Меры разброса

Не каждую выборку имеет смысл оценивать с помощью центральных тенденций. Например, если исследуется выборка

$$80,\ 80,\ 320,\ 4600 \qquad (1)$$

годовых доходов (в тысячах рублей) четверых человек, то очевидно, что ни мода (80), ни медиана (200), ни среднее (1270) не могут выступать в роли единой объективной характеристики данной выборки. Это объясняется тем, что наименьшие значения выборки (1) существенно отличаются от наибольшего — разность наибольшего и наименьшего значений соизмерима с наибольшим значением.

Определение 1. Разность наибольшего и наименьшего значений случайной величины выборки называется её *размахом* и обозначается R.

Так, для выборки (1) размах $R = 4600 - 80 = 4520$. Размах показывает, как велик разброс значений случайной величины в выборке. Однако, зная только размах выборки, невозможно охарактеризовать отличие её элементов друг от друга, отличие каждого элемента от среднего значения. Возникает вопрос: как сравнить, например, две выборки, имеющие одинаковые размахи и одинаковые средние значения? Рассмотрим реальную ситуацию на примере.

На место токаря претендуют двое рабочих. Для каждого из них установили испытательный срок, в течение которого они должны были изготавливать одинаковые детали. Результаты работы претендентов представлены в таблице 12.

Каждый из рабочих за 5 дней изготовил 250 деталей, значит, средняя производительность труда за день у обоих рабочих одинаковая:

$$\overline{X} = \overline{Y} = \frac{250}{5} = 50 \text{ (дет./день).}$$

Таблица 12

День недели	Дневная выработка	
	первого рабочего (X)	второго рабочего (Y)
Понедельник	52	61
Вторник	54	40
Среда	50	55
Четверг	48	50
Пятница	46	44
	$\Sigma X = 250$	$\Sigma Y = 250$

Моды у предложенных совокупностей отсутствуют, а медианы одинаковые (50 и 50). Кого же из этих рабочих предпочтительнее взять на работу? В данном случае в качестве критерия сравнения совокупностей может выступать *стабильность* производительности труда рабочего. Её можно оценивать с помощью отклонений от среднего значения элементов совокупности.

Определение 2. ***Отклонением от среднего*** **называют разность между рассматриваемым значением случайной величины и средним значением выборки.**

Например, если значение величины $X_1 = 52$, а значение среднего $\overline{X} = 50$, то отклонение X_1 от среднего будет равно $X_1 - \overline{X} = 52 - 50 = 2$.

Очевидно, отклонение от среднего может быть как положительным, так и отрицательным числом. Нетрудно показать, что *сумма отклонений всех значений выборки от среднего значения равна нулю.* Поэтому характеристикой стабильности элементов совокупности может служить *сумма квадратов отклонений от среднего.*

Из предложенной ниже таблицы 13 видно, что у второго рабочего сумма квадратов отклонений от среднего больше, чем у первого рабочего:

$$\Sigma (X - \overline{X})^2 < \Sigma (Y - \overline{Y})^2.$$

На практике это означает, что второй рабочий имеет нестабильную производительность труда: в какие-то дни работает не в полную силу, а в какие-то

Таблица 13

День недели	Значение случайной величины		Отклонение от среднего $\overline{X} = \overline{Y} = 50$		Квадраты отклонений	
	X	Y	$X - \overline{X}$	$Y - \overline{Y}$	$(X - \overline{X})^2$	$(Y - \overline{Y})^2$
Понедельник	48	50	−2	0	4	0
Вторник	54	40	4	−10	16	100
Среда	50	55	0	5	0	25
Четверг	52	61	2	11	4	121
Пятница	46	44	−4	−6	16	36
Сумма	**250**	**250**	**0**	**0**	**40**	**282**

навёрстывает упущенное, что всегда сказывается на качестве продукции. Очевидно, что работодатель предпочтёт взять на место токаря первого рабочего (у которого сумма квадратов отклонений от средней производительности меньше).

Если бы рабочие работали разное количество дней и производили в среднем за день одинаковое число деталей, то стабильность работы каждого из них можно было бы оценить по величине *среднего арифметического квадратов отклонений.*

Такая величина называется *дисперсией* (от лат. *dispersio* — рассеяние) и обозначается буквой D.

Для случайной величины X, принимающей N различных значений и имеющей среднее значение $\overline{X}$, дисперсия находится по формуле

$$D = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \ldots + (X_N - \overline{X})^2}{N}. \quad (1)$$

Задача 1 Два токаря вытачивали одинаковые детали, причём первый трудился полную рабочую неделю, а второй по распоряжению начальника — 4 дня. Сведения об их дневной выработке представлены в таблице 14. Сравнить стабильность работы токарей.

▶ Найдём средние значения выборок данных величин X и Y:

$$\overline{X} = \frac{53 + 54 + 49 + 48 + 46}{5} = \frac{250}{5} = 50,$$

$$\overline{Y} = \frac{52 + 46 + 53 + 49}{4} = \frac{200}{4} = 50.$$

Очевидно, $\overline{X} = \overline{Y}$.

Таблица 14

День недели	Дневная выработка первого токаря (X)	Дневная выработка второго токаря (Y)
Понедельник	53	52
Вторник	54	46
Среда	49	53
Четверг	48	49
Пятница	46	—

С помощью таблицы 15 найдём суммы квадратов отклонений от средних всех значений величин X и Y.

Таблица 15

День недели	Значение случайной величины		Отклонение от среднего		Квадрат отклонения от среднего	
	X	Y	$X - 50$	$Y - 50$	$(X - 50)^2$	$(Y - 50)^2$
Понедельник	53	52	3	2	9	4
Вторник	54	46	4	−4	16	16
Среда	49	53	−1	3	1	9
Четверг	48	49	−2	−1	4	1
Пятница	46	—	−4	—	16	—
				Сумма:	46	30

$$D_X = \frac{46}{5} = 9{,}2, \text{ а } D_Y = \frac{30}{4} = 7{,}5, \text{ т. е. } D_X > D_Y.$$

Ответ Второй токарь работает стабильнее первого. ◁

Если значения $X_1, X_2, \ldots, X_k$ случайной величины X повторяются с частотами $M_1, M_2, \ldots, M_k$ соответственно, то дисперсию величины X можно вычислить по формуле

$$D = \frac{(X_1 - \overline{X})^2 M_1 + (X_2 - \overline{X})^2 M_2 + \ldots + (X_k - \overline{X})^2 M_k}{M_1 + M_2 + \ldots + M_k}, \quad (1)$$

где $\overline{X} = \dfrac{X_1 M_1 + X_2 M_2 + \ldots + X_k M_k}{M_1 + M_2 + \ldots + M_k}$.

Используя знак суммы Σ, формулу (1) можно записать компактнее:

$$D = \frac{\Sigma\left((X - \overline{X})^2 M\right)}{\Sigma M}, \text{ где } \overline{X} = \frac{\Sigma(XM)}{\Sigma M}.$$

Пусть величина X имеет некоторую размерность (например, сантиметры). Тогда её среднее значение $\overline{X}$ и отклонение от среднего $X - \overline{X}$ имеют ту же размерность, что и сама величина (в сантиметрах). Квадрат же отклонения $(X - \overline{X})^2$ и дисперсия D имеют размерности квадрата этой величины (в квадратных сантиметрах).

Для оценки степени отклонения от среднего значения удобно иметь дело с величиной той же размерности, что и сама величина X. С этой целью используют значения корня квадратного из дисперсии $\sqrt{D}$.

Определение. Корень квадратный из дисперсии называют *средним квадратичным отклонением* и обозначают σ, т. е. $\sigma = \sqrt{D}$.

Задача 2 Распределение по частотам значений величины X — числа забитых голов игроками футбольной команды за период соревнований показано в таблице 16. Найти среднее квадратичное отклонение от среднего значения числа всех забитых голов.

Таблица 16

X	0	1	2	3
M	4	2	3	1

▶ Результаты последовательных вычислений будем заносить в таблицу 17, при этом:

$$\Sigma M = 10,\ \overline{X} = \frac{\Sigma(X \cdot M)}{\Sigma M} = \frac{11}{10} = 1{,}1.$$

Таблица 17

X	0	1	2	3
M	4	2	3	1
$X - \overline{X}$	−1,1	−0,1	0,9	1,9
$(X - \overline{X})^2$	1,21	0,01	0,81	3,61
$(X - \overline{X})^2 \cdot M$	4,84	0,02	2,43	3,61

$$D = \frac{\Sigma\left((X - \overline{X})^2 \cdot M\right)}{\Sigma M} = \frac{4{,}84 + 0{,}02 + 2{,}43 + 3{,}61}{10} =$$

$$= \frac{10{,}9}{10} = 1{,}09,$$

$$\sigma = \sqrt{D} = \sqrt{1{,}09} \approx 1{,}04.$$

Ответ $\sigma \approx 1{,}04$. ◁