АРХАНГЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ МНОГОПРОФИЛЬНЫЙ КОЛЛЕДЖ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ АРХАНГЕЛЬСКОЙ ОБЛАСТИ
государственное бюджетное профессиональное образовательное учреждение
Архангельской области «Архангельский государственный многопрофильный колледж»

ЕН.01 МАТЕМАТИКА

ТЕМА 04. ПРОСТЕЙШИЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. СТАТИСТИЧЕСКАЯ ОБРАБОТКА ИНФОРМАЦИИ И РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Математическая статистика — наука, разрабатывающая математические методы систематизации и использования статистических данных для научных и практических выводов.

Статистические данные представляют собой данные, полученные в результате исследования большого числа объектов или явлений.

В математической статистике можно выделить два направления: описательную статистику и индуктивную статистику (статистический вывод). Описательная статистика занимается накоплением, систематизацией и представлением опытных данных в удобной форме. Индуктивная статистика на основе этих данных позволяет сделать определенные выводы относительно объектов, о которых собраны данные, или оценки их параметров.

Пусть требуется изучить совокупность однородных объектов относительно некоторого качественного или количественного признака, характеризующего эти объекты. **Выборочной совокупностью** или просто выборкой называют совокупность случайно отобранных объектов.

Генеральной совокупностью называют совокупность объектов, из которых производится выборка. Объёмом совокупности (выборочной или генеральной) называют число объектов этой совокупности. В математической статистике принято правило: даже если генеральная совокупность конечна, все равно она считается бесконечной по объёму.

К выборке предъявляют основное требование – выборка должна быть репрезентативна – правильно представлять генеральную совокупность. Для репрезентативности выборки требуется, чтобы:

- отбор был случайным;
- все объекты генеральной совокупности должны иметь одинаковую вероятность попасть в выборку.

Вариационный и статистический ряды

Выборка является труднообозримым множеством. Для дальнейшего изучения выборку подвергают перегруппировке.

<u>Определение:</u> Вариационным рядом называется последовательность всех элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются.

Запись вариационного ряда: $x_1, x_2, x_3, \dots, x_n$. Ему соответствует следующая таблица:

i	1	2	3	…	n
x_i	x_1	x_2	x_3	…	x_n

Элементы вариационного ряда x_i называют его **вариантами** или порядковыми статистиками.

<u>Пример 2:</u> Студенты получили следующие баллы по тесту: 11, 8, 9, 10, 8, 6, 7, 7, 9, 11, 10, 6, 5, 11, 10. Записать статистический и вариационный ряды.

Р е ш е н и е :

11, 8, 9, 10, 8, 6, 7, 7, 9, 11, 10, 6, 5, 11, 10 – это статистический ряд.

Расположим данные в порядке возрастания:

5, 6, 6, 7, 7, 8, 8, 9, 9, 10, 10, 10, 11, 11, 11 – это вариационный ряд.

Представим данный ряд в виде таблицы (с учетом повторений) и в порядке возрастания значений признака, получим **ранжированный вариационный ряд**.

x_i	5	6	7	8	9	10	11
n_i	1	2	2	2	2	3	3

Здесь x_i – значение признака (варианта), n_i – его частота («вес» значения признака, количество повторений). Сумма всех частот значений признака равна объему выборки: n = 1 + 2 + 2 + 2 + 2 + 3 + 3 = 15.

Дискретный статистический ряд

Вариационный ряд называется *дискретным*, если любые его варианты отличаются на конечную постоянную величину, и называется *непрерывным (или интервальным),* если его варианты могут отличаться друг от друга на сколь угодно малую величину.

Определение: Дискретным статистическим рядом называется последовательность различных вариант x_i с указанием частот повторения элементов. При этом вместо абсолютных частот n_i можно задавать распределение относительных частот ω_i.

$\boldsymbol{\omega_i = \frac{n_i}{n}}$ – относительные частоты значения выборки

где n – объем выборки.

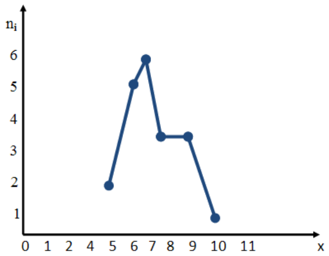
Дискретный статистический ряд (выборочное распределение) можно записать в виде таблицы:

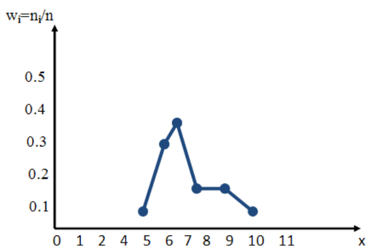
x_i	x_1	x_2	…	x_n
ω_i	ω_1	ω_2	…	ω_n

Для наглядного представления выборки часто используют различные графические изображения. Простейшими графическими изображениями являются **полигон** и **гистограмма** выборки.

Полигон (для дискретной случайной величины) - ломаная, соединяющая точки (x_i, n_i - полигон частот или точки (x_i, w_i) - полигон относительных частот.

Полигон частот:





Гистограмма — ступенчатая фигура, состоящая из прямоугольников, основаниями которых являются отрезки длиной x_i-x_{i-1}, а их высоты равны:

$$\frac{n_i}{n(x_i - x_{i-1})}$$

Пример

Измерения напряжения электросети (в вольтах) дали следующие результаты: 210, 198, 215, 212 194 213 199 191, 205, 211, 189, 206, 204, 205, 201, 194, 190, 200, 202, 196, 200, 216, 214, 200, 196, 210, 206, 200, 215, 204.

Построить гистограмму относительных частот выборки и гистограмму частот выборки.

Решение.

Объем выборки n=30. Составим вариационный ряд, расположив данные выборки в возрастающем порядке: 189 190 191 194, 194, 196, 196, 198, 199, 200, 200, 200, 200, 201, 202, 204 204 105, 205, 206, 206, 210, 210, 211, 212, 213, 214, 215, 215, 216.

Размах выборки равен 216-189=27.

Гистограмма относительных частот

Определим количество интервалов, на которые необходимо разбить выборку: $k=\log_2 30+1=5,8$. Округлим это число до ближайшего целого k=6. Так как размах выборки равен 27, то длина каждого интервала Δ=27/6=4,5.

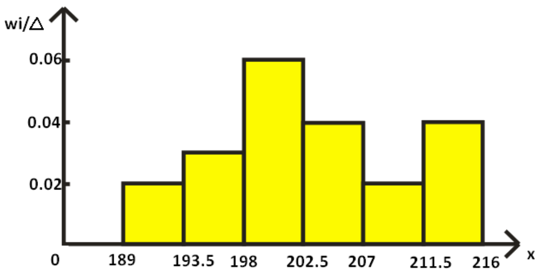
Подсчитаем, сколько измеренных значений попало в каждый из полученных интервалов:

Частичный интервал	Частота	Частичный интервал	Частота	Частичный интервал	Частота
J1=[189;193.5)	3	J3=[198;202.5)	8	J5=[207;211.5)	3
J2=[193.5;198)	4	J4=[202.5;207)	6	J6=[211.5;217]	6

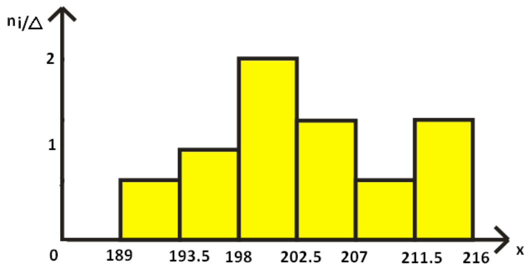
Сведем полученные данные в таблицу:

Часстичный интервал длиной Δ=4.5	Частота ni	wi=ni/n	Эмпирическая плотность распределения частоты ni/Δ	wi/Δ
[189;193.5)	3	0.1	0.66	0.02
[193.5;198)	4	0.13	0.89	0.03
[198;202.5)	8	0.27	1.78	0.06
[202.5;207)	6	0.2	1.31	0.04
[207;211.5)	3	0.1	0.66	0.02
[211.5;217]	6	0.2	1.31	0.04

Гистограмма относительных частот



Гистограмма частот



Числовые характеристики вариационных рядов

Среднее арифметическое **(*M*) –** одна из основных характеристик выборки. Этот показатель характеризуется тем, что сумма отклонений от него выборочных значений (с учетом знака) равна нулю.

Для вычисления среднего арифметического сумму всех значений признака делим на объем выборки.

Пример: x_i : 20, 15, 15, 20, 30, среднее арифметическое равно 20. При этом сумма отклонений вариант от среднего арифметического равна нулю: сумма отклонений= 0 +(-5) + (-5) + 0 + 10 = 0.

M=(20+15+15+20+30)/5=20

Следует заметить, что среднее арифметическое измеряется в тех же единицах, что и признак. Например, если масса человека измеряется в кг, то и среднее арифметическое измеряется в кг.

Среднее арифметическое, вычисленное на основе выборочных данных, то есть данных, полученных на выборке, называется **выборочным средним арифметическим**. Оно обозначается как М. Среднее арифметическое генеральной совокупности называется генеральным средним. Оно обозначается буквой мю (μ).

Мода (Мо) – характеристика положения. Представляет собой значение признака, встречающееся в выборке наиболее часто.

В качестве примера рассмотрим выборку: x_i :3; 3; 3; 5; 5; 3; 4; 6; 7; 5; 3.

В выборке цифра «3» встречается 5 раз, поэтому Мо = 3.

Медиана (Ме)- характеристика положения, представляет собой такое значение признака, при котором одна половина значений меньше ее, а другая – больше.

В качестве примера рассмотрим выборку: x_i :3; 3; 3; 5; 5; 3; 4; 6; 7; 5; 3.

Чтобы легко было определить медиану расположим варианты па возрастанию.

x_i :3; 3; 3; 3; 3; **4;** 5; 5; 5; 6; 7. Варианта со значением «4» стоит в середине этой выборки. Это и есть медиана.

__Определение:__ **Выборочной дисперсией** называют среднее арифметическое квадратов отклонений значений выборки от выборочного среднего:

$$D = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \cdots + (x_n - \bar{X})^2}{n}$$

Если выборка задана статистическим (ранжированным) рядом, то выборочную дисперсию можно найти по формуле:

$$\boldsymbol{D = \frac{n_1(x_1 - \bar{X})^2 + n_2(x_2 - \bar{X})^2 + \cdots + n_i(x_i - \bar{X})^2}{n}}$$

__Пример :__ При измерении роста девушек некоторого института была получена следующая выборка: 179, 160, 155, 183, 155, 153, 167, 186, 163, 155, 157, 175, 170, 166, 160, 173, 182, 167, 171, 169, 179, 165, 156, 179, 158, 171, 175, 173, 165, 171. Сделайте статистический анализ данной выборки.

Р е ш е н и е :

Вариационный ряд – 153, 155, 155, 155, 156, 157, 158, 160, 160, 163, 165, 165, 166, 167, 167, 169, 170, 171, 171, 171, 173, 173, 175, 175, 179, 179, 179, 182, 183, 186.

Объем выборки: $n = 30$.

Размах выборки: $R = x_{max} - x_{min} = 186 - 153 = 33$.

Медиана: $med = \frac{167+169}{2} = 168$, т.к. ряд четный и для нахождения медианы мы должны взять два числа в середине этого ряда, т.е. числа, стоящие на 15 и 16 местах, и найти среднее арифметическое.

Мода: $mod = 155; 171; 179.$

Ранжированный вариационный ряд:

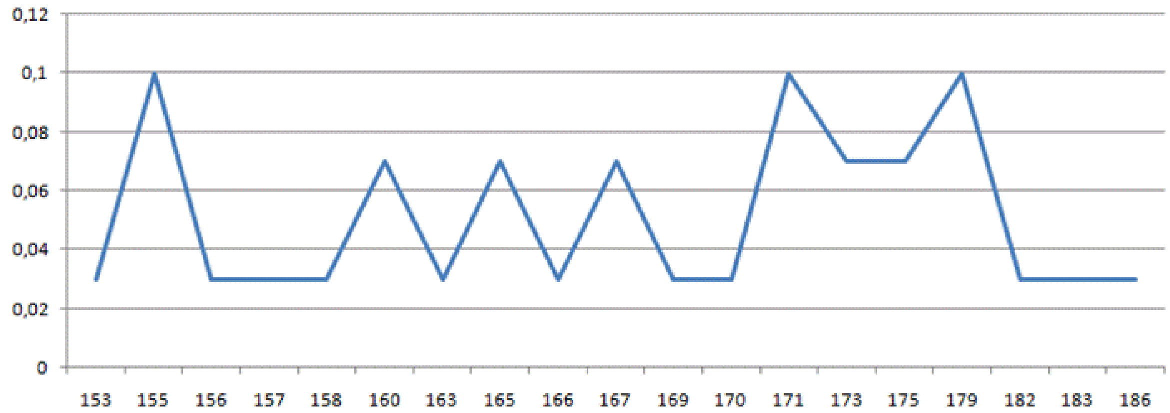
x_i	153	155	156	157	158	160	163	165	166	167	169	170	171	173	175	179	182	183	186

n_i	1	3	1	1	1	2	1	2	1	2	1	1	3	2	2	3	1	1	1

Выборочное распределение:

x_i	153	155	156	157	158	160	163	165	166	167	169	170	171	173	175	179	182	183	186
ω_i	0,03	0,1	0,03	0,03	0,03	0,07	0,03	0,07	0,03	0,07	0,03	0,03	0,1	0,07	0,07	0,1	0,03	0,03	0,03

Полигон относительных частот:



Среднее значение выборки:

$$\bar{X} = \frac{x_1 n_1 + x_2 n_2 + \dots + x_i n_i}{n} =$$

$$= \frac{153 + 155 \cdot 3 + 156 + 157 + 158 + 160 \cdot 2 + 163 + 165 \cdot 2 + 166 + 167 \cdot 2 + 169 + 170 + 171 \cdot 3 + 173 \cdot 2 + 175 \cdot 2 + 179 \cdot 3 + 182 + 183 + 186}{30} =$$

$$= \frac{5038}{30} \approx 167{,}9$$

Выборочная дисперсия:

$$D = \frac{n_1(x_1 - \bar{X})^2 + n_2(x_2 - \bar{X})^2 + \dots + n_i(x_i - \bar{X})^2}{n} =$$

$$= \frac{(153 - 167{,}9)^2 + 3 \cdot (155 - 167{,}9)^2 + (156 - 167{,}9)^2 + \dots + (186 - 167{,}9)^2}{30} =$$

$$= \frac{2561{,}9}{30} \approx 85{,}4$$

Список литературы по теме:

1. Кацман, Ю. Я. Теория вероятностей и математическая статистика. Примеры с решениями : учебник для среднего профессионального образования / Ю. Я. Кацман. — Москва : Издательство Юрайт, 2022. — 130 с. — (Профессиональное образование). — ISBN 978-5-534-10083-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/490334
2. Энатская, Н. Ю. Теория вероятностей и математическая статистика : учебник и практикум для среднего профессионального образования / Н. Ю. Энатская, Е. Р. Хакимуллин. — Москва : Издательство Юрайт, 2022. — 399 с. — (Профессиональное образование). — ISBN 978-5-534-11917-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/489852