

**ОБРАБОТКА
ЭКСПЕРИМЕНТАЛЬНЫХ
ДАнных**

**СЧЕТНАЯ ОБРАБОТКА
ПОЛУЧЕННЫХ ДАнных
Расчет средних величин**

Актуальность темы. В практической деятельности часто возникает необходимость обобщения больших массивов числовых данных.

средних величин.

Широко используются средние величины при:

- изучении физического развития различных групп населения (средний рост, вес, окружность грудной клетки и т.д.);
- характеристике физиологического состояния органов и систем организма человека (средняя частота пульса, средняя величина артериального давления, жизненной емкости легких, среднее содержание белка крови и т.д.);
- изучении закономерностей течения различных процессов в здоровом и больном организме;
- оценке эффективности применения лекарственных препаратов;
- гигиенической характеристике внешней среды (среднее содержание пыли и газов в воздухе производственных помещений и в атмосфере, средний уровень шума, вибрации и т.д.).

Средние величины удобно сравнивать между собой и выявлять закономерности.

Различают несколько видов средних величин:

- средняя арифметическая,
- средняя геометрическая,
- средняя гармоническая,
- средняя квадратическая,
- средняя прогрессивная,
- мода,
- медиана и д.р.

К показателям разнообразия (вариации, колеблемости) относятся:

- *амплитуда (A_m),*
- *лимит (lim)*
- *среднее квадратическое отклонение (δ)*
- *дисперсия (δ^2)*
- *коэффициент вариации (C_v)*

- ***Вариационный ряд*** – это ряд числовых значений изучаемого признака, отличающихся друг от друга по своей величине и расположенных в определенной последовательности (в восходящем или убывающем порядке).
- Каждое числовое значение ряда называют **вариантой** (*V*).
- Числа, показывающие, как часто встречается та или иная варианта в составе данного ряда, называется **частотой** (*p*).
- **Общее число случаев наблюдений**, из которых вариационный ряд состоит, обозначают буквой *n*.
- Различие в значении изучаемых признаков называется ***вариацией***.

Если варьирующий признак не имеет *количественной меры*, вариацию называют качественной, а ряд распределения – **атрибутивным** (например, распределение по исходу заболевания, по состоянию здоровья и т.д.).

Если варьирующий признак имеет количественное выражение, такую вариацию называют количественной, а ряд распределения – **вариационным**.

Построить вариационный ряд – означает имеющиеся числовые значения (варианты) систематизировать, упорядочить, т.е. расположить в определенной последовательности (в восходящем или убывающем порядке) с соответствующими им частотами.

- по характеру количественного признака (прерывные и непрерывные).

Если количественный признак носит непрерывный характер, т.е. между целыми величинами имеются промежуточные дробные величины, вариационный ряд называется **непрерывным**.

Если количественный признак носит прерывный характер, т.е. отдельные его значения (варианты) отличаются друг от друга на целое число и не имеют промежуточных дробных значений, вариационный ряд называют **прерывным** или **дискретным**. В простом вариационном ряду каждая варианта встречается только один раз ($p=1$), во взвешенном – одна и та же варианта встречается несколько раз ($p>1$).

Например, у 21 студентов-медиков исследовалась частота пульса (число ударов в минуту), которая составила: 80, 66, 74, 70, 64, 80, 80, 74, 68, 70, 74, 64, 68, 68, 66, 84, 84, 80, 70, 74, 84.

Распределение студентов-медиков по частоте пульса (ул/мин)

Пульс (число ударов в минуту) (V)	Число студентов (p)
64	2
66	2
68	3
70	3
74	4
80	4
84	3
	$\Sigma=n=21$

При большом количестве наблюдений ($n > 30$) число встречающихся вариантов может быть очень большим, в этом случае составляется **интервальный** или **сгруппированный вариационный ряд**, в котором для упрощения последующей обработки и выяснения характера распределения варианты *объединены в группы*.

Обычно число групповых вариантов колеблется от 8 до 15.

Их должно быть не меньше 5, т.к. иначе это будет слишком грубое, но не более 20-25, т.к. существенно искажаются особенности варьирования признака и усложняется математическая обработка.

При составлении сгруппированного ряда необходимо учесть, что:

- группы вариант должны располагаться в определенном порядке (в восходящем или нисходящем);
- интервалы в группах вариант должны быть одинаковыми;
- значения границ интервалов не должны совпадать, т.к. неясно будет, в какие группы относить отдельные варианты;
- не рекомендуется оставлять открытых интервалов (50 лет и старше, до 0.6 мг % и т.д.).

Построим **сгруппированный (интервальный) ряд**, характеризующий данные о частоте пульса (число ударов в минуту) у 55 студентов-медиков перед экзаменом:

64, 66, 60, 62, 64, 68, 70, 66, 70, 68, 62, 68, 70, 72, 60, 70, 74, 62, 70, 72, 72, 64, 70, 72, 76, 76, 68, 70, 58, 76, 74, 76, 76, 82, 76, 72, 76, 74, 79, 78, 74, 78, 74, 78, 74, 74, 78, 76, 78, 76, 80, 80, 80, 78, 78.

Для построения сгруппированного ряда необходимо:

1. Определить величину интервала;
2. Определить середину, начало и конец групп вариант вариационного ряда.

• Величина интервала (i) определяется по числу предполагаемых групп (r), количество которых устанавливается в зависимости от числа наблюдений (n) по специальной таблице .

Число групп в зависимости от числа наблюдений

n (число наблюдений)	31 – 45	46 – 100	101 – 200	201 – 500
r (число групп)	6 – 7	8 – 10	11 – 12	12 – 17

Величина интервала (i) определяется по следующей формуле:

$$i = \frac{V_{max} - V_{min}}{r}$$

в нашем примере величина интервала равна $(82 - 58) / 8 = 3$

Если величина интервала представляет собой дробное число, полученный результат следует округлить до

Оптимальное число групп, на которое следует разбить конкретную совокупность, можно определить и по формуле Стерджеса:

$$i = \frac{V_{max} - V_{min}}{1 + 3,32 \lg n}$$

Где $\lg n$ – десятичный логарифм общего число единиц данной совокупности.

Для того, чтобы правильно сгруппировать варианты, необходимо определить середину 1 -ой группы вариант, величина которой должна быть ближайшей к максимальному значению изучаемого признака и должна делиться на размер интервала.

В нашем примере, размер максимальной варианты равен 82, но эта величина не делится на интервал, равный 3, поэтому серединой 1ой группы будет значение 81, т.к. эта величина близка к максимальному значению ряда (82) и делится на 3.

Чтобы найти середины для других групп необходимо от середины каждой предыдущей группы отнять величину интервала.

Для определения *начала группы* к ее середине прибавляют величину $(i - 1)/2$, вычитая же ее из середины, получают *конец группы*.

Распределение студентов-медиков по частоте пульса перед экзаменами

Начало группы	Середина группы	Конец группы	Варианты (V)	Частоты (p)
2-е действие	1-е действие	3-е действие	4-е действие	5-е действие
82	81	80	82 – 80	4
79	78	77	79 – 77	8
76	75	74	76 – 74	16
73	72	71	73 – 71	5
70	69	68	70 – 68	11
67	66	65	67 – 65	2
64	63	62	64 – 62	6
61	60	59	61 – 59	2
58	57	56	58 – 56	1

Таким образом, мы научились составлять, строить вариационные ряды, в том числе сгруппированные, без которых нельзя определить среднюю величину изучаемого количественного признака.

Различают несколько видов средних величин:

- средняя арифметическая,
- средняя геометрическая,
- средняя гармоническая,
- средняя квадратическая,
- средняя прогрессивная,
- мода,
- медиана и д.р.

Средняя арифметическая величина (M или X) является обобщающей величиной, которая определяет то типичное, что характерно для всей совокупности. Основными способами расчета M (X) являются: **среднеарифметический способ и способ моментов (условных отклонений)**.

Среднеарифметический способ применяется для вычисления средней арифметической простой и средней арифметической взвешенной.

В случае **простого вариационного ряда**, в котором каждая варианта встречается только один раз, определяется средняя арифметическая по формуле:

$$M = \frac{\sum V}{n}$$

где: M – средняя арифметическая величина;
 V – значение варьирующего признака (варианты);
 Σ – указывает действие – суммирование;
 n – общее число наблюдений.

Средняя арифметическая взвешенная определяется по формуле:

$$M = \frac{\sum Vp}{n}$$

где n – число наблюдений, равное сумме частот – Σp .

Способ моментов. Этот более простой способ вычисления средней арифметической взвешенной величины применяется при большом числе наблюдений и вариантах, выраженных большими числами. Он основан на том, что алгебраическая сумма отклонений отдельных вариантов вариационного ряда от средней арифметической равна нулю, т.е. $\Sigma(-d) = \Sigma(+d)$, где d – истинные отклонения варианты от истинной средней арифметической величины.

Средняя арифметическая по способу моментов определяется по формуле:

$$M = A + i \frac{\Sigma ap}{n},$$

где:

A – условно принятая средняя;

a – условное отклонение каждой варианты от условной средней ($V - A$);

i – величина интервала, т.е. разность между соседними вариантами.

Для расчета средней арифметической взвешенной по способу моментов :

1. Построить вариационный ряд, расположив варианты в возрастающем или убывающем порядке с соответствующими им частотами.
2. Выбрать условную среднюю (A). За условную среднюю можно взять любую варианту ряда, но чаще всего принимают наиболее часто встречающуюся варианту.
3. Определить условные отклонения. Условное отклонение (a) вычисляется как разность между каждой вариантой и условной средней ($V-A$).
4. Перемножить значение каждого условного отклонения на соответствующую частоту (ap), и найти
5. Подставить все значения в формулу:

$$M = A + i \frac{\sum ap}{n},$$

Средняя величина может быть
рассчитана не только на основе
абсолютных данных, но и **среди**
относительных показателей. При
одинаковых числах наблюдений ее находят
как **среднюю простую**, т.е. достаточно
суммировать размеры показателей и затем
поделить на их число. При разных числах
наблюдений **среднюю величину** среди
показателей следует определять всегда как
среднюю взвешенную.

Пример 1. В результате измерения длины тела (в см) при рождении у 47 девочек были получены следующие данные:
48, 51, 53, 49, 51, 53, 51, 48, 52, 51, 53, 49, 50, 53, 48, 52, 50, 52, 50,
52, 50, 51, 52, 53, 47, 52, 48, 48, 50, 52, 46, 46, 54, 55, 56, 48, 52, 52,
51, 53, 53, 48, 50, 54, 48, 50, 50.

Пример 2. Результаты измерения температуры тела у 22 новорожденных были следующими: 37,0; 36,6; 37,2; 36,9; 36,6;
37,0; 37,1; 36,8; 37,0; 36,9; 37,2; 37,1; 36,8; 36,7; 36,9; 36,6; 37,0;
36,9; 36,7; 36,8; 37,0; 36,6.

Используя методику расчета средней арифметической взвешенной по способу моментов, определим среднюю длину тела у девочек при рождении и среднюю температуру тела у новорожденных. Для этого необходимо:

- Построить вариационный ряд, расположив варианты в возрастающем или убывающем порядке с соответствующими им частотами. В нашем примере варианты расположены в убывающем порядке (табл. 6, графы 1, 2).

ПРИМЕР 1				ПРИМЕР 2			
Длина тела (см), V_1	Число девочек, p	a ($V-A$)	ap ($V-A$) p	Температура тела, ($^{\circ}$ C), V_2	Число девочек, p	a ($V-A$)	ap ($V-A$) p
56	1	4	4	36,6	4	-4	-16
55	1	3	3	36,7	2	-3	-6
54	2	2	4	36,8	3	-2	-6
53	7	1	7	36,9	4	-1	-4
A=52	9	0	0	A=37,0	5	0	0
51	6	-1	-6	37,1	2	+1	+2
50	8	-2	-16	37,2	2	+2	+4
49	2	-3	-6				
48	8	-4	-32				
47	1	-5	-5				
46	2	-6	-12				
	$\Sigma p=n$ =47		$\Sigma ap=$ -59		$\Sigma p=n$ =22		$\Sigma ap=$ -26
$M_1 = A + \frac{\Sigma ap}{n} = 52 + \left(\frac{-59}{47} \right) = 52 - 1,26 = 50,74 \text{ CM}$				$M_2 = A + i \frac{\Sigma ap}{n} = 37,0 + 0,1 \left(\frac{-26}{22} \right) = 37,0 - 0,1 = 36,9 \text{ } ^{\circ}\text{C}$			

• Выбрать условную среднюю (A). За условную среднюю можно взять любую варианту ряда, но чаще всего принимают наиболее часто встречающуюся варианту. В примере №1 наиболее часто встречается варианта 52, она встречается у 9 девочек, т.е. $A=52$. В примере №2 условная средняя равна 37°C .

• Определить условные отклонения (графа 3). Условное отклонение (a) вычисляется как разность между каждой вариантой и условной средней ($V-A$). Вычисленные значения условных отклонений занесем в графу 3 табл. 6 с учетом алгебраических знаков. Условным отклонениям (a) в графе 7 приданы порядковые номера.

• Перемножить значение каждого условного отклонения с соответствующей частотой (ap), результаты занести в графу 4 и 8 табл. 6 и найти их сумму (Σap).

• Подставить все значения в формулы:

$$M_1 = A + \frac{\Sigma ap}{n} = 52 + \frac{-59}{47} = 52 - 1,26 = 50,74 \text{ CM}$$

Средняя арифметическая взвешенная по способу моментов в случае сгруппированного (интервального) вариационного ряда

В сгруппированном ряду расчет средней арифметической начинается с определения середины интервала (центральной варианты). Центральная варианта в непрерывных вариационных рядах определяется как полусумма наименьших значений двух соседних групп.

Например:

Группы вариант

10, - 10,9

Центральная варианта

В качестве примера рассчитаем среднюю частоту пульса перед экзаменом у студентов-медиков (по способу моментов), используя данные полученные нами ранее.

Начало группы	Середина группы	Конец группы	Варианты (V)	Частоты (p)
2-е действие	1-е действие	3-е действие	4-е действие	5-е действие
82	81	80	82 – 80	4
79	78	77	79 – 77	8
76	75	74	76 – 74	16
73	72	71	73 – 71	5
70	69	68	70 – 68	11
67	66	65	67 – 65	2
64	63	62	64 – 62	6
61	60	59	61 – 59	2
58	57	56	58 – 56	1

Ниже сведены некоторые данные причем, для упрощения расчетов разность между соседними центральными вариантами принята за 1, вместо действительной разности, равной 3 (81 – 78)

Пульс, уд/мин (V)	Число студентов (p)	Центральная варианта ($V_{\text{центр.}}$)	Условное отклонение (a)	ap
1	2	3	4	5
82-80	4	81	2	8
79-77	8	78	1	8
76-74	16	75	0	0
73-71	5	72	-1	-5
70-68	11	69	-2	-22
67-65	2	66	-3	-6
64-62	6	63	-4	-24
61-59	2	60	-5	-10
58-56	1	57	-6	-6
	$\Sigma p = n = 55$			$\Sigma ap = -57$

$$M = A + i \frac{\sum ap}{n},$$

где:

A – условная средняя (наиболее часто встречающаяся варианта, в нашем примере $A=75$, такая частота пульса встречалась у 16 студентов);

i – величина интервала, т.е. разность между соседними центральными вариантами, в нашем примере $i=3$.

Остальные обозначения известны.

$$M = A + i \frac{\sum ap}{n} = 75 + 3 \frac{-57}{55} = 75 + (-3,1) = 71,9 \text{ уд/мин}$$

Вывод. Частота пульса у студентов-медиков перед экзаменом составляла в среднем 71,9 (≈ 72) удара в минуту.

Мода (Mo) – наиболее часто встречающаяся варианта в вариационном ряду. Для распределения, представленного в таблице, моде соответствует варианта, равная 10, она встречается чаще других – 6 раз.

Распределение больных по длительности пребывания

<i>V</i>	3	6	7	8	10	12	13	15	17
<i>p</i>	2	3	4	5	6	5	4	3	2

Медиана (Me) – непараметрический показатель, делящий вариационный ряд на две равные половины: в обе стороны от медианы располагается одинаковое число вариантов.

Например, для распределения, указанного в таблице выше, медиана равна 10, т.к. по обе стороны от этой величины располагается по 14 вариантов, т.е. число 10 занимает центральное положение в этом ряду и является его медианой.

Учитывая, что *число наблюдений* в этом примере *четное* ($n=34$), медиану можно определить таким образом:

$$Me = \frac{2 + 3 + 4 + 5 + 6 + 5 + 4 + 3 + 2}{2} = \frac{34}{2} = 17$$

По характеру количественного признака

ПРЕРЫВНЫЙ

варианты отличаются друг от друга на целое число и не имеют промежуточных дробных значений

НЕПРЕРЫВНЫЙ

значение вариант выражены дробными числами и носят непрерывный характер

непараметрические средние

МОДА (M_o)

наиболее часто встречающаяся варианта

МЕДИАНА (M_e)

показатель, делящий вариационный ряд на две равные половины

♦ при четном числе наблюдений

$$M_e = \frac{n}{2}$$

♦ при нечетном числе наблюдений

$$M_e = \frac{n+1}{2}$$

♦ с учетом центральных вариант

По частоте встречаемости вариант

ПРОСТОЙ

каждая варианта (V) встречается только один раз

ВЗВЕШЕННЫЙ

одинаковые значения вариант встречаются несколько раз, т.е. имеют определенную частоту повторения (P)

СТРУПШИРОВАННЫЙ

одинаковые значения вариант встречаются несколько раз, т.е. имеют определенную частоту повторения (P)

параметрические средние

СРЕДНЯЯ

АРИФМЕТИЧЕСКАЯ (M)

является обобщающей величиной, которая определяет то типичное, что характерно для всей совокупности

СРЕДНЯЯ ПРОГРЕССИВНАЯ (M)

определяется среди оптимальных показателей, т.е. среди показателей более благоприятных по своим размерам в сравнении

среднеарифметический способ

способ моментов

♦ простая средняя $M = \frac{\sum V}{n}$

♦ взвешенная средняя $M = \frac{\sum VP}{n}$

♦ при взвешенном ряде $M = A + \frac{\sum ap}{n}$

♦ при сгруппированном ряде $M = A + i \frac{\sum ap}{n}$

Таблица 9

Содержание и применение средних величин

Средняя арифметическая величина

является обобщающей величиной, результивной суммой всех влияний, в формировании ее принимают участие все без исключения варианты, в том числе и крайние, часто нетипичные для данного явления или совокупности;

– характеризует всю массу наблюдений;

– занимает срединное положение в вариационном ряду. В строго симметричном ряду

$$M = M_o = M_e$$

Мода и Медиана

– не зависят от величины всех индивидуальных значений варьирующего признака (значений крайних вариант и степени рассеяния ряда);

– характеризуют основную массу наблюдений;

– применяются в случае незамкнутой совокупности, т.е. когда не имеют точной количественной характеристики наименьшая или наибольшая варианты (до, выше).

В этом случае нельзя рассчитать параметрическую среднюю.

– применение медианы целесообразно, когда ничего неизвестно о характере распределения результатов эксперимента, т.е. нет достаточных оснований для выбора

При изучении варьирующего признака, особенно в биологии и медицине, где изучаются живые организмы и их жизнедеятельность в норме и патологии, нельзя ограничиваться вычислением только средних величин, какими универсальными они бы ни были.

Средняя величина, рассчитанная математическим путем, – это величина, вокруг которой расположены на разном удалении варианты, вошедшие в вариационный ряд, из которого она была рассчитана.

Чем ближе друг к другу по значению отдельные варианты, тем меньше колеблемость (рассеянность) вариационного ряда, тем типичнее для характеристики изучаемого признака его средняя величина.

Различают показатели колеблемости, характеризующие:

- границы изучаемой совокупности (lim, Am);
- внутреннюю ее структуру (δ, δ^2, C_v).

Лимиты (пределы) – минимальная и максимальная варианты изучаемой совокупности, определяются крайними значениями вариант в вариационном ряду. Показывая фактические границы варьирования признака, лимиты имеют определенное значение в метеорологии, где показывают минимальную и максимальную температуру, а также в микробиологии для характеристики размеров микроорганизмов. Записываются лимиты следующим образом:

$$Lim = V_{max} \div V_{min}$$

Амплитуда (размах вариации) – разность лимитов (крайних вариантов) ($A_m = V_{\max} - V_{\min}$). С амплитуды можно оценить колеблемость одного вариационного ряда с амплитудой другого вариационного ряда.

если A_m первого вариационного ряда равна 5, а второго – 11, можно сделать вывод о том, что колеблемость второго вариационного ряда вдвое больше первого, при одинаковом значении средних величин, средняя рассчитанная из второго вариационного ряда, менее типична из-за резкой колеблемости.

Наиболее точной мерой варьирования, колеблемости вариационного ряда (изучаемого признака) являются дисперсия и среднее квадратическое отклонение (δ).

СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ – показатель рассеивания значений случайной величины относительно её математического ожидания (среднего арифметического совокупности выборок), это именованная величина, поэтому она должна иметь размерность общую для вариант и средней арифметической величины.

Способы расчета среднего квадратического отклонения:

- среднеарифметический
- способ моментов
- и по амплитуде вариационного ряда

1. СРЕДНЕАРИФМЕТИЧЕСКИЙ СПОСОБ РАСЧЕТА

Когда число наблюдений небольшое ($n \leq 30$), а все частоты в вариационном ряду $p=1$, применяется формула:

$$\delta = \pm \sqrt{\frac{\sum d^2}{n-1}},$$

где d – истинные отклонения вариант от истинной средней ($V - M$).

Последовательность расчета δ :

1. Построить вариационный ряд .
2. Определить среднеарифметическую величину (M) :

$$M = \frac{\Sigma V_p}{n}$$

3. Найти истинные отклонения d ($d=V - M$). Например, $d_1 = 2 - 7 = -5$ и т.д. Возвести каждое отклонение в квадрат (d^2).
5. Найти произведение ($d^2 P$) по всем строкам ряда .
6. Определить сумму $\Sigma d^2 P$.
7. Рассчитать δ по формуле:

$$\delta = \pm \sqrt{\frac{\Sigma p d^2}{n - 1}}$$

Пример

Распределение больных с острыми респираторными заболеваниями по длительности нетрудоспособности (в днях)

Алгоритм расчета среднего квадратического отклонения					
Длительность нетрудоспособности (в днях), V	Число больных p	Vp	d	d^2	d^2p
2	1	2	-5	25	25
3	2	6	-4	16	32
4	2	8	-3	9	18
5	6	30	-2	4	24
6	8	48	-1	1	8
7	6	42	0	0	0
8	3	24	+1	1	3
9	3	27	+2	4	12
10	1	10	+3	9	9
11	1	11	+4	16	16
12	1	12	+5	25	25
13	1	13	+6	36	36
	$\Sigma p = n = 35$	$\Sigma Vp = 233$			$\Sigma d^2p = 208$

Пример расчета δ :

1. Построить вариационный ряд (графы 1, 2).
2. Найти произведение вариантов и их частоты встречаемости (графа 3).
Определить среднеарифметическую величину (M) :

$$M = \frac{\sum vp}{n} = \frac{233}{35} = 6,7 \text{ дней}$$

3. Найти истинные отклонения d ($d=V - M$). Например, $d_1=2-7= -5$ и т.д., данные записать в графу 4.
4. Возвести каждое отклонение в квадрат(d^2), графа 5.
5. Найти произведение (d^2P) по всем строкам ряда (графа 6).
6. Определить сумму $\sum d^2P$, графа 6.
7. Рассчитать δ по формуле:

$$\delta = \pm \sqrt{\frac{\sum ad^2}{n}} = \sqrt{\frac{208}{35}} = \pm 2,4 \text{ дня}$$

2. СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ ПО СПОСОБУ МОМЕНТОВ

$$\delta = \sqrt{\frac{\sum a^2 p}{n} - \left(\frac{\sum ap}{n}\right)^2}$$

где: a – условное отклонение вариант от условной средней ($a = V - A$).

При числе наблюдений, равном 30 и менее, в формуле n заменяют на $(n - 1)$ и тогда δ определяется по формуле

$$\delta = \sqrt{\frac{\sum a^2 p}{n - 1} - \left(\frac{\sum ap}{n - 1}\right)^2}$$

Последовательность расчета δ по способу моментов:

1. Найти условную среднюю A .
2. Определить условные отклонения (a) каждой варианты от условной средней ($a = V - A$).
3. Получить произведения (ap), а затем их просуммировать (Σap).
4. Рассчитать истинную среднюю арифметическую по формуле:

$$M = A + i \frac{\Sigma ap}{n} .$$

5. Получить произведения a^2p по всем строкам вариационного ряда и просуммировать их (Σa^2p).
6. Рассчитать δ по способу моментов по формуле:

$$\delta = \sqrt{\frac{\Sigma a^2 p}{n} - \left(\frac{\Sigma ap}{n}\right)^2}$$

Пример

Распределение больных с острыми респираторными заболеваниями по длительности нетрудоспособности

Длительность нетрудоспособности (в днях), V	Число больных, p	a	ap	a^2p
1	2	3	4	5
2	1	-4	-4	16
3	2	-3	-6	18
4	2	-2	-4	8
5	6	-1	-6	6
A=6	8	0	0	0
7	6	+1	+6	6
8	3	+2	+6	12
9	3	+3	+9	18
10	1	+4	+4	16
11	1	+5	+5	25
12	1	+6	+6	36
13	1	+7	+7	49
	$\Sigma p = n = 35$		$\Sigma ap = 23$	$\Sigma a^2p = 210$

Пример расчета δ по способу моментов:

1. Найти условную среднюю $A(A=6)$.
2. Определить условные отклонения (a) каждой варианты(графа 3) от условной средней ($a = V - A$).
3. Получить произведения (ap), а затем их просуммировать(графа 4). В нашем примере $\Sigma ap=23$.
4. Рассчитать истинную среднюю арифметическую по формуле:

$$M = A + i \frac{\Sigma ap}{n} .$$

В нашем примере $M=6,7$ дней.

5. Получить произведения a^2p по всем строкам вариационного ряда и просуммировать их(графа 5). В нашем примере $\Sigma a^2p=210$.
6. Рассчитать δ по способу моментов по формуле:

$$\delta = \sqrt{\frac{\Sigma a^2 p}{n} - \left(\frac{\Sigma ap}{n}\right)^2} = 2,4 \text{ дня.}$$

3. МЕТОД ОПРЕДЕЛЕНИЯ СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ – ПО АМПЛИТУДЕ РЯДА

Применяется, если отсутствуют необходимые данные для вычисления среднего квадратического отклонения обычным путем или нет необходимости в получении высокой точности показателя колеблемости вариационного ряда:

$$\delta = \pm \frac{V_{\max} - V_{\min}}{K},$$

где: K – коэффициент, определяемый по таблице С.И. Ермолаева в зависимости от числа наблюдений

n	Значения коэффициента K									
0	–	–	1,13	1,69	2,06	2,33	2,53	2,70	2,85	2,97
10	3,08	3,17	3,26	3,34	3,41	3,47	3,53	3,59	3,64	3,69
20	3,73	3,78	3,82	3,86	3,90	3,93	3,96	4,00	4,03	4,06
30	4,09	4,11	4,14	4,16	4,19	4,21	4,24	4,26	4,28	4,30
40	4,32	4,34	4,36	4,38	4,40	4,42	4,43	4,45	4,47	4,48
50	4,50	4,51	4,53	4,54	4,56	4,57	4,59	4,60	4,61	4,63
60	4,64	4,65	4,66	4,68	4,69	4,70	4,71	4,72	4,73	4,74
70	4,75	4,77	4,78	4,79	4,80	4,81	4,82	4,83	4,83	4,84
80	4,85	4,86	4,87	4,88	4,89	4,90	4,91	4,91	4,92	4,93
90	4,94	4,95	4,96	4,96	4,97	4,98	4,99	4,99	5,00	5,01

4. СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ ДЛЯ ОТНОСИТЕЛЬНОЙ ВЕЛИЧИНЫ (P)

$$\delta = \sqrt{Pq} ,$$

где: P – величина относительного показателя, выраженного в % или ‰ и т.д.;

q – величина альтернативы (обратная величина P), т. е. $q=100 - P$ или $q=1000 - P$.

Одним из обязательных этапов методики статистической обработки вариационных рядов является **графическое изображение вариационного ряда**, которое позволяет определить, какому закону распределения подчиняется данное явление. В качестве меры колеблемости показатель среднеквадратического отклонения. Чаще всего встречается **нормальное распределение**, подчиняющееся закону Гаусса-Лапласа. **Для нормального распределения** характерна симметричность, т.е. крайние варианты (наибольшие и наименьшие) встречаются редко. Чем ближе значения варьирующего признака к величине средней арифметической, тем чаще они встречаются.

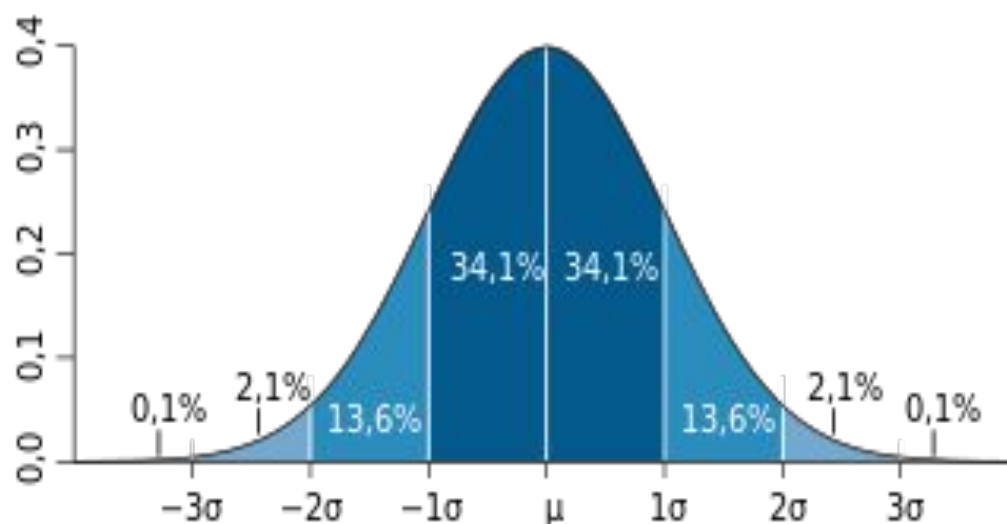


График плотности вероятности нормального распределения и процент попадания случайной величины на отрезки, равные среднеквадратическому отклонению.

Правило трёх сигм 3σ — практически все значения нормально распределённой случайной величины лежат в интервале $(\bar{x} - 3\sigma; \bar{x} + 3\sigma)$

Более строго — приблизительно с 0,9973 вероятностью значение нормально распределённой случайной величины лежит в указанном интервале.

Таким образом, при нормальном распределении при различных значениях средней и среднеквадратического отклонения, всегда 68,3% наблюдений находятся в пределах $\pm 1\delta$; 95,5% наблюдений находятся в пределах $\pm 2\delta$; 99,7% — в пределах $\pm 3\delta$. И только 0,3% (3 случая на 1000) наблюдений имеют значения, отличные от среднего больше чем на 3δ .

При помощи δ определяют типичность средней величины и меру ее точности. Если 95% всех вариантов находятся в пределах $M \pm 2\delta$, то средняя является характерной для данного ряда, и не требуется увеличивать число наблюдений в выборочной совокупности.

Для нормального распределения характерна **симметричность**. Наиболее точным показателем, характеризующим симметричность распределения, является **коэффициент асимметрии**, который рассчитывается по формуле:

$$As = \frac{\sum d^2 p}{n\delta^3}$$

Коэффициент асимметрии оценивается по специальной таблице.

Критические значения коэффициента асимметрии As ($P_{+}=0,95$)

n	As	n	As	n	As
25	0,711	70	0,459	200	0,280
30	0,661	80	0,432	250	0,251
35	0,621	90	0,409	300	0,230
40	0,587	100	0,389	350	0,213
45	0,558	125	0,350	400	0,200
50	0,533	150	0,321	450	0,188
60	0,492	175	0,298	500	0,179

Если рассчитанный $As \leq As_{0,95}$ (табличного), отвергается предположение о наличии асимметрии, т.е. распределение можно считать нормальным. При $As \geq As_{0,95}$ распределение асимметрично. Знак As указывает направление асимметрии («-» – левосторонняя, «+» – правосторонняя).

В качестве относительной меры вариабельности применяется **коэффициент вариации** если:

1. Сравниваются **не только** однородные совокупности (одноименные) или признаки.
2. Средние уровни сравниваемых признаков **незначительно** отличаются друг от друга.

$$C_v = \frac{\delta}{M} \times 100\%$$

Коэффициент вариации рассчитывается по форму

Коэффициент вариации является критерием надежности средней арифметической. Если $C_v \geq 40\%$, то средняя арифметическая неустойчива и ненадежна.

Оценка степени колеблемости изучаемых признаков по коэффициенту вариации:

C_v (в %)	Степень колеблемости (рассеяние вариантов около средней арифметической величины)
менее 10	малая
от 10 до 20	средняя
более 20	сильная

