Практическое занятие 6: Формула для вычисления дисперсии. Среднее квадратическое отклонение. Коэффициент вариации

Цель занятия:

- приобретение навыков вычисления дисперсии и среднего квадратического отклонения;
- вычисление коэффициента вариации.

Вычисление дисперсии

Генеральная дисперсия - среднее арифметическое квадратов отклонений всех вариант генеральной совокупности от её средней:

$$D_Г = \frac{\sum_{i=1}^{N}(x_i - \bar{x}_Г)^2}{N},$$

где N – объём генеральной совокупности.

Генеральная дисперсия для сформированного вариационного ряда формула принимает вид:

$$D_Г = \frac{\sum_{i=1}^{K}(x_i - \bar{x}_Г)^2 \cdot n_i}{N},$$

Выборочная дисперсия - среднее арифметическое квадратов отклонений всех вариант выборки от её средней:

$$D_в = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_в)^2}{n}$$ – для не сгруппированных данных:

$$D_в = \frac{\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 \cdot n_i}{n}$$ – для сформированного вариационного ряда,

где x_i – кратные (одинаковые по значению) *варианты* в дискретном случае либо середины *частичных интервалов* – в интервальном;

n_i – соответствующие частоты.

Расчёт дисперсии по определению прост и реально используется на практике, но существует **ещё более простой и удобный способ вычисления – по формуле** – дисперсия равна разности *средней арифметической* квадратов всех *вариант* статистической совокупности и квадрата *средней* самих этих вариант.

$$D = \overline{x^2} - \overline{x}^2$$

$$D(X) = M(X^2) — [M(X)]^2.$$

<u>*Доказательство*</u>. Математическое ожидание $M(X)$ — это постоянная величина, поэтому, $2M(X)$ и $M^2(X)$ будут также постоянными величинами. Зная это и применяя свойства математического ожидания, которые говорят, что постоянный множитель можно вынести за знак математического ожидания, а математическое ожидание суммы равно сумме математических ожиданий слагаемых, можно упростить формулу, определяющую дисперсию:

$$D(X)= M[X-M(X)]^2 = M[X^2 – 2XM(X) + M^2(X)] =$$
$$= M(X^2) – 2M(X)M(X) + M^2(X) = M(X^2) – M^2(X)$$

Для несгруппированных *вариант* $x_1, x_2, \ldots, x_n$ выборочной совокупности формула детализируется следующим образом:

$$D_e = \frac{\sum_{i=1}^{n} {x_i}^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2$$

А для готового вариационного ряда

$$D_e = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - \left(\frac{\sum_{i=1}^{k} x_i n_i}{n} \right)^2,$$

где $\boldsymbol{x_i}$ – кратные (одинаковые) варианты дискретного ряда, либо середины интервалов интервального ряда;

$\boldsymbol{n_i}$ – соответствующие частоты.

Для генеральной дисперсии $\boldsymbol{D_Г}$ формулы те же, только с буквами ***N, K*** вместо ***n, k***.

Коэффициент вариации

Рассмотренные выше показатели (размах вариации, среднее линейное отклонение, дисперсия, стандартное отклонение) входят в группу ***абсолютных*** **показателей вариации**, которые обладают рядом неудобств. Так, если в предыдущей задаче не уменьшать варианты в 1000 раз, то дисперсия получится в миллион раз больше! Да-да, не $\boldsymbol{D_в}$=0,0576, а 57600. И возникает естественное желание привести результаты к некому единому стандарту.

Для этого существуют показатели *относительные*, и самым известным из них является **коэффициент вариации** – отношение *стандартного отклонения* к *средней*, выраженное в процентах:

$$V = \frac{\sigma}{\overline{x}} \bullet 100\% .$$

И вот теперь совершенно без разницы, в д.е. мы считали:

$V = \frac{\sigma_e}{\overline{x}_e} \bullet 100\%$ = (240/780)*100% ≈ 30,77 %

или в тысячах д.е.:

V = (0,24/0,78)*100 % ≈ 30,77 %

Примечание*: на практике часто считают именно через* $\boldsymbol{\sigma_в}$*, но для оценки коэффициента вариации всей генеральной совокупности, конечно же, корректнее использовать исправленное стандартное отклонение.*

В статистике существует следующий эмпирический ориентир:

- если показатель вариации составляет примерно 30% и меньше, то статистическая совокупность считается ***однородной***. Это означает, что большинство *вариант* находится недалеко от *средней*, и найденное значение $\bar{x}$ хорошо характеризует центральную тенденцию совокупности;

- если показатель вариации составляет существенно больше 30%, то выборка ***неоднородна***, то есть, значительное количество *вариант* находятся далеко от $\bar{x}$, и выборочная средняя плохо характеризует типичную варианту.

Другое **преимущество относительных показателей** – это ***возможность сравнивать разнородные статистические совокупности***.

Например, множество слонов и множество хомячков. Совершенно понятно, что *дисперсия* веса слонов по отношению к дисперсии веса хомяков будет просто огромной, и их сопоставление не имеет смысла. Но вот анализ *коэффициентов вариации* веса вполне осмыслен, и может статься, что у слонов он составляет 10%, а у хомячков 40% *(пример, конечно, условный)*. Это говорит о сбалансированном питании и размеренной жизни слонов. А вот хомяки там, то носятся с голодухи по полям, то отъедаются и спят в норах, и поэтому среди них есть много худощавых и много упитанных особей.

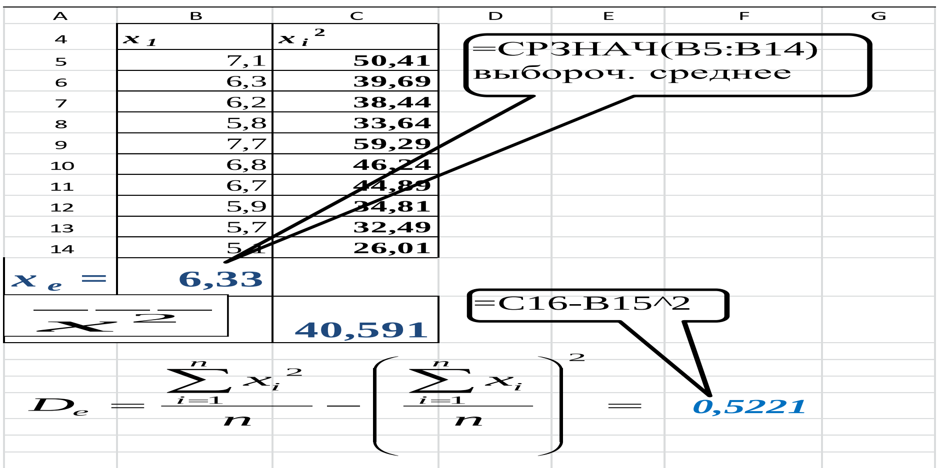
Задание 1

В результате 10 независимых измерений получены опытные данные, которые представлены в таблице:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
7,1	6,3	6,2	5,8	7,7	6,8	6,7	5,9	5,7	5,1

Требуется вычислить дисперсию.

Методика выполнения



Варианты задания 1

№ вар.	Результаты измерений, x_i									
1	3,65	2,09	3,68	2,26	5,94	3,90	4,22	2,85	3,09	5,01
2	2,48	4,06	3,92	3,00	3,43	5,66	2,06	4,48	4,28	3,37
3	3,65	4,93	4,57	4,75	5,47	5,89	4,15	3,74	5,82	4,02
4	2,92	2,99	4,06	5,22	2,25	4,42	5,65	4,12	5,48	3,09
5	5,51	4,69	5,08	5,28	5,61	5,49	4,39	4,86	3,64	2,53
6	4,77	3,09	3,35	5,21	5,92	2,18	2,32	5,20	5,57	5,95
7	2,66	5,11	3,03	5,34	5,12	2,63	3,12	3,61	4,77	2,05
8	2,61	2,22	2,14	3,83	5,34	4,05	5,69	3,51	3,15	3,62
9	2,58	5,93	3,51	5,52	5,90	5,47	3,27	2,77	3,07	4,32
10	4,37	2,85	5,63	4,98	2,28	4,45	2,38	5,44	2,59	4,45
11	4,17	5,22	2,92	3,81	5,04	5,25	3,85	4,73	4,77	5,59
12	5,85	3,09	2,63	3,34	2,22	2,73	5,91	4,25	4,47	6,00
13	3,12	4,90	4,20	3,43	4,59	3,34	4,92	2,98	4,58	3,88
14	4,73	4,19	4,64	2,39	4,72	4,73	5,39	4,48	4,05	4,94
15	3,93	3,43	3,70	5,98	3,29	3,34	2,16	4,22	3,33	4,88
16	4,46	2,70	4,78	4,47	2,02	3,83	5,37	2,65	3,49	2,84
17	4,33	2,32	5,89	3,13	3,28	5,99	3,04	2,22	2,14	2,24
18	2,88	2,63	5,49	3,01	3,24	4,86	3,87	3,06	3,68	5,47
19	3,02	5,30	2,21	3,11	5,64	5,08	5,89	3,21	5,80	2,37
20	5,05	2,08	2,87	5,37	3,75	3,59	4,98	2,68	3,96	2,88
21	4,33	5,65	3,16	5,35	2,34	2,72	3,23	4,62	5,30	3,50
22	5,97	4,82	2,35	2,26	3,28	3,37	4,60	4,66	4,57	2,76
23	2,74	3,24	2,27	5,32	4,51	4,92	2,10	3,94	4,65	3,35
24	4,66	3,63	2,91	5,18	2,10	5,04	4,23	5,03	2,83	2,15
25	2,57	4,76	4,63	4,81	4,16	3,82	2,02	4,09	2,14	5,80

Задание 2

С целью изучения вкладов в Сбербанке города проведено выборочное исследование, в результате которого получены следующие данные.

Вычислить выборочную дисперсию и среднее квадратическое отклонение, оценить соответствующие показатели генеральной совокупности.

Размер вклада	Число вкладчиков
до 400	32
400-600	56
600-800	120
800-1000	104
свыше 1000	88
Σ=	400

Методика выполнения

Проблема: подсчитан объем выборки $\boldsymbol{n}$=400, но не «закрыты» крайние интервалы.

Поскольку длины внутренних интервалов составляют $\boldsymbol{h}$=200 д.е., то логично рассмотреть такую же длину и по краям, то есть, интервалы от 200 до 400 и от 1000 до 1200 денежных единиц.

Интервалы	Середины, x_i	Частоты, n_i
200-400	300	32
400-600	500	56
600-800	700	120
800-1000	900	104
1000-1200	1100	88
	Σ=	400

Вопрос: а как быть, если даны интервалы разной длины? В этом случае принимаем за «эталон» среднюю длину известных интервалов.

Для расчёта числовых характеристик перейдём к **дискретному вариационному ряду**, выбрав в качестве *вариант* x_i середины интервалов.

Кроме того, варианты целесообразно уменьшить в 1000 раз, поскольку в ходе дальнейших вычислений будут получаться достаточно большие числа.

A	B	C	D	E	F
9	Интервалы	Середины, x_i	Частоты, n_i	$x_i n_i$	$x_i^2 n_i$
10	200-400	0,3	32	9,6	2,88
11	400-600	0,5	56	28	14
12	600-800	0,7	120	84	58,8
13	800-1000	0,9	104	93,6	84,24
14	1000-1200	1,1	88	96,8	106,48
15		Σ=	400	312	266,4
16			$\overline{x_e} =$	**0,78**	
17				**(тыс.д.е.)**	
18				$\overline{x^2} =$	**0,666**
19					**(тыс.д.е.)^2**
20	$D_e = \dfrac{\sum_{i=1}^{k} x_i^2 n_i}{n} - (\overline{x_e})^2 = 0{,}666 - (0{,}78)^2 =$			***0,0576***	***(тыс.д.е.)^2***
21					
22	$\delta_e = \sqrt{D_e} =$		***0,24***	**(тыс.д.е.)**	

Теперь следует корректно оценить генеральную дисперсию $D_Г$ и генеральное среднее квадратическое отклонение $\sigma_Г$.

Исправленная выборочная дисперсия

$$D = \frac{n}{n-1} D_e = \frac{400}{399} 0{,}0576 \approx 0{,}057744$$

Исправленное среднее квадратическое отклонение $\sigma = D^{1/2} \approx 0{,}2403$.

Варианты задания 2

№ вар.	*Размер вклада, ден.ед.*				
	до 400	*400-600*	*600-800*	*800-1000*	*свыше 1000*
	Число вкладчиков				
1	51	33	68	68	84
2	92	53	54	57	90
3	68	68	90	58	108
4	94	51	89	33	76
5	31	105	76	111	45
6	73	68	101	51	118
7	114	71	71	77	81
8	31	41	55	63	52
9	86	106	85	104	97
10	62	90	41	116	93
11	102	65	36	78	52
12	110	118	30	49	72
13	87	82	71	81	105
14	96	117	68	107	42
15	57	34	48	84	72
16	101	105	68	90	92
17	106	59	46	38	74
18	30	58	73	96	75
19	73	90	32	92	106
20	97	103	107	107	59
21	64	58	46	97	118
22	31	70	90	102	64
23	41	70	42	64	59
24	67	92	84	118	91
25	73	78	56	41	59

Задание 3

Стандартное отклонение выборочной совокупности равно $\sigma_в$ =5, а средний квадрат её вариант $\overline{x^2}$ = 250. Найти выборочную среднюю.

Методика выполнения

Используем формулу $D_в = \overline{x^2} - \overline{x_в}^2$.

По условию $D_в = \sigma_в^2 = 5^2 = 25$. $\overline{x^2}$ =250.

Тогда 25=250-$\overline{x_в}^2$. ⇒ $x_в = \sqrt{225} = 15$.

Варианты задания 3

№ вар.	*СКО*	$\overline{x^2}$	*№ вар.*	*СКО*	$\overline{x^2}$	*№ вар.*	*СКО*	$\overline{x^2}$
1	5	205	9	1	232	17	1	315
2	8	294	10	5	261	18	7	345
3	6	264	11	8	230	19	6	316
4	8	323	12	1	253	20	5	223
5	8	249	13	3	263	21	8	293
6	8	320	14	1	324	22	5	310
7	4	257	15	5	254	23	8	265
8	1	266	16	8	238	24	4	242

Задание 4

Определите среднее квадратическое отклонение, если известно, что средняя равна 260, а коэффициент вариации составляет 30%.

Методика выполнения

$V = \frac{\sigma}{\overline{x}} \bullet 100\%$ По условию $\overline{x}_e = 260$, V=30 %.

30 % = (σ/260)*100 % $\Rightarrow$ $\boldsymbol{\sigma}$=(30%*260)/100% = 78

Варианты задания 4

№ вар.	***V***	$\overline{x_e}$	***№ вар.***	***V***	$\overline{x_e}$	***№ вар.***	***V***	$\overline{x_e}$
1	37	240	9	47	235	17	20	223
2	64	308	10	36	200	18	57	227
3	38	207	11	68	320	19	53	254
4	52	223	12	24	203	20	52	284
5	35	311	13	31	239	21	42	323
6	50	265	14	61	332	22	68	296
7	67	327	15	61	326	23	20	228
8	21	297	16	69	242	24	58	242

Задание 5

Производство стальных труб на предприятии (тонн) в 1-м полугодии составило:

январь	февраль	март	апрель	май	июнь
263	284	310	296	288	251

Определить:

– среднемесячный объем производства;

– среднее квадратическое отклонение;

– коэффициент вариации.

Методика выполнения

	x_i	x_i^2
	263	***69169***
	284	***80656***
	310	***96100***
	296	***87616***
	288	***82944***
	251	***63001***
Σ=	***1692***	***479486***
$\overline{x}$ =	***282***	

$$D = \overline{x^2} - \overline{x}^2$$

$$D_e = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 = \mathbf{79914{,}333\text{-}282^2 = 390{,}333}$$

$$\frac{\sum_{i=1}^{n} x_i^2}{n} = \mathbf{79914{,}333}$$

Среднее квадратическое отклонение: ***19,757 тонн***

$$V = \frac{\delta}{\overline{x}} \bullet 100\% = \mathbf{19{,}757/282*100\%} \approx \mathbf{7\%}$$

Краткие выводы: за первое полугодие среднемесячный объём производства труб составил 282 тонны. Низкие показатели вариации говорят о стабильной ситуации на производстве

Варианты задания 5

№ вар.	*январь*	*февраль*	*март*	*апрель*	*май*	*июнь*
1	289	276	265	284	274	286
2	291	275	266	280	261	277
3	287	296	297	262	267	293
4	295	288	298	282	250	272
5	295	267	282	267	285	277
6	254	257	252	295	263	272
7	265	257	250	300	270	272
8	265	278	276	284	293	295
9	279	287	259	299	300	282
10	266	273	254	292	279	263
11	295	277	279	290	295	279
12	259	281	260	264	262	293
13	270	274	264	292	274	286
14	278	285	257	251	263	290
15	298	286	252	296	295	282
16	261	280	274	273	255	285
17	261	300	256	298	274	278
18	288	261	263	253	259	272
19	277	273	269	283	277	255
20	266	254	266	300	300	250
21	268	275	253	253	262	269
22	278	265	291	270	250	295
23	271	283	267	286	272	285
24	289	287	258	293	286	269
25	268	274	261	278	270	252