СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

§ 1. Точечные оценки

Статистической оценкой Θ^* неизвестного параметра Θ теоретического распределения называют функцию $f(X_1, X_2, \ldots, X_n)$ от наблюдаемых случайных величин $X_1, X_2, \ldots, X_n$.

Точечной называют статистическую оценку, которая определяется одним числом $\Theta^* = f(x_1, x_2, \ldots, x_n)$, где $x_1, x_2, \ldots, x_n$ — результаты n наблюдений над количественным признаком X (выборка).

Несмещенной называют точечную оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки.

Смещенной называют точечную оценку, математическое ожидание которой не равно оцениваемому параметру.

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя

$$\bar{x}_в = \left(\sum_{i=1}^{k} n_i x_i\right) \Big/ n,$$

где x_i — варианта выборки, n_i — частота варианты x_i, $n = \sum_{i=1}^{k} n_i$ — объем выборки.

Замечание 1. Если первоначальные варианты x_i — большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число C, т. е. перейти к условным вариантам $u_i = x_i - C$ (в качестве C выгодно принять число, близкое к выборочной средней; поскольку выборочная средняя неизвестна, число C выбирают «на глаз»). Тогда

$$\bar{x}_в = C + \left(\sum n_i u_i\right) / n.$$

Смещенной оценкой генеральной дисперсии служит выборочная дисперсия

$$D_в = \left(\sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2\right) \Big/ n;$$

эта оценка является смещенной, так как

$$M[D_в] = \frac{n-1}{n} D_г.$$

Более удобна формула

$$D_в = \overline{x^2} - [\bar{x}]^2 = \frac{\sum n_i x_i^2}{n} - \left[\frac{\sum n_i x_i}{n}\right]^2.$$

Замечание 2. Если первоначальные варианты x_i — большие числа, то целесообразно вычесть из всех вариант одно и то же число C, равное выборочной средней или близкое к ней, т. е. перейти к условным вариантам $u_i = x_i - C$ (дисперсия при этом не изменится). Тогда

$$D_{\text{в}}(X)=D_{\text{в}}(u)=\overline{u^2}-[\bar{u}]^2=\frac{\sum n_i u_i^2}{n}-\left[\frac{\sum n_i u_i}{n}\right]^2.$$

Замечание 3. Если первоначальные варианты являются десятичными дробями с k десятичными знаками после запятой, то, чтобы избежать действий с дробями, умножают первоначальные варианты на постоянное число $C=10^k$, т. е. переходят к условным вариантам $u_i=Cx_i$. При этом дисперсия увеличится в C^2 раз. Поэтому, найдя дисперсию условных вариант, надо разделить ее на C^2:

$$D_{\text{в}}(X)=D_{\text{в}}(u)/C^2.$$

Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия

$$s^2=\frac{n}{n-1}\cdot D_{\text{в}}=\frac{\sum n_i(x_i-\bar{x}_{\text{в}})^2}{n-1}.$$

Более удобна формула

$$s_X^2=\frac{\sum n_i x_i^2-\left[\sum n_i x_i\right]^2/n}{n-1}.$$

В условных вариантах она имеет вид

$$s_u^2=\frac{\sum n_i u_i^2-\left[\sum n_i u_i\right]^2/n}{n-1},$$

причем если $u_i=x_i-C$, то $s_X^2=s_u^2$; если $u_i=Cx_i$, то $s_X^2=s_u^2/C^2$.

1. Из генеральной совокупности извлечена выборка объема $n=50$:

варианта x_i	2	5	7	10
частота n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

Решение. Несмещенной оценкой генеральной средней является выборочная средняя

$$\bar{x}_{\text{в}}=\left(\sum n_i x_i\right)/n=(16\cdot 2+12\cdot 5+8\cdot 7+14\cdot 10)/50=5{,}76.$$

2. Из генеральной совокупности извлечена выборка объема $n=60$:

x_i	1	3	6	26
n_i	8	40	10	2

Найти несмещенную оценку генеральной средней.

3. Найти выборочную среднюю по данному распределению выборки объема $n=10$:

x_i	1250	1270	1280
n_i	2	5	3

Р е ш е н и е. Первоначальные варианты — большие числа, поэтому перейдем к условным вариантам $u_i = x_i - 1270$. В итоге получим распределение условных вариант:

u_i	-20	0	10
n_i	2	5	3

Найдем искомую выборочную среднюю:

$$\bar{x}_в = C + \frac{\sum n_i u_i}{n} = 1270 + \frac{2\cdot(-20) + 5\cdot 0 + 3\cdot 10}{10} = 1270 - 1 = 1269.$$

4. Найти выборочную среднюю по данному распределению выборки объема $n = 20$:

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

У к а з а н и е. Перейти к условным вариантам $u_i = x_i - 2620$.

5. По выборке объема $n = 41$ найдена смещенная оценка $D_в = 3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

Р е ш е н и е. Искомая несмещенная оценка равна исправленной дисперсии:

$$s^2 = \frac{n}{n-1} D_в = \frac{41}{40}\cdot 3 = 3{,}075.$$

6. По выборке объема $n = 51$ найдена смещенная оценка $D_в = 5$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

7. В итоге пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): 92; 94; 103; 105; 106. Найти: а) выборочную среднюю длину стержня; б) выборочную и исправленную дисперсии ошибок прибора.

Р е ш е н и е. а) Найдем выборочную среднюю:

$$\bar{x}_в = 92 + (0 + 2 + 11 + 13 + 14)/5 = 92 + 8 = 100.$$

б) Найдем выборочную дисперсию:

$$D_в = \frac{\sum (x_i - \bar{x}_в)^2}{n} = [(92-100)^2 + (94-100)^2 + (103-100)^2]/5 +$$
$$+ [(105-100)^2 + (106-100)^2]/5 = 34.$$

Найдем исправленную дисперсию:

$$s^2 = \frac{n}{n-1}\cdot D_в = \frac{5}{4}\cdot 34 = 42{,}5.$$

8. В итоге четырех измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 8; 9; 11; 12. Найти: а) выборочную среднюю результатов измерений; б) выбо-

рочную и исправленную дисперсии ошибок прибора.

9. Ниже приведены результаты измерения роста (в см) случайно отобранных 100 студентов.

Рост	154—158	158—162	162—166	166—170	170—174	174—178	178—182
Число студентов	10	14	26	28	12	8	2

Найти выборочную среднюю и выборочную дисперсию роста обследованных студентов.

Указание. Найти середины интервала и принять их в качестве вариант.

10. Найти выборочную дисперсию по данному распределению выборки объема $n=10$:

$$\begin{array}{cccc} x_i & 186 & 192 & 194 \\ n_i & 2 & 5 & 3 \end{array}$$

Решение. Варианты—сравнительно большие числа, поэтому перейдем к условным вариантам $u_i=x_i-191$ (мы вычли из вариант число $C=191$, близкое к выборочной средней). В итоге получим распределение условных вариант:

$$\begin{array}{cccc} u_i & -5 & 1 & 3 \\ n_i & 2 & 5 & 3 \end{array}$$

Найдем искомую выборочную дисперсию:

$$D_{\text{в}}=(\sum n_iu_i^2)/n-[(\sum n_iu_i)/n]^2=(2\cdot5^2+5\cdot1^2+3\cdot3^2)/10-$$
$$-[(2\cdot(-5)+5\cdot1+3\cdot3)/10]^2=8,2-0,16=8,04.$$

11. Найти выборочную дисперсию по данному распределению выборки объема $n=100$:

$$\begin{array}{ccccc} x_i & 340 & 360 & 375 & 380 \\ n_i & 20 & 50 & 18 & 12 \end{array}$$

Указание. Перейти к условным вариантам $u_i=x_i-360$.

12. Найти выборочную дисперсию по данному распределению выборки объема $n=100$:

$$\begin{array}{ccccc} x_i & 2502 & 2804 & 2903 & 3028 \\ n_i & 8 & 30 & 60 & 2 \end{array}$$

Указание. Перейти к условным вариантам $u_i=x_i-2844$.

13. Найти выборочную дисперсию по данному распределению выборки объема $n=10$:

$$\begin{array}{cccc} x_i & 0,01 & 0,04 & 0,08 \\ n_i & 5 & 3 & 2 \end{array}$$

Р е ш е н и е. Для того чтобы избежать действий с дробями, перейдем к условным вариантам $u_i=100x_i$. В итоге получим распределение

u_i	1	4	8
n_i	5	3	2

Найдем выборочную дисперсию условных вариант:

$$D_в(u)=(\sum n_iu_i^2)/n-[(\sum n_iu_i)/n]^2.$$

Подставив в эту формулу условные варианты и их частоты, получим

$$D_в(u)=7,21.$$

Найдем искомую выборочную дисперсию первоначальных вариант:

$$D_в(X)=D_в(u)/100^2=7,21/10\,000=0,0007.$$

14. Найти выборочную дисперсию по данному распределению выборки объема $n=50$:

x_i	0,1	0,5	0,6	0,8
n_i	5	15	20	10

У к а з а н и е. Перейти к условным вариантам $u_i=10x_i$.

15. Найти выборочную дисперсию по данному распределению выборки объема $n=50$:

x_i	18,4	18,9	19,3	19,6
n_i	5	10	20	15

У к а з а н и е. Перейти к условным вариантам $u_i=10x_i-195$.

16. Найти исправленную выборочную дисперсию по данному распределению выборки $n=10$:

x_i	102	104	108
n_i	2	3	5

Р е ш е н и е. Перейдем к условным вариантам $u_i=x_i-104$. В итоге получим распределение

u_i	−2	0	4
n_i	2	3	5

Найдем исправленную выборочную дисперсию условных вариант:

$$s_u^2=\frac{\sum n_iu_i^2-[\sum n_iu_i]^2/n}{n-1}.$$

Подставив в эту формулу условные варианты, их частоты и объем выборки, получим $s_u^2=6,93$.

Все первоначальные варианты были уменьшены на одно и то же постоянное число $C=104$, поэтому дисперсия не изменилась, т. е. искомая дисперсия равна дисперсии условных вариант: $s_X^2=s_u^2=6,93$.

17. Найти исправленную выборочную дисперсию по данному распределению выборки объема $n=100$:

x_i	1250	1275	1280	1300
n_i	20	25	50	5

Указание. Перейти к условным вариантам $u_i=x_i-1275$.

18. Найти исправленную выборочную дисперсию по данному распределению выборки объема $n=10$:

x_i	0,01	0,05	0,09
n_i	2	3	5

Решение. Для того чтобы избежать действий с дробями, перейдем к условным вариантам $u_i=100x_i$. В итоге получим распределение

u_i	1	5	9
n_i	2	3	5

Найдем исправленную выборочную дисперсию условных вариант

$$s_u^2=\frac{\sum n_i u_i^2-\left[\sum n_i u_i\right]^2/n}{n-1}.$$

Подставив в эту формулу данные задачи, получим

$$s_u^2=10,844.$$

Найдем искомую исправленную дисперсию первоначальных вариант:

$$s_X^2=s_u^2/100^2=10,844/10\,000\simeq 0,0085.$$

19. Найти исправленную выборочную дисперсию по данному распределению выборки объема $n=20$:

x_i	0,1	0,5	0,7	0,9
n_i	6	12	1	1

Указание. Перейти к условным вариантам $u_i=10x_i$.

20. Составлена таблица результатов сессии для *100* студентов, успешно сдавших *4* экзамена:

Сумма баллов z_i	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
Число студентов n_i	*1*	*3*	*7*	*15*	*21*	*30*	*12*	*8*	*3*

Найти выборочное среднее $\overline{x}$, выборочную дисперсию $D_в$ и исправленную выборочную дисперсию s^2 для величины X – суммы баллов.

Решение. Выборка записана в виде статистического ряда ($k = 9$), объём выборки

$$n = 1 + 3 + 7 + 15 + 21 + 30 + 12 + 8 + 3 = 100.$$

Вычисляем выборочное среднее:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{k} z_i n_i=\frac{1}{100}(12\cdot 1+13\cdot 3+14\cdot 7+15\cdot 15+16\cdot 21+$$

$$+16\cdot 21+17\cdot 30+18\cdot 12+19\cdot 8+20\cdot 3)=16,48$$

Для вычисления выборочной дисперсии найдём вначале

$$\overline{x^2}=\frac{1}{n}\sum_{i=1}^{k} z_i^2 n_i=\frac{1}{100}\left(12^2\cdot 1+13^2\cdot 3+14^2\cdot 7+15^2\cdot 15+\right.$$

$$\left.+16^2\cdot 21+17^2\cdot 30+18^2\cdot 12+19^2\cdot 8+20^2\cdot 3\right)=\frac{27420}{100}=274,2.$$

Тогда $D_в=\overline{x^2}-\left(\overline{x}\right)^2=274,2-(16,48)^2=2,6096\approx 2,61$.

Исправленная выборочная дисперсия

$$s^2=\frac{n}{n-1}D_в=\frac{100}{100-1}\cdot 2,6096\approx 2,64.$$

Замечание. Если варианты z_i (или z_i^*, если речь идёт о группированной выборке) – *большие числа*, то для упрощения расчёта выборочного среднего $\overline{x}$ и исправленной выборочной дисперсии s^2 целесообразно перейти к так называемым ***условным вариантам*** $u_i=\frac{z_i-c}{h}$, где в качестве c обычно берут варианту с наибольшей частотой, а h – некая масштабная величина. Например, если варианты z_i равноотстоящие, то удобно взять $h=z_i-z_{i-1}$. Тогда, легко проверить, $\overline{x}=c+\overline{u}h$, $s_x^2=h^2 s_u^2$.

Так в рассмотренном примере 3.6 можно было рассмотреть условные варианты $u_i = z_i - 17$. Для вычисления удобно составить таблицу:

i	z_i	u_i	n_i	$u_i n_i$	$u_i^2 n_i$
1	12	–5	1	–5	25
2	13	–4	3	–12	48
3	14	–3	7	–21	63
4	15	–2	15	–30	60
5	16	–1	21	–21	21
6	17	0	30	0	0
7	18	1	12	12	12
8	19	2	8	16	32
9	20	3	3	9	27
Σ	–	–	100	–52	288

Итак,

$$\overline{u}=\frac{1}{n}\sum n_i u_i=\frac{-52}{100}=-0,52\,,\quad \overline{u^2}=\frac{1}{n}\sum n_i u_i^2=\frac{288}{100}=2,88\,,$$

$$s_u^2=\frac{n}{n-1}\left(\overline{u^2}-(\overline{u})^2\right)=\frac{100}{99}\left(2,88-(-0,52)^2\right)\approx 2,64\,.$$

Окончательно, $\overline{x}=17+(-0,52)=16,48$, $s_x^2=s_u^2\approx 2,64$, что совпадает с полученными результатами в примере 20.

21. Вычислить среднее $\overline{x}$ и исправленную выборочную дисперсию s_x^2 группированной выборки

Интервалы	$[134, 138)$	$[138, 142)$	$[142, 146)$	$[146, 150)$	$[150, 154)$	$[154, 158]$
Частоты	*1*	*3*	*14*	*18*	*12*	*2*

Решение. Здесь шаг разбиения $h=4$, объём выборки $n=50$, значение середины интервала, встречающегося с наибольшей частотой, $c=148$. Таким образом, преобразование последовательности середин интервалов выполняется по формуле

$$u_i=\frac{z_i^*-148}{4},\quad i=1,\,2,\,\dots,\,6\,.$$

Вычисления удобно свести в таблицу:

i	z_i^*	u_i	n_i^*	$u_i\,n_i^*$	$u_i^2 n_i^*$
1	136	–3	1	–3	9
2	140	–2	3	–6	12
3	144	–1	14	–14	14
4	148	0	18	0	0
5	152	1	12	12	12
6	156	2	2	4	8
Σ	–	–	50	–7	55

Проводим вычисления: $\overline{u}=\dfrac{1}{50}\sum\limits_{i=1}^{6}u_i n_i^*=\dfrac{-7}{50}=-0,14$;

$$s_u^2=\frac{50}{49}\left(\frac{1}{50}\sum_{i=1}^{6}u_i^2 n_i^*-(\overline{u})^2\right)=\frac{50}{49}\left(\frac{55}{50}-(-0,14)^2\right)\approx 1,10\,;$$

$$\overline{x}=c+\overline{u}h=148+(-0,14)\cdot 4=147,44\,;\quad s_x^2=h^2 s_u^2=4^2\cdot 1,10\approx 17,63\,.$$