Практическое занятие 5: Размах вариации. Среднее линейное отклонение. Генеральная и выборочная дисперсия

Цель занятия:

- приобретение навыков вычисления размаха вариации;
- вычисление среднего линейного отклонения;
- вычисление генеральной и выборочной дисперсии;
- вычисление среднего линейного отклонения;
- вычисление исправленной выборочной дисперсии.

Показатели вариации показывают, как варьируются статистические данные, а именно – насколько далеко «разбросаны» варианты относительно средних значений, да и просто друг от друга.

Систематизируем информацию о том, какие статистические данные могут оказаться в нашем распоряжении:

- первичные (не обработанные), грубо говоря – это неупорядоченный список чисел, либо вторичными – это уже сформированный дискретный или интервальный вариационный ряд;
- рассматриваемая статистическая совокупность может быть генеральной либо выборочной (чаще всего).

Размах вариации

Это разность между самым большим и самым малым значением статической совокупности: $\boldsymbol{R = x_{max} - x_{min}}$,

при этом не имеет значения, генеральная ли это совокупность или выборочная, сгруппированы данные или нет.

Очевидно, что все варианты x_i исследуемой совокупности заключены в отрезке [x_{min}; x_{max}], а размах R – есть не что иное, как его длина.

Пример
Дана статистическая совокупность 15, 17, 13, 10, 21, 17, 23, 9, 14, 19.
Найти размах вариации.
Решить задачу можно несколькими способами:

- вручную;
- с использованием программного обеспечения, при этом числа можно просто отсортировать (по возрастанию либо убыванию) или использовать специальные функции.

Решение.

A	B	C	D	E	F	G	H	I	J	K	L
2	x_i	15	17	13	10	21	17	23	9	14	19
3	min=	9	=МИН(C2:L2)					=ABS(C4-C3)			
4	max=	23	=МАКС(C2:L2)				$R = xmax - xmin =$				14

Среднее линейное отклонение

$$\bar{l} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$ – есть среднее арифметическое абсолютных отклонений всех значений статистической совокупности от средней. Это формула для не сгруппированных статистических данных.

Если имеется сформированный дискретный либо интервальный вариационный ряд, то формула будет такой:

$$\bar{l} = \frac{\sum_{i=1}^{k} |x_i - \vec{x}| * n_i}{n},$$

где x_i – *варианты* (для дискретного ряда) либо середины частичных интервалов (для интервального ряда), а n_i – соответствующие частоты.

Маленькая буква n обычно используется для *выборочной* совокупности, а большая – для *генеральной*: N – объём ген. совокупности, Ni – частоты.

Генеральная дисперсия

Это среднее арифметическое квадратов отклонений всех вариант генеральной совокупности от её средней:

$$D_\Gamma = \frac{\sum_{i=1}^{N} (x_i - \bar{x}_\Gamma)^2}{N},$$

где N – объём генеральной совокупности.

Для сформированного вариационного ряда формула принимает вид:

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

где x_i – либо варианты дискретного ряда, либо середины частичных интервалов интервального ряда, а n_i – соответствующие частоты.

Выборочная дисперсия

Это среднее арифметическое квадратов отклонений всех вариант выборки от её средней:

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

$$D_в = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_в)^2 \cdot n_i}{n}$$ – 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.

A	B	C	D	E	F	G
4	x_1	$(x_i - x_в)^2$				
5	7,1	**0,5929**				
6	6,3	**0,0009**				
7	6,2	**0,0169**				
8	5,8	**0,2809**				
9	7,7	**1,8769**				
10	6,8	**0,2209**				
11	6,7	**0,1369**				
12	5,9	**0,1849**				
13	5,7	**0,3969**				
14	5,1	**1,5129**				
$x_в =$	**6,33**					
$D_в =$		**0,5221**				

=ABS(B5-B15)

=СРЗНАЧ(B5:B14)

=СУММ(C5:C14)/10 Выборочная дисперсия

При проведении серии каких-то испытаний (измерений) выборочные значения $\bar{x}_в$ *без какой-либо закономерности* (в общем случае) будут варьироваться вокруг истинного значения показателя $\bar{x}_Г$ (роль генеральной средней может играть некий теоретический эталон). Это свойство (отсутствие закономерности) называется ***несмещённостью*** оценки генеральной средней, но справедливо оно, не для всех показателей.

Теперь об отклонениях. Всё просто: у кого эти показатели ниже, тот качественнее проводит опыты *(плавнее выполняет действия, точнее снимает показания с приборов, засекает время и т.п.)*. В идеале эти отклонения равны нулю, но это только в идеале – сам эмпиризм ситуации порождает генеральное линейное отклонение $\bar{l}_{Г}$ и генеральную дисперсию $\boldsymbol{D_Г}$, которые обусловлены человеческим фактором, погрешностью приборов и так далее.

В случае с полученными линейными отклонениями $\bar{l}$ – всё то же самое, они будут безо всякой закономерности варьироваться вокруг генерального значения $\bar{l}_{Г}$.

Но вот с дисперсией всё не так. Полученные значения выборочной дисперсии $\boldsymbol{D_в}$ будут давать *систематически* заниженную оценку генеральной дисперсии $\boldsymbol{D_Г}$.

Дисперсия, как и доля или средняя арифметическая, также меняет свое значение от выборки к выборке, но здесь есть интересная особенность. Дисперсия ведь рассчитывается от средней величины, а она в свою очередь, тоже рассчитывается по выборке, то есть является ошибочной. Как же это обстоятельство влияет на саму дисперсию?

Если бы мы знали истинную среднюю величину (по генеральной совокупности), то ошибка дисперсии была бы связана только с нерепрезентативностью, то есть с тем, что данные в выборке оказались бы ближе или дальше от средней, чем в целом по генеральной совокупности. При этом при многократном повторении данные стремились бы к своему реальному расположению относительно средней.

Выборочный показатель, который при многократном повторении выборки стремится к своему теоретическому значению, называется ***несмещенной оценкой***.

Почему оценкой? Потому неизвестно реальное значение показателя (по генеральной совокупности), и с помощью выборочного наблюдения делается попытка его оценки. ***Оценка показателя – это есть его характеристика, рассчитанная по выборке***.

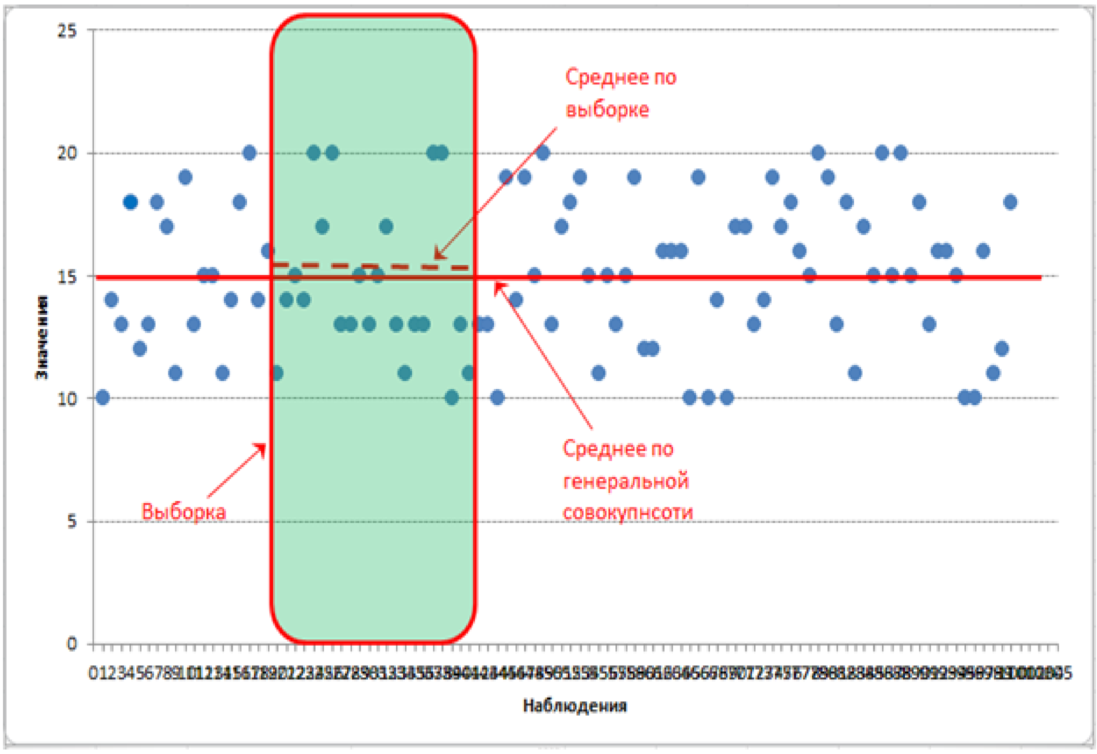
Выборочная средняя – это **несмещенная оценка** математического ожидания, так как средняя из выборочных средних стремится к своему теоретическому значению по генеральной совокупности. Она расположена в центре выборки. Средняя всегда находится в центре значений, по которым рассчитана – на то она и средняя.

А раз выборочная средняя находится в центре выборки, то из этого следует, что ***сумма квадратов расстояний от каждого значения выборки до выборочной средней всегда меньше, чем до любой другой точки, в том числе и до генеральной средней***.

Это ключевой момент. А раз так, то ***дисперсия в каждой выборке будет занижена***.

Средняя из заниженных дисперсий также даст заниженное значение. То есть при многократном повторении эксперимента выборочная дисперсия не будет стремиться к своему истинному значению (как выборочная средняя), а будет смещена относительно истинного значения по генеральной совокупности.

Отклонение выборочной средней от генеральной показано на рисунке.



Чтобы решить проблему смещенности выборочной дисперсии, в ее расчет вносят корректировку – умножают на ***n/(n-1)***, либо сразу при расчете в знаменатель ставят не ***n***, а ***n-1***.

Исправленная выборочная дисперсия
(она же ***несмещённая*** оценка генеральной дисперсии):

$$\varepsilon^2 = \frac{n}{n-1} D_e$$

Следует отметить, что для большой выборки (от 100 и даже от 30 вариант) этой поправкой можно пренебречь, так как при $n \to \infty$ дробь ***n/(n-1)*** стремится к единице и $D_e \to \varepsilon^2$.

Более того, встречаются задачи, где вообще не понятно – выборочная ли дана совокупность или генеральная, и тогда разумно проявить аккуратность и использовать обозначения без подстрочных индексов, в частности, $\bar{x}$ и ***D***.

Задание 1

В результате 10 независимых измерений некоторой величины, выполненных с одинаковой точностью, полученные опытные данные, которые представлены в таблице. Требуется вычислить среднее линейное отклонение

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*

Методика выполнения

4	x_1	$\lvert x_i - x_e \rvert$
5	7,1	**0,77**
6	6,3	**0,03**
7	6,2	**0,13**
8	5,8	**0,53**
9	7,7	**1,37**
10	6,8	**0,47**
11	6,7	**0,37**
12	5,9	**0,43**
13	5,7	**0,63**
14	5,1	**1,23**
$x_e =$	**6,33**	
$\overleftarrow{l} =$		**0,596**

=ABS(B5-B15)

=СРЗНАЧ(B5:B14)

=СУММ(C5:C14)/10
Ср. линейное
отклонение

Вычисленное ***среднее линейное отклонение*** означает, что измеренные значения $\boldsymbol{x_i}$ в среднем отличаются от $\bar{x}_в = 6{,}33$ примерно на 0,6 ед.

Варианты задания 1

№ вар.	Результаты измерений, x_i									
1	5,79	3,15	4,02	3,80	5,77	2,09	4,50	2,71	5,68	3,97
2	3,31	4,96	5,36	3,52	5,73	2,13	3,94	2,59	3,00	4,14
3	4,57	5,54	4,88	3,43	4,44	5,14	2,36	2,13	4,28	5,58
4	3,22	3,42	4,27	5,70	4,52	3,63	4,52	4,79	4,37	5,42
5	3,28	4,54	3,32	4,74	3,12	2,42	4,76	5,35	2,02	4,62
6	3,89	4,27	5,23	4,47	3,78	4,53	4,24	4,15	5,12	5,34
7	4,76	3,10	5,46	2,16	5,44	4,87	5,16	5,21	5,74	4,68
8	2,41	5,21	4,19	5,74	4,26	2,07	3,66	5,42	3,76	2,89
9	4,05	2,89	4,09	3,00	4,90	2,82	2,06	4,83	5,76	2,58
10	2,14	2,14	3,15	5,70	5,62	5,52	5,42	5,09	4,92	4,30
11	3,56	2,31	4,15	5,41	4,94	3,72	3,66	2,08	4,85	3,22
12	3,02	5,58	5,90	4,51	5,76	4,46	2,14	4,16	4,44	4,94
13	5,75	5,23	3,09	4,19	3,55	5,41	2,92	5,41	5,67	4,15
14	2,32	2,81	3,94	2,38	5,18	2,63	2,98	2,65	3,88	2,28
15	3,80	2,50	3,84	2,05	2,71	2,57	5,79	5,88	2,18	5,84
16	4,39	4,18	3,06	3,96	5,19	5,67	2,63	4,08	4,31	5,88
17	5,98	5,58	4,25	3,84	5,88	2,80	3,46	3,07	4,68	4,92
18	3,52	5,23	2,12	3,40	5,26	4,68	4,77	2,18	5,55	3,34
19	5,35	3,32	4,18	4,81	5,51	4,71	4,63	2,54	2,38	3,55
20	2,89	5,76	3,80	2,25	5,85	5,61	3,06	3,29	4,65	2,19
21	4,45	5,10	2,91	3,22	5,92	2,60	2,46	2,52	5,06	5,65
22	2,03	4,47	3,40	5,92	5,99	4,05	2,99	5,49	5,22	3,90
23	4,84	2,40	3,16	3,34	2,06	5,56	2,55	2,07	5,97	5,88
24	4,22	4,77	4,55	5,54	2,93	3,15	5,06	2,22	4,13	3,26
25	2,03	5,07	3,35	2,31	2,03	5,02	3,66	2,15	4,56	3,18

Задание 2

В результате выборочного исследования звонков, статистик МТС получил следующие данные за некоторый временной промежуток (дан готовый вариационный ряд).

Найти размах вариации, среднее линейное отклонение и выборочную дисперсию. Дать несмещённую оценку генеральной дисперсии.

Длительность соединеия, мин., t_i	Количество звонков, n_i
0	8
1	18
2	11
3	7
4	4
5	2

Методика выполнения

$\boldsymbol{R} = \boldsymbol{t_{max}} - \boldsymbol{t_{min}}$ = 5-0 = 5 мин.

Длит. соед, мин., t_i	Кол. звонков, n_i	$t_i\, n_i$	$\lvert t_i - \bar{t}_e \rvert$	$\lvert t_i - \bar{t}_e \rvert \cdot n_i$	$(t_i - \bar{t}_e)^2$	$(t_i - \bar{t}_e)^2 \cdot n_i$
0	8	0	1,74	13,92	3,0276	24,2208
1	18	18	0,74	13,32	0,5476	9,8568
2	11	22	0,26	2,86	0,0676	0,7436
3	7	21	1,26	8,82	1,5876	11,1132
4	4	16	2,26	9,04	5,1076	20,4304
5	2	10	3,26	6,52	10,6276	21,2552
∑=	50	87		54,48		87,62
$\bar{t}_e = \frac{\sum_{i=1}^{k=5} t_i n_i}{n} =$	***1,74***		$\bar{l} = \frac{\sum_{i=1}^{k=5} \lvert t_i - \bar{t}_e \rvert \bullet n_i}{n} =$	***1,0896***	$De = \frac{\sum_{i=1}^{k=5} (t_i - \bar{t}_e)^2 \bullet n_i}{n} =$	***1,7524***
Выбороч. средняя, мин.			Ср. линейное отклонение, мин.		Выбороч. дисперсия, мин.^{2}	

Несмещённой оценкой генеральной дисперсии является исправленная выборочная дисперсия:

$\varepsilon^2 = \frac{n}{n-1} D_e$ = (50/49)*1,7524 =1,7882 мин.^{2}

Варианты задания 2

№ вар.	*Длительность соединения, мин., t_i*						
	0	*1*	*2*	*3*	*4*	*5*	*6*
	Количество звонков, n_i						
1	7	16	7	11	10	8	16
2	7	2	17	17	10	9	14
3	8	16	2	8	16	18	4
4	10	4	4	3	14	6	18
5	11	3	4	14	7	9	11
6	12	4	7	16	3	6	5
7	16	11	18	12	13	12	14
8	11	7	7	14	8	12	16
9	16	6	16	3	14	14	8
10	2	15	10	2	13	3	16
11	7	6	16	9	13	11	6
12	17	10	11	16	13	13	6
13	16	10	8	8	10	7	12
14	16	7	8	16	18	13	16
15	7	11	7	10	13	16	9
16	6	15	17	8	18	8	4
17	12	5	7	2	13	7	13
18	17	17	2	8	3	13	10
19	4	15	17	10	12	6	13
20	5	12	12	7	5	15	4
21	10	14	11	17	10	18	17
22	10	7	16	18	15	16	16
23	17	10	13	10	4	10	14
24	11	8	4	16	7	13	12
25	18	17	3	7	9	11	8