) Предмет изучения математической статистики.
- 2) Назовите основные задачи математической статистики.
- 3) Что называется генеральной совокупностью?
- 4) Дайте понятие выборочной совокупности и ее объема.
- 5) Что понимают под ранжированием опытных данных?
- 6) Дайте понятия варианты, ее частоты и относительной частоты.
- 7) Что такое дискретный вариационный ряд? интервальный вариационный ряд?
- 8) Дайте понятие эмпирической функции распределения и перечислите ее свойства.
- 9) Как называется график эмпирической функции распределения?
- 10) Что такое полигон частот (относительных частот)? Для изображения каких вариационных рядов он используется?
- 11) Дайте определение гистограммы частот (относительных частот). Какие вариационные ряды можно изобразить с их помощью? Чему равна площадь гистограммы частот (относительных частот)?
- 12) Что понимают под размахом выборки?
- 13) Дайте определение среднего арифметического наблюдаемых значений случайной величины.
- 14) Перечислите основные свойства среднего арифметического.
- 15) Что называется выборочной дисперсией? выборочным среднеквадратичным отклонением?
- 16) Назовите свойства выборочной дисперсии.
- 17) Что такое исправленная выборочная дисперсия? исправленное среднеквадратичное отклонение?
- 18) Для чего используется поправка Шеппарда?
- 19) Что понимают под точечной статистической оценкой? Какими свойствами она должна обладать?

- 20) Что такое доверительный интервал, его границы и надежность интервальной оценки?
- 21) Какую зависимость между величинами называют стохастической? корреляционной? Приведите примеры.
- 22) Дайте понятие коэффициента корреляции. Каковы его границы?
- 23) Какие существуют виды корреляции?
- 24) Дайте понятия функции и уравнения регрессии.
- 25) Определите линейную корреляцию.

Задания для самостоятельного решения

1. Задано распределение частот выборки объема $n = 20$:

x_i	2	6	12
m_i	3	10	7

Напишите распределение относительных частот.

2. Постройте вариационный ряд и полигон относительных частот по данным выборки: 5, 8, 7, 6, 7, 8, 9, 5, 10, 8, 7, 5, 6, 6, 9, 7, 8, 6, 7, 10.
3. Игральную кость бросили 8 раз. При этом 1 очко выпало 1 раз, 2 очка – 1 раз, 3 очка – 1 раз, 4 очка – 2 раза, 5 очков – 2 раза, 6 очков – 1 раз. Найдите эмпирическую функцию распределения числа очков, выпавших при бросании игральной кости. Постройте график этой функции.
4. Найдите эмпирическую функцию по данному распределению выборки.

x_i	2	3	5	6
m_i	10	15	5	20

Постройте полигон частот.

5. Постройте гистограмму относительных частот по данному распределению выборки

i	$x_i - x_{i+1}$	m_i
1	0 - 3	30
2	3 - 6	20
3	6 - 9	50

6. Найдите выборочную среднюю, выборочную и исправленную дисперсии по данному распределению выборки:

А)

1	2	5	8	9
3	4	6	4	3

Б)

0,2	0,5	0,7	0,8
5	10	25	10

В)

1580	1610	1620	1660	1710
20	10	30	25	15

7. Зарботная плата сотрудников фирмы за неделю составила 152, 176, 162, 167, 181, 155, 196, 169, 172, 182, 181, 186, 190, 186, 190, 176, 190, 176, 192, 178, 167, 210, 175, 160, 200, 150, 196, 203, 153 и 202 у.е. Составьте интервальный вариационный ряд, постройте гистограмму частот, замените его дискретным рядом и найдите методом произведений выборочную среднюю и выборочную дисперсию по найденному распределению выборки.
8. Найдите методом сумм выборочную среднюю и выборочную дисперсии по заданному распределению выборки объёма $n = 100$.

x_i	12	14	16	18	20	22
m_i	5	15	50	16	10	4

Тест для самоконтроля

Задание 1. (выберите один вариант ответа) Совокупность наблюдений, отобранных случайным образом из генеральной совокупности, называется...

Варианты ответов:

- 1) вариантой 2) частотой 3) частотью 4) выборкой

Задание 2. (выберите один вариант ответа) Чему равна частота варианты $x_2 = 3$ для выборки

x_i	2	3	5	6
m_i	10	15	5	20

Варианты ответов:

- 1) 0,15 2) 15 3) 0,3

Задание 3. (выберите один вариант ответа) Найдите вес варианты $x_4 = 7$, если объем выборки

x	0,5	2	6	7	8
m	2	5	7	?	2

 равен 20.

Варианты ответов:

- 1) 4 2) 5 3) 12

Задание 4. (выберите один вариант ответа) Дан дискретный вариационный ряд

x_i	11	15	16	22	27	30	33
m_i	5	10	16	8	6	4	1

Укажите относительную частоту варианты $x_2 = 15$.

Варианты ответов:

- 1) 10 2) 0,1 3) 0,2 4) 2 / 3

Задание 5. (выберите варианты ответов согласно тексту задания) По двадцати случайно отобранным рабочим дням года проверялось количество автобусов парка, выходящих на линию. Данные записаны в виде вариационного ряда:

x_i	40	42	46	48	50
m_i	2	3	5	7	3

Установите соответствие между числовой характеристикой вариационного ряда и ее значением:

А) размах выборки
В) выборочная мода

Б) объем выборки
Г) выборочная медиана

Варианты ответов:

- 1) 46 2) 20 3) 10 4) 48

Задание 6. (выберите один вариант ответа) Анализируются объемы ежедневных продаж некоторого товара за 11 дней. Данные записаны в виде ранжированного ряда: 4; 5; 5; 5; 5; 6; 6; 7; 7; 8; 8. Найдите среднее арифметическое собранных данных.

Варианты ответов:

- 1) 5 2) 5,5 3) 6 4) 6,5

Задание № 7. (выберите один вариант ответа) Найдите выборочную дисперсию по данному распределению выборки

x	-1	0	2
p*	0,5	0,2	0,3

Варианты ответов:

- 1) 1,69 2) 1,6 3) 0,6 4) 1,71

Задание № 8. (выберите один вариант ответа) Выборочная дисперсия вариационного ряда равна 3,5. Объем выборки равен 50. Чему равна исправленная выборочная дисперсия?

Варианты ответов:

- 1) 3,43 2) 3,57 3) 0,07 4) 3,5

Задание № 9. (выберите один вариант ответа) Что произойдет с выборочной дисперсией, если все частоты вариант умножить на одно и то же число λ ?

Варианты ответов:

- 1) увеличится в λ раз 2) увеличится на λ 3) не изменится

Задание № 10. (выберите один вариант ответа) Чему равна выборочная средняя константы С?

Варианты ответов:

- 1) 0 2) С 3) C^2

Задание № 11. (выберите один вариант ответа) Дана выборка объема n . Если каждый элемент выборки увеличить в 5 раз, то выборочное среднее \bar{X} ...

Варианты ответов:

- 1) не изменится
3) увеличится в 25 раз

- 2) уменьшится в 5 раз
4) увеличится в 5 раз

Задание № 12. (выберите один вариант ответа) Для каких вариационных рядов строится гистограмма относительных частот?

Варианты ответов:

- 1) только для дискретных
2) только для интервальных
3) и для дискретных, и для интервальных

Задание № 13. (выберите один вариант ответа) Дан вариационный ряд

варианта	1	3	6
частота	10	8	12

Чему равно значение эмпирической функции распределения $F^*(x)$ в точке $x = 5$?

Варианты ответов:

- 1) 18 2) 8 3) 0,8 4) 0,6