

Математическая статистика

Борисова Елена Викторовна,
кандидат экономических наук, доцент,
доцент кафедры МЭиФО ФГБОУ ВО РГАИС

План

1. Основы статистического описания.
2. Гистограмма и полигон частот.
3. Эмпирическое распределение и его свойства

ОСНОВЫ СТАТИСТИЧЕСКОГО ОПИСАНИЯ

Математическая статистика используется в различных областях знаний:

в экономике, опытно-деловом, земледелии, животноводстве и т.д.,

т. е. там, где для изучения процессов и явлений недостаточно только качественной характеристики.

Чтобы глубоко познать сущность процессов, необходимы количественные характеристики в виде измерений, наблюдений с их последующим анализом, обобщением и выводами.

Математическая статистика

§1. Предмет и задачи математической статистики.

Определение 1. Математическая статистика – это наука, занимающаяся разработкой методов сбора, регистрации и обработки результатов наблюдений (измерений) с целью познания закономерностей случайных массовых явлений.

Математическая статистика

*Результаты измерений (наблюдений) называют **статистическими данными**. В зависимости от поставленной цели все задачи математической статистики могут быть сформулированы в различных формах, среди которых типичными являются:*

- 1) приближенное определение неизвестного закона распределения случайной величины;*
- 2) приближенное определение неизвестных параметров распределения, т.е. их статистические оценки;*
- 3) проверка правдоподобия гипотез о распределении.*

Математическая статистика

Определение 2. Вся исследуемая совокупность однородных объектов называется генеральной совокупностью.

Если предположить, что над всеми объектами проведено наблюдение (измерение), то результаты можно рассматривать как значения случайной величины с функцией распределения $F(x)$.

Математическая статистика

Как и в теории вероятностей, вероятность для всех значений X в генеральной совокупности, меньших чем x_0 , равна $F(x_0)$.

$$F(x_0) = P(X < x_0)$$

Математическая статистика

Определение 3. Множество из n - объектов, отобранных случайным образом из генеральной совокупности, называется **выборочной совокупностью** или **выборкой** (n - объем выборки).

Одним из основных способов сбора статистических данных является **выборочный метод**.

Математическая статистика

Определение 4. Метод, основанный на том, что по данным обследования выборки, выделенной из данной генеральной совокупности, делается заключение обо всей генеральной совокупности, называется **выборочным методом**.

Определение 5. Выборка называется **репрезентативной**, если каждый объект генеральной совокупности имеет одинаковую возможность попасть в выборку.

Математическая статистика

В реальных социально - экономических системах нельзя проводить эксперименты, поэтому данные обычно представляют собой пассивные наблюдения за происходящим процессом, например: курс валюты на бирже в течение месяца, урожайность пшеницы в хозяйстве за 30 лет, производительность труда рабочих за смену и т.д.

В результате наблюдений мы получаем сведения о численной величине изучаемого признака у каждого члена данной совокупности.

Математическая статистика

Вариационные ряды.

Определение 1. Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется **вариантом** (x_i), а изменения этого значения – **варьированием**.

Результаты наблюдений, в общем случае – ряд чисел, расположены в беспорядке, поэтому их необходимо упорядочить.

Математическая статистика

Вариационные ряды.

Определение 2. Вариационным рядом называется ранжирование в порядке возрастания вариант с соответствующими им частотами (ранжир - в переводе с фр.- «ставить в ряд по росту»).

Определение 3. Операция, заключающаяся в том, что результаты наблюдений над случайной величиной располагают в порядке неубывания, называется **ранжированием опытных данных**.

Математическая статистика

Для каждой группы сгруппированного ряда данных можно подсчитать их численность, т.е. определить число, которое показывает, сколько раз встречается соответствующий вариант в ряде наблюдений.

Определение 4. Численность отдельной группы сгруппированного ряда наблюдаемых данных называется **частотой** или **весом** соответствующего варианта и обозначается m_i , где i - индекс варианта.

Математическая статистика

Определение 5. Отношение частоты данного варианта к объему совокупности называется **относительной частотой** p_i или **частотью** этого варианта.

$$p_i = \frac{m_i}{n}$$

Математическая статистика

Частота является статистической вероятностью появления варианта x_i . Она обладает свойством устойчивости, или, иначе, при выполнении определенных условий (см. предельные теоремы - теорема Бернулли) стремится по вероятности к вероятности p_i .

Математическая статистика

Пример 1: Пусть мы интересуемся размерами проданной в магазине мужской обуви за некоторый отрезок времени. Получены данные в порядке продажи:

41	39	40	38	43	41	38	41	42	40	42	41	40	42	39	41	41
36	43	42	41	38	41	40	42	41	42	42	42	40	41	41	39	42
40	40	39	41	39	38	40	41	41	40	40	40	39	42	43	37	40
42	43	42	38	40	40	41	41	41	40	43	42	42	39	43	41	40
41	42	42	43	39	41	42	43	41	42	40						

Математическая статистика

Интересующий нас признак принимает различные и притом только целые значения, причем он постоянно меняется, как говорят, варьирует.

Упорядочим записанный ряд:

¹	²	³⁻⁷	⁸⁻¹⁵	¹⁶⁻³²	³³⁻⁵³	⁵⁴⁻⁷¹	⁷²⁻⁷⁹
36	37	<u>38...38</u>	<u>39...39</u>	<u>40...40</u>	<u>41...41</u>	<u>42...42</u>	<u>43...43</u>
		5	8	17	21	18	8

Данные о количествах и размерах проданной мужской обуви будут более наглядными, если их представить в виде таблицы.

a)

b)

Таблица 1

Размер обуви (варианты) x_i	Число проданных пар (частота) m_i		Доля покупок Δ (частотность) P_i	
	частота m_i	накопленная частота S_i	частотность P_i	накопленная частотность
36	1	1	0,013	0,013
37	1	2	0,013	0,026
38	5	7	0,063	0,089
39	8	15	0,101	0,190
40	17	32	0,215	0,405
41	21	53	0,266	0,671
42	18	71	0,228	0,899
43	8	79	0,101	1,0
<i>Всего</i>	<i>n = 79</i>	-	<i>1,0</i>	-

Математическая статистика

Получен вариационный ряд. Он может быть записан с указанием числа проданных пар (частот каждого варианта) (а) или указанием доли каждого из них во всей совокупности (частостей) (б).

*Рассмотренный нами вариационный ряд называется **дискретным**.*

Математическая статистика

Определение 6. *Дискретным* вариационным рядом распределения называется ранжированная совокупность вариантов с соответствующими им частотами m_i или частотами p_i .

В общем виде его можно записать так:

x_i	x_1	x_2	\dots	x_n
m_i	m_1	m_2	\dots	m_n

Вариационный ряд часто дополнительно характеризуется накопленными частотами или накопленными частотами (таблица 1).

Математическая статистика

Определение 7. **Накопленные частоты** характеризуют число членов данной совокупности, у которых рассматриваемый признак принимает значения, не превышающие данного варианта.

Определение 8. **Накопленные частоты** – результаты последовательного суммирования частот всех вариантов, включая частоту данного варианта. Накопленная частота показывает долю членов совокупности, у которых интересующий нас признак не превосходит данного значения.

Математическая статистика

Кроме дискретных вариационных рядов широкое применение имеют *непрерывные (интервальные)* вариационные ряды.

Определение 9. *Интервальным* вариационным рядом называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частостями попаданий в каждый из них значений случайной величины.

Математическая статистика

Интервальный ряд целесообразно построить, если число возможных значений дискретной величины велико или признак является непрерывным, т.е. может принимать любые значения в пределах некоторого интервала.

Для построения интервального ряда необходимо определить величину частичных интервалов, на которые разбивается весь интервал варьирования наблюдаемых значений случайной величины.

Математическая статистика

Считая, что все частичные интервалы имеют одну и ту же длину, для каждого интервала следует установить его верхнюю и нижнюю границы, а затем в соответствии с полученной упорядоченной совокупностью частичных интервалов сгруппировать результаты наблюдений. Т.е. промежуток изменения признака разбивается на ряд отдельных интервалов и подсчитывается количество значений величины в каждом из них.

Математическая статистика

Размах варьирования определяется по формуле:

$$W = X_{\max} - X_{\min}$$

Для определения величины частичного интервала воспользуемся формулой Стерджесса:

$$h = \frac{W}{k} \quad (*), \text{ где } k - \text{ число интервалов}$$

$$k = 1 + 3,322 \lg n \quad k \in [6; 12]$$

W-размах варьирования

Математическая статистика

Тогда формулу () можно записать:*

$$h = \frac{X_{\max} - X_{\min}}{1 + 3,322 \lg n}$$

Если окажется, что h - дробное число, то за длину частичного интервала следует брать, либо ближайшее целое число, либо ближайшую простую дробь.

Математическая статистика

За начало первого интервала рекомендуется брать величину:

$$X_{\text{нач.}} = X_{\text{min}} - \frac{h}{2}$$

Конец последнего интервала $X_{\text{кон}}$ должен удовлетворять условию:

$$X_{\text{кон.}} - h \leq X_{\text{max}} < X_{\text{кон.}}$$

Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала h .

Математическая статистика

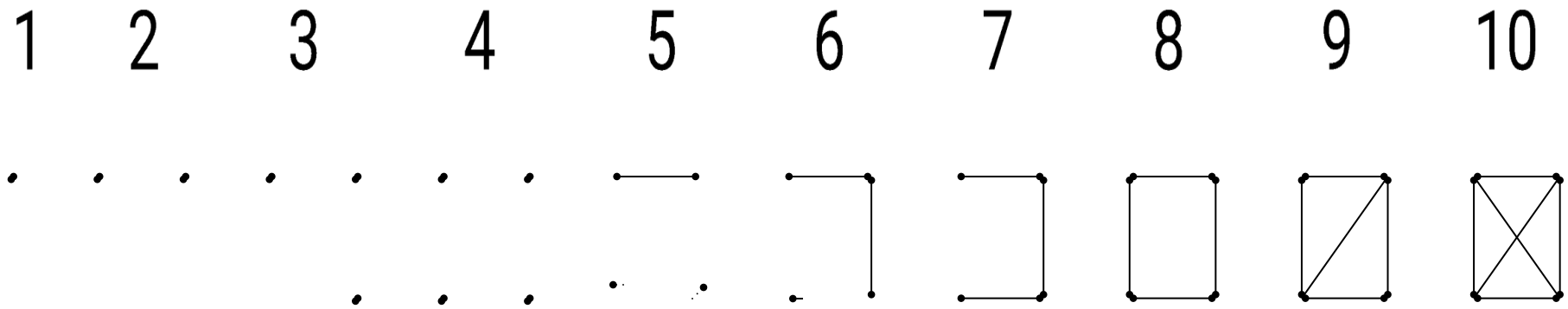
Теперь, просматривая, результаты наблюдений, определяем, сколько значений признака попало в каждый конкретный интервал. При этом в интервал включают значение случайной величины, большие или равные нижней границе и меньше верхней границы. Это можно выполнить следующим образом.

Математическая статистика

Границы полученной последовательности интервалов записывают в столбец, а затем, просматривая данные в том порядке, в котором они были получены, проставляют справа от соответствующего интервала точки или черточки. Подсчет частот для каждого интервала удобно проводить методом «конвертиков». В результате каждому десятку будет соответствовать фигура, похожая на конверт:

Математическая статистика

.



Математическая статистика

Пример 2: Пусть дан ряд распределения хозяйств по количеству рабочих на 100 га с/х угодий ($n=60$):

12 6 8 6 10 11 7 10 12 8 7 7 6 7 8 6 11 9 11 9 10
11 9 10 7 8 8 8 11 9 8 7 5 9 7 7 14 11 9 8 7 4
7 5 5 10 7 7 5 8 10 10 15 10 10 13 12 11 15 6

Построить интервальный вариационный ряд.

Математическая статистика

Решение. Для определения числа групп подставим значение $n=60$ в формулу Стерджесса:

$$k = 1 + 3,322 \lg 60 \approx 6,907; \quad k = 7$$

Найдем длину частичного интервала

$$h = \frac{X_{\max} - X_{\min}}{k} = \frac{15 - 4}{7} = \frac{11}{7} \approx 1,6$$

Построим интервальный вариационный ряд, для этого в качестве начального значения используем

X_{\min}

Группы хозяйств по численности работников на 100 га с/х угодий	Рабочее поле	Число хоз-в в группе m_i	Накопленное число хоз-в S_i	Относительная частота $\frac{m_i}{n}$ P_i
4- 5,6	— ..	5	5	5/60
5,61- 7,2	☒ ☐	17	22	17/60
7,21- 8,8	☒	9	31	9/60
8,81- 10,4	☒ — ..	15	46	15/60
10,41- 12,0	☒	10	56	10/60
12,01- 13,6	·	1	57	1/60
13,61- 15,2	· · ·	3	60	3/60
Итого	-	60	-	1

Математическая статистика

Иногда интервальный вариационный ряд для простоты исследований условно заменяют дискретным.

В этом случае серединное значение i -го интервала принимают за вариант x_i , а соответствующую интервальную частоту m_i - за частоту этого интервала.

Математическая статистика

Графическое изображение вариационных рядов.

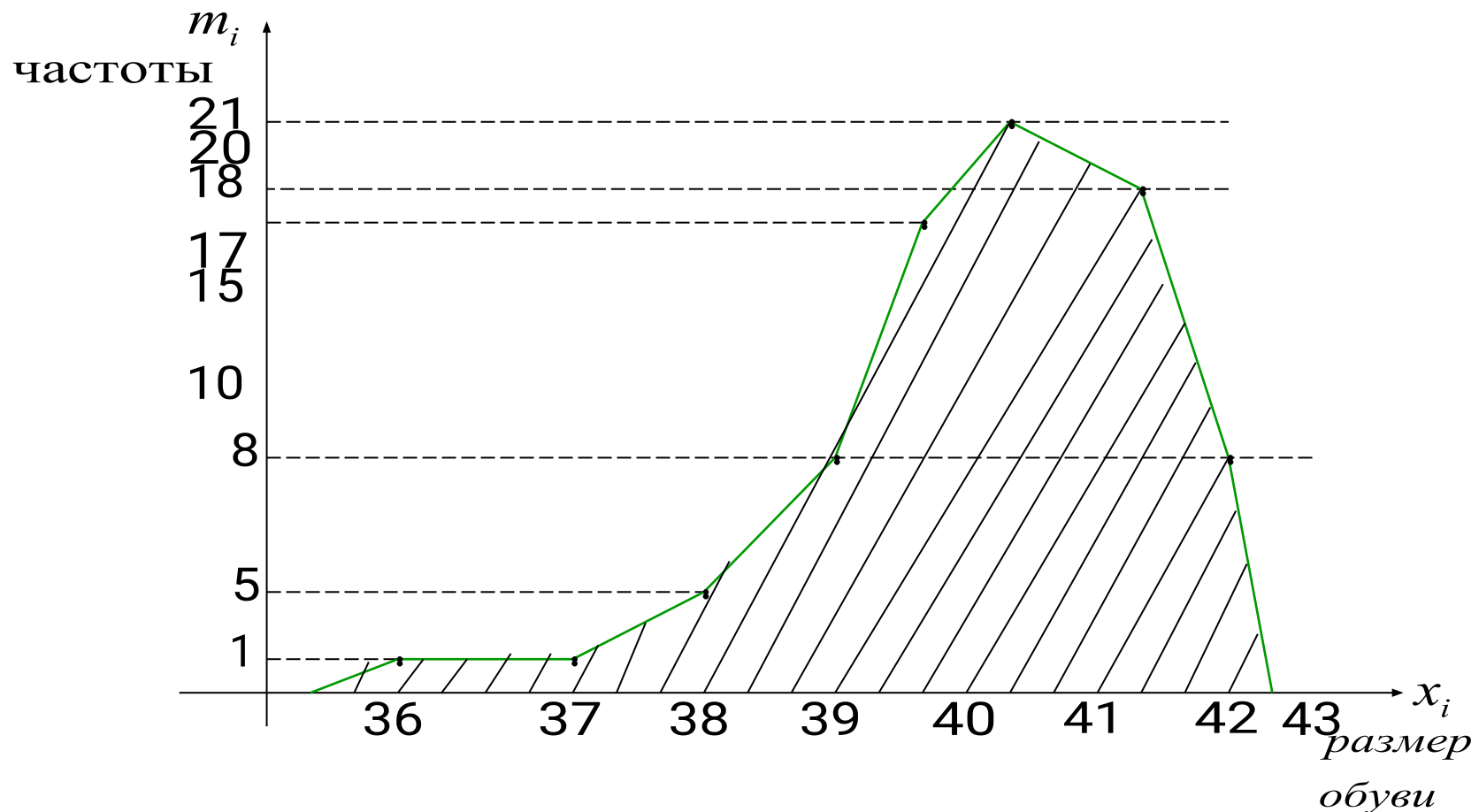
Графическое изображение позволяет представить в наглядной форме закономерности варьирования значений признаков с помощью полигона, гистограммы, кумуляты и огивы.

***Определение 1.** Полигоном (для дискретного вариационного ряда) называется ломанная, соединяющая на плоскости точки с координатами $(x_i; m_i)$.*

Математическая статистика

Графическое изображение вариационных рядов.

Пример: Построить полигон распределения по условию задачи №1.



Математическая статистика

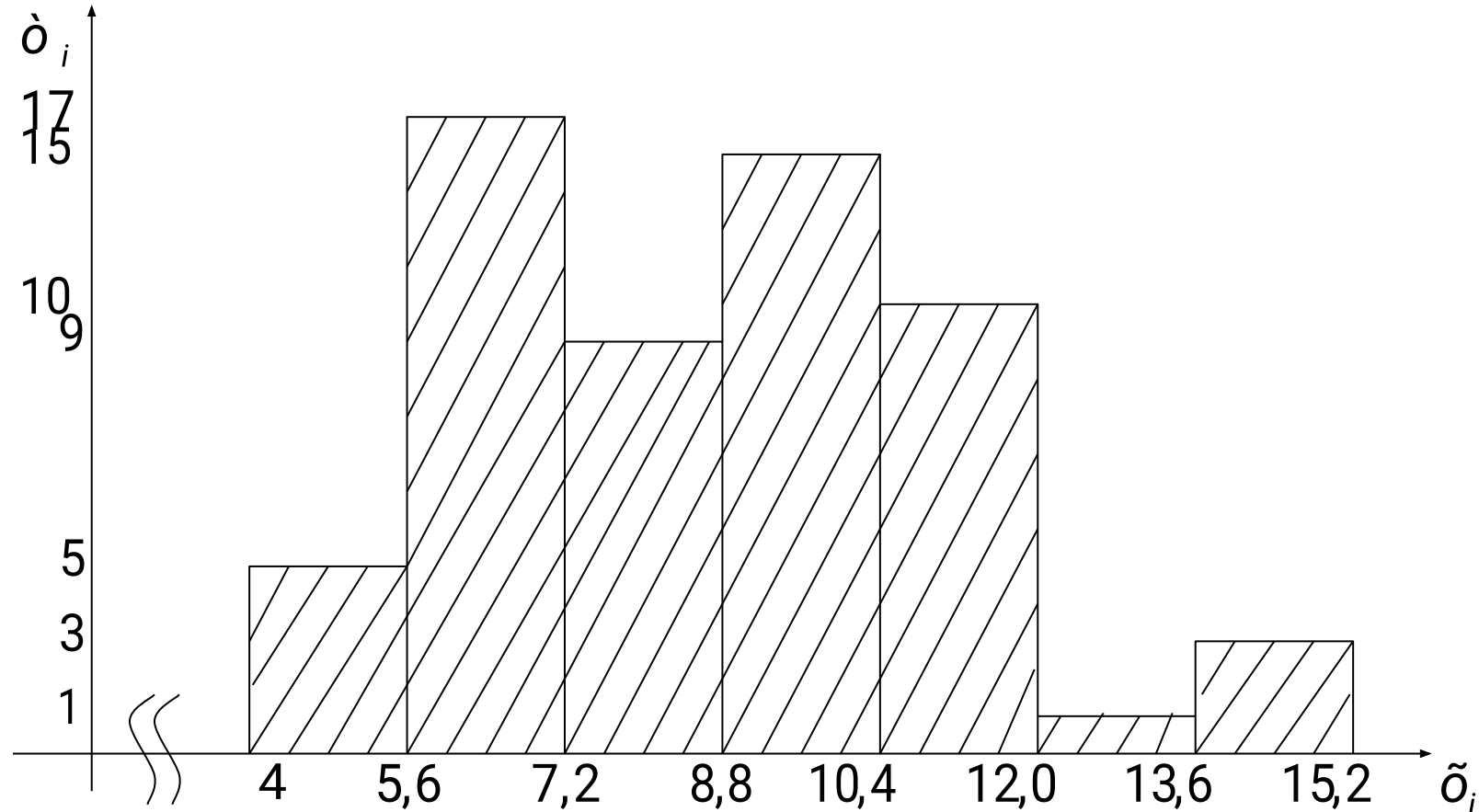
Графическое изображение вариационных рядов.

Определение 2. Гистограммой (для интервального вариационного ряда) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат интервалы $(x_{i-1}; x_i)$, а высотами - частоты m_i .

Математическая статистика

Графическое изображение вариационных рядов.

Пример: Построить гистограмму по данным примера 2.



Математическая статистика

Графическое изображение вариационных рядов.

Если в вариационном ряду вместо частот взяты соответственно накопленные частоты, то полученный ряд называется **кумулятивным рядом** (кумуляция – от латинского «скопление»).

Определение 3. **Кумулятой** называется ломанная, соединяющая на плоскости точки вида (x_i, S_i) .

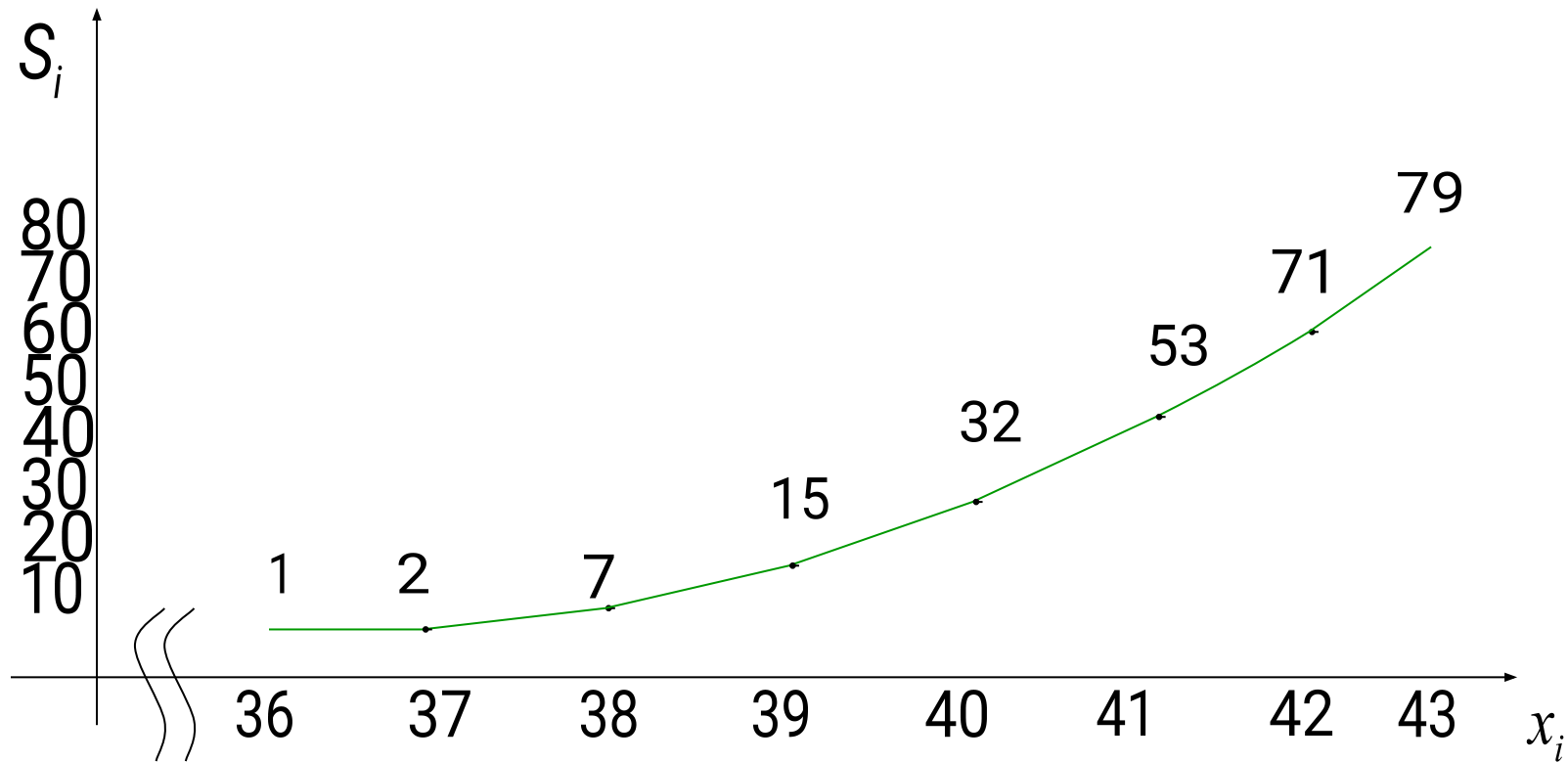
Кумуляту иначе называют **полигоном накопленных частот**.

Математическая статистика

•

Графическое изображение вариационных рядов.

Пример: Построить кумулятивную кривую по условию задачи №1.



Математическая статистика

Графическое изображение вариационных рядов.

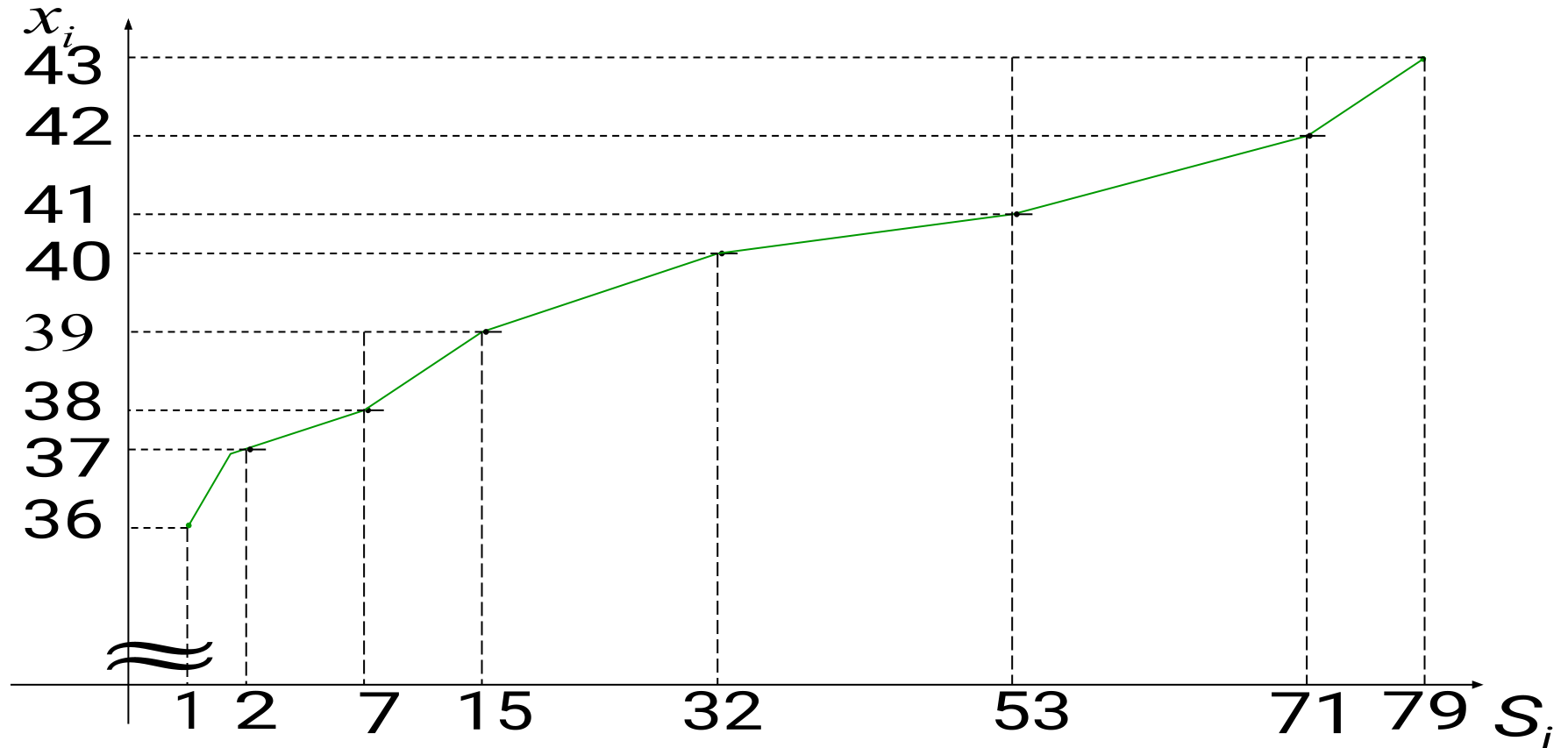
Определение 4. Если по оси абсцисс откладывают накопленные частоты, а по оси ординат - значение признака, затем полученные точки соединить отрезками, то получится *огива*.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§3. Графическое изображение вариационных рядов.

Пример: Построить огиву по условию задачи №1.



Математическая статистика

Числовые характеристики вариационных рядов.

Вариационные ряды позволяют получить первое представление об изучаемом распределении. Далее необходимо исследовать числовые характеристики распределения (аналогичные характеристикам распределения теории вероятностей): характеристики **положения** (средняя арифметическая, мода, медиана); характеристики **рассеивания** (дисперсия, среднее квадратическое отклонение, коэффициент вариации); характеристики **меры скошенности** (коэффициент асимметрии) и **островершинности** (эксцесс) распределения.

Математическая статистика

Числовые характеристики вариационных рядов.

Определение 1. Средней арифметической дискретного вариационного ряда называется отношение суммы произведений вариантов на соответствующие частоты к объему совокупности:

$$\bar{X} = \frac{\sum x_i m_i}{n}$$

(1)

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Найти (\bar{X}) по условию задачи 1.

$$\bar{X} = \frac{36 \cdot 1 + 37 \cdot 1 + 38 \cdot 5 + 39 \cdot 8 + 40 \cdot 17 + 41 \cdot 21 + 42 \cdot 18 + 43 \cdot 8}{79} = \frac{3206}{79} = 40,5$$

Вычисленное по формуле (1) среднее арифметическое называется **взвешенным**, так как частоты m_i называются **весами**, а операция умножения x_i на m_i - **взвешиванием**.

Для интервального вариационного ряда за x_i принимают середину i -го интервала, а за m_i - соответствующую интервальную частоту.

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Найти (\bar{X}) по условию задачи 2.

$$\bar{X} = \frac{4,8 \cdot 5 + 6,4 \cdot 17 + 8,0 \cdot 9 + 9,6 \cdot 15 + 11,2 \cdot 10 + 12,8 \cdot 1 + 14,4 \cdot 3}{60} = \frac{516,8}{60} \approx 8,613$$

Математическая статистика

Числовые характеристики вариационных рядов.

Определение 2. Модой $M_0(X)$ дискретного вариационного ряда называют вариант, имеющий наибольшую частоту.

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Найти $\overset{\Delta}{M}_0(x)$ по условию задачи 1.

$$m_{ax} = 20 \text{ соответствует } 4x = \Rightarrow \overset{\Delta}{M}_0(4) =$$

Для интервальных вариационных рядов при нахождении $\overset{\Delta}{M}_0(x)$ используют формулу:

$$\overset{\Delta}{M}_0(x)$$

Математическая статистика

Числовые характеристики вариационных рядов.

$$\boxed{M_0(x) = x_0 + h \cdot \frac{m_i - m_{i-1}}{(m_i - m_{i-1}) + (m_i - m_{i+1})}} \text{ , где}$$

x_0 - начало модального интервала;

h - длина частичного интервала;

m_i - частота модального интервала;

m_{i-1} - частота предмодального интервала;

m_{i+1} - частота послемодального интервала.

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Найти $\overset{\boxtimes}{M}_0(x)$ по условию задачи 2.

$m_{\max} = 17$ соответствует интервалу 5,61 – 7,2–

$$\overset{\boxtimes}{M}_0(x) = 5,61 + 1,6 \cdot \frac{17 - 5}{(17 - 5) + (17 - 9)} \approx 6,56$$

Математическая статистика

Числовые характеристики вариационных рядов.

Определение 3. Медианой ($\overset{\boxtimes}{M}_e(x)$) дискретного вариационного ряда называется вариант, делящий ряд на две равные части.

Если дискретный вариационный ряд имеет **четное** ($2n$) число членов, то:

$$\overset{\boxtimes}{M}_e(x) = \frac{x_n + x_{n+1}}{2}$$

Математическая статистика

Числовые характеристики вариационных рядов.

Если дискретный вариационный ряд имеет **нечетное** $(2n-1)$ число значений варьирующего признака, расположенных в порядке возрастания, то медианой этого распределения является вариант x_n

$$M_e(x) = x_n$$

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Найти $\overset{\sphericalangle}{M}_e(\mathbf{x})$ по условию задачи 1.

$$n = 79 \quad 2n - 1 = \overset{\sphericalangle}{79} \quad \Rightarrow \quad 2n = 80 \quad \Rightarrow \quad n = 40$$
$$x_{40} = 41 \quad \Rightarrow \quad M_e(x) = 41$$

Математическая статистика

Числовые характеристики вариационных рядов.

При нахождении $\overset{\boxtimes}{M}_e(x)$ для интервальных вариационных рядов используют формулу:

$$\overset{\boxtimes}{M}_e(x) = x_0 + h \cdot \frac{0,5n - S_{i-1}}{m_i}$$

где

Математическая статистика

Числовые характеристики вариационных рядов.

x_0 - начало медианного интервала;

h - длина частичного интервала;

n - объем совокупности;

S_{i-1} - накопленная частота интервала,
предшествующего медианному;

t_i - частота медианного интервала.

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Найти $\overset{\boxtimes}{M}_e(x)$ по условию задачи 2.

$n = 60 \Rightarrow \frac{n}{2} = 30 \Rightarrow$ медиана расположена в интервале $7,21 - 8,8$

$$\overset{\boxtimes}{M}_e(x) = 7,21 + 1,6 \frac{0,5 \cdot 60 - 22}{9} \approx 8,62$$

Математическая статистика

Числовые характеристики вариационных рядов.

Определение 4. Дисперсия вариационного ряда (как дискретного, так и интервального) характеризует средний квадрат отклонения значения признака от его среднего значения.

$$D(\) = \frac{\sum (x_i - \bar{x})^2 \cdot m_i}{n}$$

Математическая статистика

Числовые характеристики вариационных рядов.

Определение 5. Среднее квадратическое отклонение вариационного ряда распределения характеризует те же значения, что и дисперсия, но измеряется в единицах варьирующего признака.

$$\sigma(x) = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot m_i}{n}}$$

Математическая статистика

Числовые характеристики вариационных рядов.

Определение 6. Коэффициент вариации характеризует относительное значение среднего квадратического отклонения и служит для сравнения колеблемости несоизмеримых показателей.

$$V = \frac{\hat{\sigma}(x)}{\bar{X}} \cdot 100\%$$

Математическая статистика

Числовые характеристики вариационных рядов.

Соотношения между начальными и центральными моментами:

Коэффициент асимметрии -

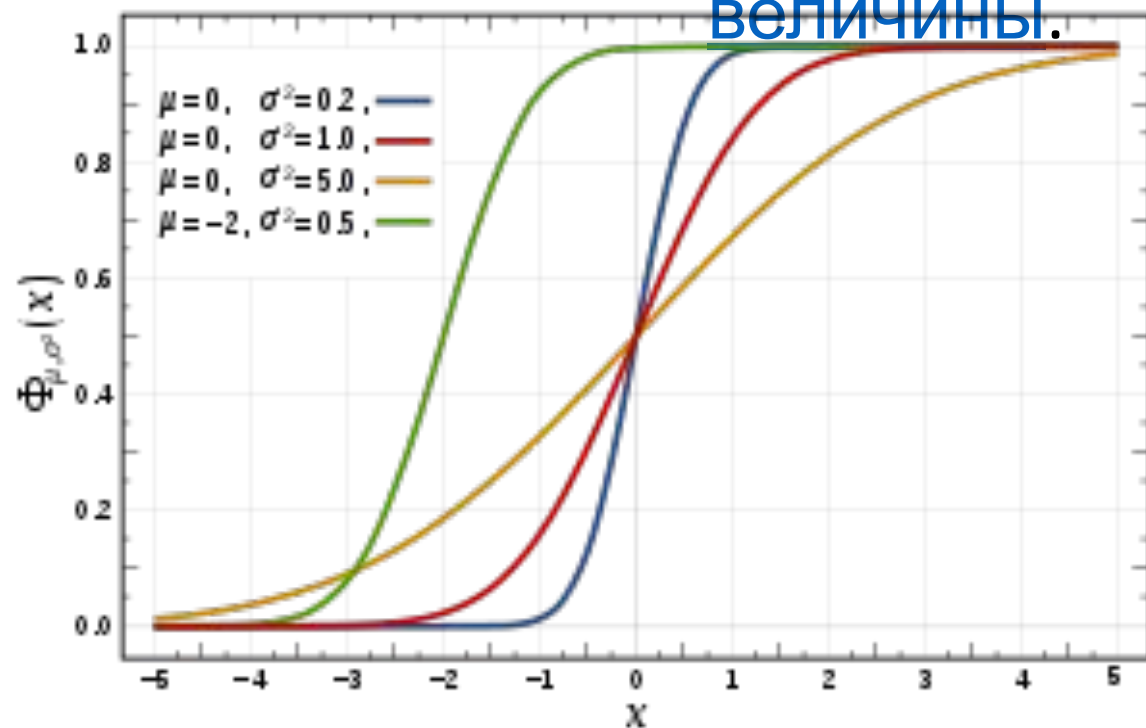
$$\boxed{A = \frac{\sum (x_i - \bar{x})^3 \cdot m_i}{n \cdot \hat{\sigma}^3(x)}}$$

Эксцесс -

$$\boxed{E = \frac{\sum (x_i - \bar{x})^4 \cdot m_i}{n \cdot \hat{\sigma}^4(x)} - 3}$$

Момент случайной величины

Момент случайной величины — числовая характеристика распределения данной случайной величины.



Функция распределения в теории вероятностей — функция, характеризующая распределение случайной величины или случайного вектора; вероятность того, что случайная величина X примет значение, меньшее или равное x , где x — произвольное действительное число.

Математическая статистика

Числовые характеристики вариационных рядов.

Моменты для вариационных рядов в математической статистике находятся по формулам, аналогичным формулам из теории вероятностей:

$$\overset{\boxtimes}{V}_k = \frac{\sum x_i^k \cdot m_i}{n} = \sum x_i^k \cdot p_i$$

- начальный момент k -го порядка.

$$\overset{\boxtimes}{M}_k = \frac{\sum (x_i - \bar{x})^k \cdot m_i}{n}$$

- центральный момент k -го порядка.

Математическая статистика

Числовые характеристики вариационных рядов.

Пример: Рассчитать дисперсию, среднее квадратическое отклонение, коэффициенты вариации, асимметрии и эксцесс для задачи 2. Сделать выводы.

Построим вспомогательную таблицу.

Группы хоз-в по численност и работников на 100 га с/х угодий, чех.	Сред- нее значе- ние инте- вала, x_i	Чис- ло хоз-в в груп- пе m_i	$x_i \cdot m_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2 \cdot m_i$	$\frac{x_i - \bar{x}}{\sigma(x)}$	$\left(\frac{x_i - \bar{x}}{\sigma(x)}\right)^3 \cdot m_i$	$\left(\frac{x_i - \bar{x}}{\sigma(x)}\right)^4 \cdot m_i$
4-5,6	4,8	5	24	-3,813	72,708	-1,559	-18,954	29,554
5,61-7,2	6,4	17	108,8	- 2,213	83,280	- 0,905	-12,601	11,404
7,21-8,8	8	9	72	- 0,613	3,386	- 0,251	- 0,142	0,036
8,81-10,4	9,6	15	144	0,987	14,603	0,403	0,985	0,397
10,41-12,0	11,2	10	112	2,587	66,908	1,058	11,832	12,514
12,01-13,6	12,8	1	12,8	4,187	17,528	1,712	5,017	8,588
13,61-15,2	14,4	3	43,2	5,787	100,457	2,366	39,740	94,030
<i>Итого</i>	-	60	516,8	0	358,869	-	25,876	156,523

Математическая статистика

Глава 1. Анализ вариационных рядов.

§4. Числовые характеристики вариационных рядов.

$$\bar{x} = \frac{516,8}{60} = 8,613$$

$$D(x) = \frac{\sum (x_i - \bar{x})^2 \cdot m_i}{n} = \frac{358,869}{60} = 5,981$$

$$\sigma(x) = \sqrt{D(x)} = \sqrt{5,981} \approx 2,446$$

$$V = \frac{\hat{\sigma}(x)}{\bar{x}} \cdot 100\% = \frac{2,446}{8,613} \cdot 100\% = 28,4\%$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§4. Числовые характеристики вариационных рядов.

Таким образом, средняя численность работников на 100 га с/х угодий по исследуемой совокупности хозяйств составила 8,61 чел. Плотность работников в среднем колебалась в промежутке $\boxed{\bar{x} \pm \sigma(x)} = 8,61 \pm 2,45$, т.е. от 6,16 до 11,06 чел. на 100 га с/х угодий.

Этот интервал, а так же коэффициент вариации показывает, что имеются большие различия в обеспечении хозяйств рабочей силой.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§4. Числовые характеристики вариационных рядов.

$$\boxed{A} = \frac{\sum (x_i - \bar{x})^3 \cdot m_i}{n \cdot \hat{\sigma}^3(x)} = \frac{25,876}{60} = 0,43$$

$$\boxed{E} = \frac{\sum (x_i - \bar{x})^4 \cdot m_i}{n \cdot \hat{\sigma}^4(x)} - 3 = \frac{156,523}{60} - 3 = -0,39$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§4. Числовые характеристики вариационных рядов.

Найденное значение коэффициента асимметрии (не достаточно близкое к нулю) указывает, что распределение не симметрично. Эксцесс также отличен от нуля, что говорит о возможном отличии распределения от нормального.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

В реальных условиях обычно бывает трудно или экономически нецелесообразно, а иногда и невозможно, исследовать всю совокупность, характеризующую изучаемый признак (генеральную совокупность). Поэтому на практике широко применяется выборочное наблюдение, когда обрабатывается часть генеральной совокупности (выборочная совокупность).

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Свойства (закон распределения и его параметры) генеральной совокупности неизвестны, поэтому возникает задача их оценки по выборке. Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была репрезентативной (представительной). Репрезентативность в силу закона больших чисел, достигается случайностью отбора.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Различают 5 основных типов выборок:

1. Собственно - случайная:

- а) повторная (элементы после выбора возвращаются обратно);*
- б) бесповторная (выбранные элементы не возвращаются).*

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

2. *Типическая* – генеральная совокупность предварительно разбивается на группы типических элементов, и выборка осуществляется из каждой.

Следует различать:

а) *равномерные* выборки (при равенстве объемов исходных групп в генеральной совокупности выбирается одинаковое количество элементов из каждой);

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

- б) пропорциональные (численность выборок формируют пропорционально численностям или средним квадратическим отклонениям групп генеральной совокупности);*
- в) комбинированные (численность выборок пропорциональна и средним квадратическим отклонениям, и численностям групп генеральной совокупности).*

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

- 3. Механическая – отбор элементов проводится через определенный интервал.*
- 4. Серийная – отбор проводится не по одному элементу, а сериями для проведения сплошного обследования.*
- 5. Комбинированная – используются различные комбинации вышеуказанных методов, например, типическая выборка сочетается с механической и собственно случайной.*

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

После осуществления выборки возникает задача оценки числовых характеристик генеральной совокупности по элементам выборочной совокупности. Различают точечные и интервальные оценки.

***Определение 1.** Точечной оценкой характеристики генеральной совокупности называется число, определяемое по выборке.*

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Пусть $\hat{\Theta}$ — выборочная характеристика, вычисленная по n результатам n наблюдений величины X , используемая в качестве оценки Θ — характеристики генеральной совокупности (в качестве Θ может быть $M(X); D(X)$ и т.д.).

Качество оценки $\hat{\Theta}$ устанавливается по 3-м свойствам:

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

1) **Состоятельность.** Оценка $\overset{\wedge}{\Theta}$ является состоятельной оценкой генеральной совокупности Θ , если для любого $\varepsilon > 0$ выполняется неравенство:

$$\lim_{n \rightarrow \infty} P\left(\left|\overset{\wedge}{\Theta}_n - \Theta\right| < \varepsilon\right) = 1$$

Это означает, что при увеличении объема выборки n выборочная характеристика стремится к соответствующей характеристике генеральной совокупности

$$\left(\overset{\wedge}{\Theta}_n \rightarrow \Theta\right)$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

- 2) **Несмещенность.** Оценка $\hat{\Theta}$ генеральной характеристики Θ называется несмещенной, если для любого фиксированного числа наблюдений n выполняется равенство:

$$M \left(\hat{\Theta}_n \right) = \Theta$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

3) **Эффективность.** Несмещенная оценка $\overset{\Delta}{\Theta}$ генеральной характеристики Θ называется несмещенной эффективной, если среди всех подобных оценок той же характеристики она имеет наименьшую дисперсию:

$$D\left(\overset{\Delta}{\Theta}_n\right) \rightarrow \min$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Статистики \bar{X} и \hat{P}_i являются состоятельными, несмещенными и эффективными характеристиками математического ожидания $M(X)$ и вероятности P соответственно.

Выборочная дисперсия $\hat{D}(X) = \hat{\sigma}^2(X)$ не обладает свойством несмещенности.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

На практике используют *исправленную выборочную дисперсию* S^2 , которая является несмещенной оценкой дисперсии генеральной совокупности:

$$S^2 = \frac{n}{n-1} \cdot \sigma^2(x) = \frac{n}{n-1} \cdot \frac{\sum (x_i - \bar{x})^2 \cdot m_i}{n} = \frac{\sum (x_i - \bar{x})^2 \cdot m_i}{n-1} \Rightarrow$$

$$S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot m_i}{n-1} \quad , \text{ где}$$

S – стандартное отклонение.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Кроме того, в расчетах используют *стандартную ошибку выборки*:

$$S_x = \frac{S}{\sqrt{n}}$$

Определение 2. *Интервалной* называют оценку, которая определяется двумя числами - границами интервала.

Интервальная оценка позволяет ответить на вопрос: внутри какого интервала, и с какой вероятностью находится неизвестное значение оцениваемого параметра генеральной совокупности?

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Пусть $\hat{\Theta}$ точечная оценка параметра Θ . Чем меньше разность $\hat{\Theta} - \Theta$, тем точнее и лучше оценка. Обычно говорят о **доверительной вероятности** $p=1-\alpha$, с которой Θ будет находиться в интервале

$$\hat{\Theta} - \Delta < \Theta < \hat{\Theta} + \Delta, \text{ где}$$

$\Delta(\Delta > 0)$ - предельная ошибка выборки, которая может быть задана наперед, либо вычислена;
 α - риск или уровень значимости (вероятность того, что неравенство будет неверным).

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

В качестве $(1-\alpha)$ принимают значения 0,90;0,95; 0,99;0,999. Доверительная вероятность показывает, что в $(1-\alpha) \cdot 100\%$ случаев оценка будет покрываться указанным интервалом.

Точечная оценка математического ожидания $M(X)=a$ определяется как средняя арифметическая

$$\bar{x} = \frac{1}{n} \sum x_i \cdot m_i$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Точечная оценка вероятности p_i определяется как относительная частота:

$$p_i = \frac{m_i}{n}$$

Для построения доверительного интервала параметра a - математического ожидания нормального распределения, составляют выборочную характеристику (статистику), функционально зависимую от наблюдений и связанную с a , например:

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

1. для *повторного отбора*:

$$u = \frac{\bar{x} - a}{\frac{\sigma(x)}{\sqrt{n}}}$$

Статистика u распределена по нормальному закону распределения с математическим ожиданием $a=0$ и средним квадратическим отклонением $\sigma(x)=1$.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Отсюда:

$P(|u| < U_{\alpha/2}) = 1 - \sigma(x)$ или, $2\Phi(U_{\alpha/2}) = 1 - \sigma(x)$ где
 Φ - функция Лапласа.

$U_{\alpha/2}$ - квантиль нормального закона распределения,
соответствующая уровню значимости α .

Доверительный интервал для параметра a :

$$\bar{x} - U_{\alpha/2} \cdot \frac{\sigma(x)}{\sqrt{n}} < a < \bar{x} + U_{\alpha/2} \cdot \frac{\sigma(x)}{\sqrt{n}}$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

2. Для *бесповторного отбора*:

Доверительный интервал для средней:

$$\bar{x} - \Delta_{\bar{x}} < \bar{x}_0 < \bar{x} + \Delta_{\bar{x}} \quad , \text{ где}$$

\bar{x} - *выборочная средняя;*

\bar{x}_0 - *средняя генеральной совокупности;*

$\Delta_{\bar{x}}$ - *предельная ошибка выборки для средней.*

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Предельная ошибка выборки:

$$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}, \text{ где}$$

t - квантиль нормального закона распределения
(при $\alpha=0,05$; $t=1,96$);

N - объем генеральной совокупности;

n - объем выборки;

*S*² - исправленная выборочная дисперсия.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Определение 3. Квантилем или нормированным отклонением называется отношение предельной ошибки к средней ошибке.

$$t = \frac{\Delta_x^-}{M_{x_0}^-}, \text{ где } M_{x_0}^- = \frac{\sigma(x)}{\sqrt{n}}$$

Квантиль вычисляется по соответствующему уровню значимости α (при $n \geq 30$, t - квантиль нормального закона распределения, при $n < 30$, t - квантиль распределения Стьюдента).

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Существуют таблицы значений для t в зависимости от уровня значимости α .

Важной является задача определения объема выборочной совокупности n при заданном уровне значимости. В случае бесповторного отбора необходимый объем выборки определяется по формуле:

$$n = \frac{t^2 \cdot S^2 \cdot N}{t^2 \cdot S^2 + \Delta_{\frac{2}{x}}^2 \cdot N}$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Пример: По условию задачи 2. При уровне значимости $\alpha=0,05$ определить:

- 1) несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения;*
- 2) доверительный интервал для математического ожидания с доверительной вероятностью $(1-\alpha)$;*
- 3) объем выборки, при котором с доверительной вероятностью $(1-\alpha)$ предельная ошибка выборки уменьшится в 2 раза, при сохранении уровня остальных характеристик.*

Учитывая, что проводилась 10% случайная бесповторная выборка.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Решение.

1) Несмещенной оценкой $M(x)$ является выборочная средняя \bar{X} $\bar{X} = 8,613$

Несмещенной оценкой $D(x)$ является исправленная выборочная дисперсия S^2

$$S^2 = \sigma^2(x) \cdot \frac{n}{n-1} = \frac{5,981 \cdot 60}{59} = 6,082$$

Несмещенной оценкой $\sigma(x)$ является стандартное отклонение S :

$$S = \sqrt{S^2} = \sqrt{6,082} = 2,466$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

2) Средняя численность работников на 100 га с/х угодий = 8,61. Найдем доверительный интервал для средней:

$$\bar{x} - \Delta_{\bar{x}} < x_0 < \bar{x} + \Delta_{\bar{x}}$$

$$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$$

- предельная ошибка выборки для средней.

При уровне значимости $\alpha=0,05$ квантиль нормального распределения $t=1,96$.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

Учитывая, что проводилась 10% выборка,
 $N=10 \cdot 60=600 \Rightarrow$

$$\Delta_{\bar{x}} = 1,96 \cdot \sqrt{\frac{6,082}{60} \left(1 - \frac{60}{600}\right)} = 0,592$$

Значит, с доверительной вероятностью $1-\alpha=0,95$, можно утверждать, что средняя численность работников на 100 га с/х угодий во всей совокупности хозяйств находится в границах

$$\bar{x} \pm \Delta_{\bar{x}} = 8,61 \pm 0,592$$

, т.е. от 8,021 до 9,205.

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

3) Необходимый объем выборки, для того, чтобы предельная ошибка не превышала $0,5 \cdot \Delta_{\bar{x}}$ при заданном уровне значимости $\alpha=0,05$ в случае случайного бесповторного отбора, определяется по формуле:

$$n = \frac{t^2 \cdot S^2 \cdot N}{t^2 \cdot S^2 + \Delta_{\bar{x}}^2 \cdot N}$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§5. Выборочный метод.

$$\Delta_{\bar{x}}^2 = (0,5 \cdot \Delta_{\bar{x}})^2 = (0,5 \cdot 0,592)^2 = (0,296)^2$$

$$n = \frac{(1,96)^2 \cdot 6,082 \cdot 600}{(1,96)^2 \cdot 6,082 + (0,296)^2 \cdot 600} = \frac{14018,766}{23,365 + 52,576} = 185$$

Значит, для уменьшения предельной ошибки в два раза объем совокупности необходимо увеличить в 3 раза.

Математическая статистика

Глава 2. Элементы теории корреляции.

Известно, что процессы, протекающие в растениях и живых организмах, обусловлены влиянием большого числа взаимосвязанных факторов, среди которых имеются главные, определяющие **основные** свойства и характеристики процесса или явления, и **второстепенные**.

Как найти в виде формулы зависимость между двумя случайными величинами, полученными в результате наблюдений, если каждому значению одной величины соответствует несколько значений другой?

Математическая статистика

Глава 2. Элементы теории корреляции.

Как найти параметры этих формул при условии, чтобы они отражали сущность изучаемого процесса и «сглаживали» влияние случайных, не характерных для данного процесса факторов? Насколько сильно влияет изменение одной величины на изменение другой? Ответы на эти вопросы составляют содержание настоящей главы.

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

В сельскохозяйственных науках, в отличие от точных наук, полные (точные) функциональные связи встречаются редко, так как возможность искусственной изоляции влияния других факторов на изучаемые признаки в большинстве случаев неосуществима.

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

Например, связь урожайность - удобрения, имеется, но есть еще ряд факторов, влияющих на урожайность (севообороты, семена, предшественники, агротехника - субъективные факторы; метеорологические факторы- объективные).

Поэтому связь урожайность - удобрения неполная функциональная связь. Эту связь называют корреляционной (англ. correlation - соотношение, соответствие).

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

Метод корреляции применяется для того, чтобы при сложном взаимодействии посторонних влияний выяснить, какова была бы зависимость между результатом и фактором, если бы посторонние причины (факторы) не изменялись и своим изменением не искажали основную зависимость.

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

Первая задача корреляции: выявление на основе наблюдений над большим количеством фактов того, как изменяется в среднем результирующий признак в связи с изменением данного фактора (парная корреляция) или группы факторов (множественная корреляция). Эта задача решается нахождением уравнения связи.

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

Вторая задача корреляции: определение степени влияния искажающих факторов. Эта задача решается при помощи различных показателей тесноты связи: коэффициента корреляции, корреляционного отношения.

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

Определение 1. Процесс нахождения связи между признаками называется **выравниванием**.

Выравнивание ведет к нахождению переменной средней \bar{y} , исчисленной в предположении функциональной зависимости y от x , т.е. $\bar{y}_x = f(x)$, и называется **уравнением регрессии**.

Математическая статистика

Глава 2. Элементы теории корреляции.

§1. Понятие о корреляции.

*При изучении влияния одних признаков на другие выделяются два признака - **факториальный** и **результативный**. Выделение этих признаков осуществляется путем логического анализа.*

Например, в связи урожайность - осадки, урожайность - результативный признак, а осадки - факториальный.

Математическая статистика

Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

Графическое изображение связи изучаемых явлений позволяет не только установить наличие или отсутствие связи между ними, но и изучить характер этой связи (форму связи и тесноту связи).

Если имеются числовые характеристики факториальных и результативных признаков одного и того же явления, то каждую пару чисел можно изобразить графически, откладывая по оси абсцисс - факториальный признак, по оси ординат - результативный признак.

Математическая статистика

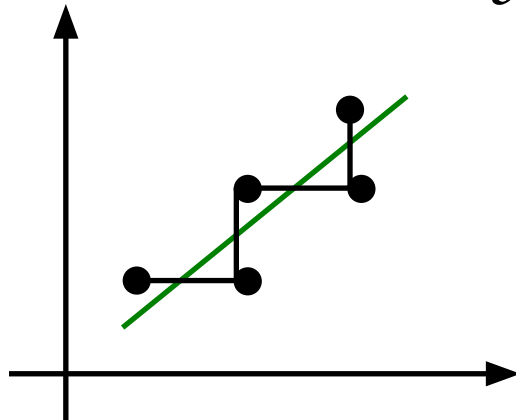
Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

Ломаная, соединяющая эти точки, называется *ломаной регрессии*.

По форме этой ломаной приближенно определяют вид зависимости.

1) Если из графика видно, что связь близка к прямой, то уравнение регрессии пишется в виде:



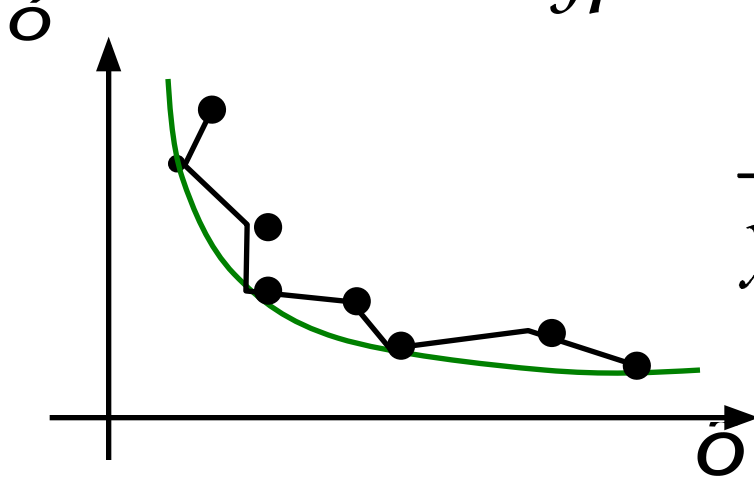
$$\bar{y} = ax + b$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

2) Если экспериментальные данные располагаются так, что через них можно провести гиперболу, то можно ожидать уравнение в виде:



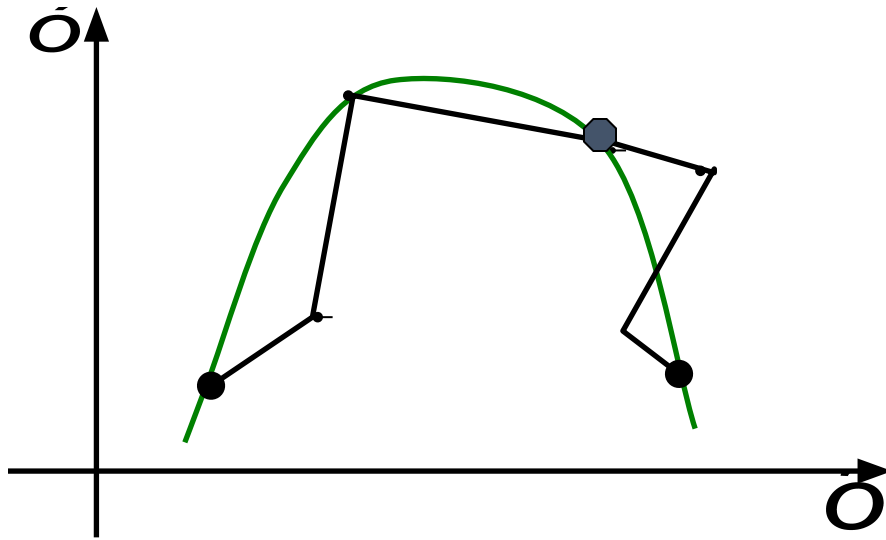
$$\bar{y} = \frac{k}{x}; \quad \bar{y} = \frac{a}{x+b}, \quad \bar{y} = \frac{a}{x+b} + c$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

3) Если кривая имеет *max* или *min*, то зависимость определяется по уравнению:



$$\bar{y} = ax^2 + bx + c$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

Для выявления функциональных зависимостей и определения неизвестных коэффициентов этой зависимости можно воспользоваться **методом наименьших квадратов**.

$$\left\{ \begin{array}{l} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i + b \cdot n = \sum_{i=1}^n y_i \end{array} \right. \Rightarrow y = ax + b$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

$$\left\{ \begin{array}{l} a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x x_i^3 \neq \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i^2 \cdot \\ a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i = \sum_{i=1}^n x_i \cdot y_i \quad \Rightarrow \quad y = ax^2 + bx + c \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i + c \cdot n = \sum_{i=1}^n y_i \end{array} \right.$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

После того, как уравнение регрессии найдено, находят так называемый **коэффициент корреляции**. Он используется для оценки тесноты связи между величинами при прямолинейной зависимости. Обозначается буквой **r** и определяется по формуле:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \text{ где}$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

\bar{X} - среднее значение факториального (причинного)

признака $\bar{x} = \frac{\sum x_i}{n}$

\bar{Y} - среднее значение результативного

признака $\bar{y} = \frac{\sum y_i}{n}$

Промежуточные вычисления удобно располагать в виде таблицы:

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

Величина коэффициента корреляции находится в пределах $-1 \leq r \leq 1$:

- 1) Чем ближе $|r|$ к 1, тем теснее связь между факториальным и результативным признаками.
- 2) при $|r|=1$ получается полная функциональная связь.
- 3) если $|r| \rightarrow 0$, то связь между признаками слабая.

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

- 4) при $|r|=0$ связи между признаками нет (линейная зависимость отсутствует).
- 5) при $r>0$ зависимость между признаками прямая (возрастающая).
- 6) при $r<0$ зависимость обратная (убывающая).

Если зависимость между признаками прямая, то можно пользоваться уравнением прямой регрессии:

$$y - \bar{y} = b_{y/x} (x - \bar{x}) \quad , \text{ где}$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

$b_{y/x}$ - коэффициент регрессии, который определяется по формуле:

$$b_{y/x} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

Задача: Для 10 петушков леггорнов 15 дневного возраста были получены следующие данные о весе их тела (x) в граммах и весе гребня (y) (в мг):

x_i	83	72	69	90	90	95	95	91	75	70
y_i	56	42	18	84	56	107	90	68	31	48

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

Требуется:

- 1) найти коэффициент корреляции и сделать вывод о тесноте и направлении линейной корреляционной связи между признаками;*
- 2) составить уравнение прямой регрессии;*
- 3) нанести на чертеж исходные данные и построить прямую регрессии.*

Решение:

Составим вспомогательную таблицу

N_o	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	83	56	0	0	-4	16	0
2	72	42	-11	121	-18	324	198
3	69	18	-14	186	-42	1764	588
4	90	84	7	49	24	576	168
5	90	56	7	49	-4	16	-28
6	95	107	12	144	47	2209	564
7	95	90	12	144	30	900	360
8	91	68	8	64	8	64	64
9	75	31	-8	64	-29	841	232
10	70	48	-13	169	12	144	156
Σ	830	600	0	990	0	6854	2302

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

Вычисляем средние:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{830}{10} = 83 \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{600}{10} = 60$$

1) найдем коэффициент корреляции:

$$r = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}} \quad r = \frac{2302}{\sqrt{990 \cdot 6854}} = 0,88$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

Вывод: между весом тела x и весом гребня y у 15-дневных петушков существует тесная положительная линейная корреляционная связь.

2) найдем коэффициент регрессии:

$$b_{y/x} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad b_{y/x} = \frac{2302}{990} \approx 2,32$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

Подставим в уравнение прямой регрессии:

$$y - \bar{y} = b_{y/x} (x - \bar{x})$$

$$y - 60 = 2,32(x - 83)$$

$$\underline{y = 2,32x - 132,56}$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§3. Коэффициент корреляции.

3) наносим исходные данные на координатную плоскость и строим найденную прямую регрессии.

