МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

И.Л. Каширина, К.В. Чудинова

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебно-методическое пособие

Воронеж
Издательский дом ВГУ
2019

Оглавление

генеральной совокупности извлечена выборка объёма n наблюдаемых значений этой случайной величины, причём значение x_1 наблюдается в выборке n_1 раз, $x_2 - n_2$ раз, … $x_k - n_k$ раз, при этом $\sum_{i=1}^{k} n_i = n$.

Различные наблюдаемые значения x_i случайной величины X называют *вариантами*. Последовательность вариант, записанная в порядке неубывания, называется *вариационным рядом*. Количество повторений n_i называется частотой варианты x_i. А отношение $\frac{n_i}{n}$ *относительной частотой* (или частостью) варианты x_i. *Статистическим распределением выборки* называется перечень вариант и соответствующих им частот (или относительных частот).

Пример 1: Рассмотрим выборку, составленную из оценок 20 студентов на экзамене по теории вероятностей.

1) Выборка: 5,4,3,3,2,4,4,4,5,4,5,4,3,3,2,4,4,4,5,4.

2) Вариационный ряд: 2, 2,3,3,3,3,4,4,4,4,4,4,4,4,4,4,5,5,5,5.

3) Статистический ряд (распределение выборки):

x_i	2	3	4	5
n_i	2	4	10	4

Обозначим через n_x – число наблюдений в выборке, в которых наблюдаемое значение признака было меньше, чем x. *Эмпирической функцией распределения* называют функцию $F^*(x) = \frac{n_x}{n}$. В отличие от эмпирической функции распределения выборки $F^*(x)$, функцию распределения *F(x)* случайной величины *X*, вычисленную по всей генеральной совокупности, называют теоретической функцией распределения. Эмпирическая функция распределения используется для оценки теоретической функции распределения генеральной совокупности. При возрастании объема выборки различия между *F*(x)* и *F(x)* уменьшаются.

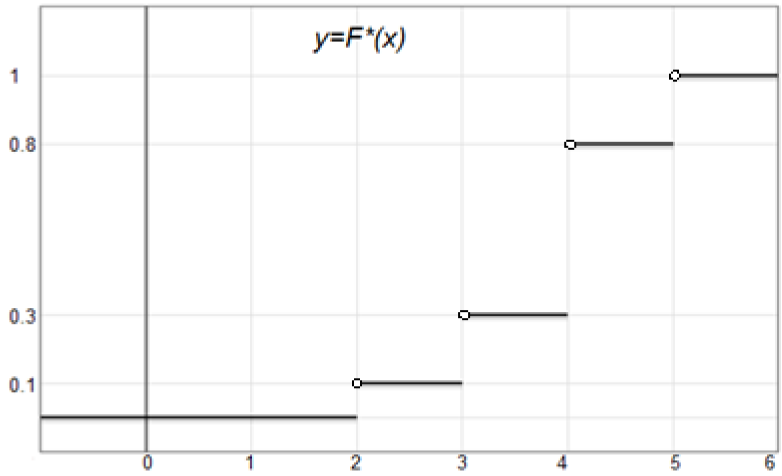

Рис. 1. График эмпирической функции распределения для примера 1.

Свойства эмпирической функции распределения аналогичны свойствам теоретической функции распределения случайной величины:

1) $F^*(x)$ – неубывающая функция;
2) $0 \leq F^*(x) \leq 1$;
3) если x_1 – наименьшая варианта, а x_k – наибольшая, то $F^*(x)=0$ при $x \leq x_1$ и $F^*(x)=1$ при $x>x_k$.

Полигон частот, гистограмма частот

Полигоном частот называют ломаную линию, которая соединяет точки (x_1, n_1); (x_2, n_2); ... (x_k, n_k).

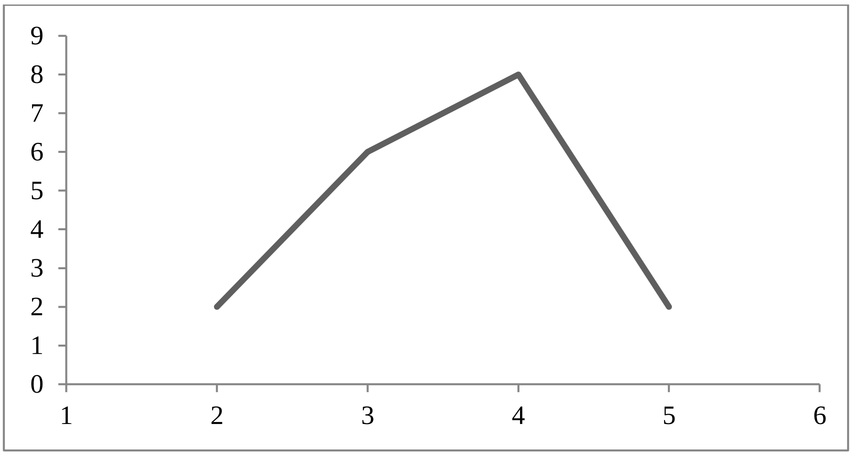

Рис. 2 Полигон частот для примера 1.

Полигоном относительных частот (частостей) называют ломаную, которая соединяет точки (x_i, ω_i); где $\omega_i = \frac{n_i}{n}$.

В тех случаях, когда различных значений наблюдаемых значений признака слишком много (когда значение признака – непрерывная случайная величина), принято строить не полигон частот, а гистограмму частостей. Для этого значение признака разбивают на несколько интервалов и для каждого интервала подсчитывают сумму частот вариант, попавших в этот интервал. *Гистограммой* относительных частот (частостей) называют ступенчатую фигуру, составленную из прямоугольников, основаниями которых служат интервалы длиной h_i, а высотой $\frac{n_i}{nh_i}$. Значения h_i чаще всего выбирают одинаковые (но это не обязательно).

Формула Стерджесса: рекомендуемое число интервалов определяется по формуле $m = [1 + 3.3221 \lg n]$, где [.] – целая часть. Величина интервала $\Delta = \frac{x_i^{max} - x_i^{min}}{1 + 3.3221 \lg n}$.

Пример 2 : Рассмотрим выборку, составленную из роста 50-ти случайно отобранных студентов-юношей факультета ПММ.

1) Выборка: 173, 175, 180,182, 170, 168, 177, 187, 180, 176, 182, 164, 170, 174, 188, 177, 168, 176, 178, 185, 172, 174, 180, 166, 172, 178, 190, 174, 177, 178, 175,170, 172, 189, 180, 192, 166,187, 180, 177, 184, 180, 176, 175, 184, 169, 193, 171, 177.

Согласно формуле Стерджесса *m=6, Δ=4.3.* С учетом того, что рост измерялся с точностью до целого значения, округлим Δ до 4-х. Получим:

x_i	[164;169)	[169;174)	[174;179)	[179;184)	[184;189)	[189;194)
n_i	6	10	16	9	6	3

Посчитаем частости по формуле $w_i = \frac{n_i}{n\Delta}$, с учетом *n=50, Δ=4,* и построим гистограмму (рис. 3)

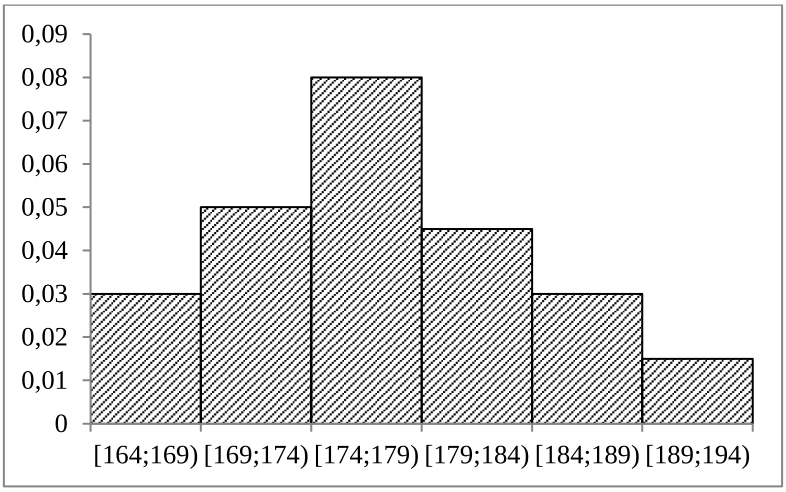

Рис. 3. Гистограмма относительных частот для примера 2.

Числовые характеристики статистического распределения выборки

Выборочной средней называют среднее арифметическое значение всех элементов выборки: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Если известно статистическое распределение выборки, то применяют сокращённую формулу:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i.$$

Если статистический ряд представлен в интервальной форме, тогда в качестве x_i берётся середина каждого интервала.

Свойства выборочной средней:

1. Если все элементы выборки увеличить (или уменьшить) в одно и то же число раз, то выборочная средняя увеличится (или уменьшится) в такое же число раз: $\overline{kx} = k\bar{x}$.

2. Если все элементы выборки увеличить (или уменьшить) на одно и то же число, то выборочная средняя увеличится (или уменьшится) на это число: $\overline{c + x} = c + \bar{x}$.

3. Выборочная средняя суммы (или разности) нескольких признаков равна сумме (или разности) выборочных средних этих признаков: $\overline{x \pm y} = \bar{x} \pm \bar{y}$.

4. Среднее отклонение от выборочной средней равно 0: $\overline{x-\bar{x}}=0$.

5. Если объединить несколько выборок, то выборочная средняя объединённого ряда является средним значением групповых средних, причём весами являются объёмы выборок: $\bar{x}=\frac{1}{n}\sum_{i=1}^{k}\bar{x_\iota}n_i$, где $\bar{x_\iota}$- выборочная средняя i-й выборки, $n=\sum_{i=1}^{k}n_i$.

Выборочная мода $M_в$– это варианта с наибольшей частотой.

Выборочная медиана – значение признака, приходящееся на середину вариационного ряда. Если объём выборки n – чётное число, то медиана $Me=\frac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}$, если n – нечётное число, то $Me=x_{\frac{n+1}{2}}$.

Вариационный размах R – разность между последним и первым элементом вариационного ряда.

Выборочная дисперсия – среднее значение квадратов отклонения вариант от выборочной средней: $D_в=\frac{1}{n}\sum_{i=1}^{k}(x_i-\bar{x})^2n_i$.

Свойства выборочной дисперсии:

1. Дисперсия признака, имеющего постоянное значение, равна 0.

2. Если все варианты увеличить или уменьшить в k раз, то выборочная дисперсия увеличится или уменьшится в k^2 раз.

$$D_в(kx)=k^2D_в(x)$$

3. Если все варианты увеличить или уменьшить на одно и то же число, то выборочная дисперсия не изменится: $D_в(x\pm c)=D_в(x)$

4. Сокращённая формула для вычисления выборочной дисперсии имеет вид: $D_в=\frac{1}{n}\sum_{i=1}^{k}x_i^2n_i-\bar{x}^2=\overline{x^2}-(\bar{x})^2$.

5. Если выборка состоит из нескольких групп наблюдения, то общая дисперсия равна сумме среднего значения групповых дисперсий и межгрупповой дисперсии: $D_в=\overline{D}_в^i+\Delta^2$, где $\overline{D}_в^i=\frac{1}{n}\sum_{i=1}^{k}D_в^in_i$, n_i – объём i-ой группы, k – количество групп, $D_в^i$ – выборочная дисперсия в i-ой группе,