ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 8.

1. ПЕРВИЧНАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА И ОЦЕНКА ОСНОВНЫХ ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Под *генеральной совокупностью* в математической статистике понимается множество (гипотетическое) всех возможных результатов измерения некоторой величины, которые могут быть получены в данных условиях. Реальная серия повторных измерений этой величины $x_1, x_2, ..., x_n$ трактуется как случайная выборка из генеральной совокупности, или просто *случайная выборка*.

В статистике принята следующая математическая модель подобных экспериментов. *Каждый элемент случайной выборки рассматривается как отдельная случайная величина.* Относительно этих случайных величин, которые в дальнейшем будем обозначать заглавными буквами, известна некоторая априорная информация.

Случайная выборка называется *повторной*, если все входящие в нее случайные величины $X_1, X_2, ..., X_n$ взаимно независимы и имеют одинаковую функцию распределения $F(x)$, причем ту же, что и наблюдаемая случайная величина X. Распределение случайной величины X характеризуется рядом *параметров* (математическое ожидание, дисперсия и т.д.). Эти параметры называют *параметрами генеральной совокупности.* Важной задачей математической статистики является нахождение по случайной выборке приближенных значений каждого из параметров, называемых *точечными оценками параметров*, или просто *оценками*. Таким образом, *оценкой параметра* β называется функция $f(X_1, X_2, ..., X_n)$ от случайной выборки, значение которой принимается в качестве приближенного для данного параметра и обозначается $\tilde{\beta}$.

$$\beta \approx \tilde{\beta} = f(X_1, X_2, ..., X_n). \qquad (1.1)$$

Пусть задана повторная случайная выборка $X_1, X_2, \ldots, X_n$. За *оценку математического ожидания* a принимается среднее арифметическое элементов выборки:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i . \qquad (1.3)$$

Оценкой дисперсии σ^2 при *неизвестном математическом ожидании* является величина S^2, которую называют *эмпирической дисперсией*:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 . \qquad (1.5)$$

Оценкой среднего квадратического отклонения σ при этом является, соответственно, величина $S = \sqrt{S^2}$.

Для практических расчетов формулу (1.5) удобно преобразовать к виду:

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right). \tag{1.7}$$

Вычисление среднего значения $\overline{X}$ и оценки дисперсии S^2 упрощается, если отсчет значений X_i вести от подходящим образом выбранного начала отсчета C и в подходящем масштабе, то есть, если сделать линейную замену:

$$X_i = C + hU_i (i = 1, 2, ..., n). \tag{1.8}$$

При такой замене формулы (1.3) и (1.5) – (1.6) принимают вид:

$$\overline{X} = C + h\overline{U}, \quad \overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i \ ; \tag{1.9}$$

$$S^2 = \frac{h^2}{n-1}\sum_{i=1}^{n}\left(U_i - \overline{U}\right)^2 = \frac{h^2}{n-1}\left(\sum_{i=1}^{n} U_i^2 - n\overline{U}^2\right). \tag{1.10}$$

Для контроля вычислений весь расчет повторяют с другим началом отсчета C, результаты должны совпадать с точностью до возможных ошибок округления.

Задача 1.1. В табл. 1.1 в первом столбце записаны результаты n = 18 независимых равноточных измерений величины заряда электрона $q = x \cdot 10^{-10}$ (в единицах *CGSE*), полученных Милликеном. Вычислить оценки математического ожидания, дисперсии и среднего квадратического отклонения величины X , провести контроль расчетов.

Решение

Выбираем C = 4,780 и, полагая h = 10^{-3}, подсчитываем значения $u_i = (x_i - C)/h = (x_i - 4,780)/10^{-3}$ и u_i^2. Суммы чисел второго и третьего столбца дают возможность рассчитать $\overline{X}$ и S^2:

$$\overline{U} = 13/18 = 0,72, \quad \overline{X} = 4,780 + 0,72 \cdot 10^{-3} = 4,7807;$$

$$S^2 = 10^{-6}\left(6239 - 13^2/18\right)/17 = 3,66 \cdot 10^{-4},$$

откуда $S = \sqrt{S^2} = \sqrt{3,66 \cdot 10^{-4}} = 1,91 \cdot 10^{-2}$. В последних двух столбцах приведены расчеты при другом начале отсчета C_1 = 4,790, то есть при замене $V_i = (X_i - 4,790)/10^{-3}$. Эти расчеты дают те же значения $\overline{X}$ и S:

$$\bar{V} = 167/18 = -9,2, \quad \overline{X} = 4,790 - 9,28 \cdot 10^{-3} = 4,7809;$$

$$S^2 = 10^{-6}\left(7779 - 167^2/18\right)/17 = 3,66 \cdot 10^{-4}.$$

Таблица 1.1.

Исходные данные и результаты расчета к задаче 1.1.

Исходные данные X	Расчет		Контроль	
	U	U^2	V	V^2
4,761	–19	361	–29	841
4,792	12	144	2	4
4,758	–22	484	–32	1024
4,764	–16	256	–26	676
4,810	30	900	20	400
4,799	19	361	9	81
4,797	17	289	7	49
4,790	10	100	0	0
4,747	–33	1089	–43	1849
4,769	–11	121	–21	441
4,806	26	676	16	256
4,779	–1	1	–11	121
4,785	5	25	–5	25
4,790	10	100	0	0
4,777	–3	9	–13	169
4,749	–31	961	–41	1686
4,781	1	1	–9	81
4,799	19	361	9	81
Сумма	13	6239	–167	7779

Ответ: $\overline{X} = 4,7809$; $S^2 = 3,66 \cdot 10^{-4}$; $S = 1,91 \cdot 10^{-2}$.

2. ОЦЕНКА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ ПО НЕРАВНОТОЧНЫМ ИЗМЕРЕНИЯМ.

Часто встречающимся на практике случаем неповторной выборки является выборка, в которой случайные величины $X_1, X_2, \ldots, X_n$ взаимно независимы, имеют одинаковые математические ожидания, но различные дисперсии. Такие измерения называют *неравноточными.* Как правило, дисперсии каждой величины X_i не известны, но известны отношения дисперсий. Числа, обратно пропорциональные дисперсиям, называют *весами измерений* и обозначают w_i:

$$D(X_1):D(X_2):\ldots:D(X_n)=(1/w_1):(1/w_2):\ldots:(1/w_n),$$

или

$$D(X_i)=\sigma^2/w_i,\ \ (i=1,2,\ldots,n). \qquad (1.13)$$

Коэффициент σ^2 в формуле (1.13) обычно не известен, он называется *дисперсией измерения с единичным весом*, веса w_i, как правило, известны.

Среднее арифметическое (1.3) для неравноточных измерений является несмещенной и состоятельной оценкой математического ожидания, но не является наилучшей линейной оценкой. Наилучшей линейной оценкой в этом случае будет *среднее взвешенное*:

$$\overline{X}_{взв}=\frac{\left(\sum\limits_{i=1}^{n}X_i w_i\right)}{\sum\limits_{i=1}^{n}w_i}, \qquad (1.14)$$

эта оценка будет несмещенной и состоятельной. Она и используется на практике для неравноточных измерений.

Задача 1.2 В табл. 1.3 в первом столбце записаны результаты n = 5 независимых случайных величин X_i, являющихся средними арифметическими пяти серий измерений. В каждой серии измерения независимы, имеют одинаковые математические ожидания и дисперсии, т.е. равноточны $D(X_{ij})=\sigma^2$, где i – номер серии; j – номер измерения в серии; n_i – число измерений в серии. Найти наилучшую оценку математического ожидания.

Таблица 1.3

Исходные данные и результаты расчета к задаче 1.3.

i	X_i	n_i	$X_i n_i$
1	2,41	5	12,05
2	2,83	2	5,66
3	2,62	4	10,48
4	2,49	6	14,94
5	2,75	3	8,25
Σ	13,10	20	51,38

Решение

Среднее арифметическое результатов X_i будет $\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ = 13,10/5 = 2,62. Получена несмещенная оценка математического ожидания, но она не является наилучшей линейной оценкой, так как результаты измерений неравноточны. Среднее арифметическое по каждой серии $X_i = \sum_{j=1}^{n_i} X_{ij} / n_i$. Если измерения $X_{ij\,j}$ равноточны, то дисперсии средних арифметических равны $D(X_i) = \sigma^2 / n_i$ и величины x_i неравноточны. Сравнивая $D(X_i)$ с формулой (1.13), делаем вывод, что весами измерений в этом случае являются числа измерений n_i, то есть $w_i = n_i$. Используя формулу (1.14), получаем $\overline{X}_{взв} = 51{,}38 / 20 = 2{,}569$. Это значение и будет наилучшей линейной оценкой, то есть, имеющей наименьшую погрешность.

3. ОЦЕНКА ДИСПЕРСИИ ПО НЕСКОЛЬКИМ СЕРИЯМ ЭКСПЕРИМЕНТОВ

Пусть заданы L независимых повторных выборок – L серий измерений. Случайные величины различных выборок имеют, в общем случае, различные математические ожидания, но дисперсии всех величин во всех выборках одинаковы. Такая ситуация возникает, когда одним и тем же прибором производят измерения различных величин (например, измерения значений функции для различных значений аргумента).

В этом случае для оценки единой дисперсии можно использовать значения измерений всех серий. По каждой выборке находят эмпирическую дисперсию S_j^2 с числом степеней свободы k_j (j = 1, 2, ..., L). В качестве оценки единой дисперсии принимают сводную эмпирическую дисперсию:

$$S_{\text{св}}^2 = \left(\sum_{j=1}^{L} S_j^2 k_j \right) \Bigg/ \sum_{j=1}^{L} k_j \qquad (1.15)$$

с числом степеней свободы $k_{\text{св}} = \sum_{j=1}^{L} k_j$. Сводная оценка дисперсии (1.15) является несмещенной, она более точная, чем каждая из эмпирических дисперсий S_j^2.

Задача 1.3. При изучении зависимости предела прочности от размера (D) зерна рекристаллизованного металла замеры производились независимо на разных образцах и предположительно с одинаковой точностью (табл. 1.4). Оценить эту точность, то есть, найти оценку дисперсии и оценку среднего квадратического отклонения. Число измерений прочности при различных значениях D различно.

Таблица 1.4

Экспериментальные данные к задаче 1.4.

D, мкм	Предел прочности, кг/мм^2				
20	48,9	48,8	48,7	49,0	49,2
60	46,1	46,2	46,6	46,4	
110	43,8	44,0	44,2		
120	43,6	44,0	43,7	43,8	43,8
160	42,0	42,4	42,2		
200	41,2	41,3	41,4	41,1	41,6

Решение

Вначале вычислим эмпирические дисперсии каждой серии замеров, то есть при каждом значении *D*. Прежде всего заметим, что приведенные в таблице 1.4 значения предела прочности удобно уменьшить на 40, полученные данные обозначим через *Y* (табл. 1.5). Чтобы вести расчет с небольшими целыми числами, закодируем значения *Y* по формуле $U = 10(Y - C)$, где за начало отсчета *C* в каждой серии примем число, набранное курсивом в табл. 1.5, например, при *D* = 20 примем *C* = 8,9, а при *D* = 60 примем *C* = 6,2.

Таблица 1.5

Результаты расчета к задаче 1.4.

D	*Y*	*U*=10(*Y* –C)	Σ*U*	Σ*U*^{2}	n	kS^2·10^2	k	S
20	***8,9*** 8,8 8,7 9,0 9,2	0 -1 -2 1 3	1	15	5	14,8	4	0,19
60	6,1 ***6,2*** 6,6 6,4	-1 0 4 2	5	21	4	14,75	3	0,22
110	3,8 ***4,0*** 4,2	-2 0 2	0	8	3	8	2	0,20
120	3,6 4,0 3,7 ***3,8*** ***3,8***	-2 2 -1 0 0	-1	9	5	8,8	4	0,15
160	2,0 2,4 ***2,2***	-2 2 0	0	8	3	8	2	0,20
200	1,2 ***1,3*** 1,4 1,1 1,6	-1 0 1 -2 3	1	15	5	14,8	4	0,19
Σ	– –	– –	–	–	–	69,15	19	

В этой таблице все расчеты ведутся по строкам. $kS^2 \cdot 10^2 = kS_u^2 = \; = \sum u^2 - \left(\sum u\right)^2 / n$,

где *n* – число замеров в серии. Суммы, подсчитанные в столбцах $kS^2 \cdot 10^2$ и *k*, позволяют получить оценки дисперсии и среднего квадратического отклонения:

$$S_{\text{св}}^2 = 0{,}1^2 \cdot 69{,}15/19 = 0{,}0364; \quad S_{\text{св}} = 0{,}1\sqrt{3{,}64} = 0{,}191.$$